Mathematical Programming
https://doi.org/10.1007/s10107-019-01451-7

FULL LENGTH PAPER

Series B

Problem-driven scenario generation: an analytical approach for stochastic programs with tail risk measure

Jamie Fairbrother[1] · **Amanda Turner**[1] · **Stein W. Wallace**[2]

Received: 15 January 2017 / Accepted: 11 November 2019

Abstract
Scenario generation is the construction of a discrete random vector to represent parameters of uncertain values in a stochastic program. Most approaches to scenario generation are *distribution-driven*, that is, they attempt to construct a random vector which captures well in a probabilistic sense the uncertainty. On the other hand, a *problem-driven* approach may be able to exploit the structure of a problem to provide a more concise representation of the uncertainty. In this paper we propose an analytic approach to problem-driven scenario generation. This approach applies to stochastic programs where a tail risk measure, such as conditional value-at-risk, is applied to a loss function. Since tail risk measures only depend on the upper tail of a distribution, standard methods of scenario generation, which typically spread their scenarios evenly across the support of the random vector, struggle to adequately represent tail risk. Our scenario generation approach works by targeting the construction of scenarios in areas of the distribution corresponding to the tails of the loss distributions. We provide conditions under which our approach is consistent with sampling, and as proof-of-concept demonstrate how our approach could be applied to two classes of problem, namely network design and portfolio selection. Numerical tests on the portfolio selection problem demonstrate that our approach yields better and more stable solutions compared to standard Monte Carlo sampling.

Mathematics Subject Classification 90C15 (stochastic programming)

1 Introduction

Stochastic programming is a tool for making decisions under uncertainty. Under this modeling paradigm, uncertain parameters are modeled as a random vector, and one

✉ Jamie Fairbrother
j.fairbrother@lancaster.ac.uk

1 STOR-i Centre for Doctoral Training, Lancaster University, Lancaster, UK

2 Department of Business and Management Science, Norwegian School of Economics, Bergen, Norway

Published online: 26 November 2019

attempts to minimize (or maximize) the expectation or risk measure of some loss function which depends on the initial decision. However, what distinguishes stochastic programming from other stochastic modeling approaches is its ability to explicitly model future decisions based on outcomes of stochastic parameters and initial decisions, and the associated costs of these future decisions. The power and flexibility of the stochastic programming approach comes at a price: stochastic programs are usually analytically intractable, and often not susceptible to solution techniques for deterministic programs.

Typically, a stochastic program can only be solved when it is *scenario-based*, that is when the random vector for the problem has a finite discrete distribution. For example, stochastic linear programs become large-scale linear programs when the underlying random vector is discrete. In the stochastic programming literature, the mass points of this random vector are referred to as *scenarios*, the discrete distribution as the *scenario set* and the construction of this set as *scenario generation*. Scenario generation can consist of discretizing a continuous probability distribution, or directly modeling the uncertain quantities as discrete random variables. The more scenarios in a set, the more computational power that is required to solve the problem. The key issue of scenario generation is therefore how to represent the uncertainty to ensure that the solution to the problem is reliable, while keeping the number of scenarios low so that the problem is computationally tractable.

A common approach to scenario generation is to fit a statistical model to the uncertain problem parameters and then generate a random sample from this for the scenario set. This has desirable asymptotic properties [22,33], but may require large sample sizes to ensure the reliability of the solutions it yields. This can be mitigated somewhat by using variance reduction techniques such as stratified sampling and importance sampling [24]. Sampling also has the advantage that it can be used to construct confidence intervals on the true solution value [25]. Another approach is to construct a scenario set whose distance from the true distribution, with respect to some probability metric, is small [12,19,28]. These approaches tend to yield better and much more stable solutions to stochastic programs than does sampling.

A characteristic of these approaches to scenario generation is that they are *distribution-driven*; that is, they only aim to approximate a distribution and are divorced from the stochastic program for which they are producing scenarios. By exploiting the structure of a problem, it may be possible to find a more parsimonious representation of the uncertainty. Note that such a *problem-driven* approach may not yield a discrete distribution which is close to the true distribution in a probabilistic sense; the aim is only to find a discrete distribution which yields a high quality solution to our problem.

Stochastic programs often have the objective of minimizing the expectation of a loss function. This is particularly appropriate when the initial decision represents a strategic decision that is going to be used again and again, and individual large losses do not matter in the long term. For example, in a stochastic facility location problem (e.g. see [5]) the locations of several facilities must be chosen subject to the unknown demands of customers in a way which minimizes fixed investment costs, and future distribution costs. In other cases, the decision may be only used once or a few times, and the occurrence of large losses may have serious consequences such as bankruptcy. This is characteristic of the portfolio selection problem [26] studied

in detail in the latter part of this paper. In this latter case, minimizing the expectation alone is not appropriate as this does not necessarily mitigate against the possibility of large losses. One possible remedy is to use a *risk measure* which penalises in some way the likelihood and severity of potential large losses.

In this paper we are interested in stochastic programs which use *tail risk measures*. A precise definition of a tail-risk measure will be given in Sect. 3 but for now, one can think of a tail risk measure as a function of a random variable which only depends on the upper tail of its distribution function. Tail risk measures are useful as they summarize the extent of potential losses in the worst possible outcomes. Examples of tail risk measures include the Value-at-Risk (VaR) [21] and the Conditional Value-at-Risk (CVaR) [29], both of which are commonly used in financial contexts. Although the methodology developed in this paper can in principle be applied to any loss function, in this work we are mainly interested in loss functions which arise in one and two-stage stochastic programs.

Distribution-driven scenario generation methods are particularly problematic for stochastic programs involving tail risk measures. This is because these methods tend to spread their scenarios evenly across the support of distribution and so struggle to adequately represent the tail risk without using a potentially prohibitively large number of scenarios.

In this paper, we propose an analytic problem-driven approach to scenario generation applicable to stochastic programs which use tail risk measures of a form made precise in Sect. 3. We observe that the value of a tail risk measure depends only on scenarios confined to an area of the distribution that we call the *risk region*. This means that all scenarios that are not in the risk region can be aggregated into a single point. By concentrating scenarios in the risk region, we can calculate the value of a tail risk measure more accurately.

Given a risk region for a problem, we propose a simple algorithm for generating scenarios which we call *aggregation sampling*. This algorithm takes samples from the random vector until a specified number of samples in the risk region have been produced, and all other scenarios are aggregated into a single scenario. We provide and give proofs of conditions under which this method is asymptotically consistent with standard Monte Carlo sampling.

In general, finding a risk region is difficult as it is determined by the loss function, problem constraints and the distribution of the uncertain parameters. Therefore, we derive risk regions for two classes of problem as a proof-of-concept of our methodology. The first class of problems are those with monotonic loss functions which, as will be shown, occur naturally in the context of network design. The second class are portfolio selection problems. For both types of risk regions we run numerical tests which demonstrate that our methodology yields better quality solutions and with greater reliability than standard Monte Carlo sampling.

This paper is organized as follows: in Sect. 2 we discuss related work; in Sect. 3 we define tail risk measures and their associated risk regions; in Sect. 4 we discuss how these risk regions can be exploited for the purposes of scenario generation; in Sect. 5 we prove that our scenario generation method is consistent with standard Monte Carlo sampling; in Sects. 6 and 7 we derive risk regions for the two classes of problems

described above; in Sect. 8 we present numerical tests; finally in Sect. 9 we summarize our results and make some concluding remarks.

Notation Throughout this paper random variables and vectors are represented by bold (mainly Greek) letters: $\boldsymbol{\theta}$, $\boldsymbol{\xi}$, $\boldsymbol{\zeta}$ and outcomes of these are represented by the corresponding non-bold letters: θ, ξ, ζ. Inequalities used with vectors and matrices always apply component-wise. $\|\cdot\|$ represents the standard Euclidean norm.

2 Related work

There are relatively few cases of problem-driven scenario generation in the literature. The earliest example of which we are aware is the importance sampling approach of [8] which constructs a sampler from the loss function. Importance sampling has been used more recently for scenario generation for problems which, like our own, concern rare events. In [23] an importance sampling scheme is used for a multistage problem involving the CVaR risk measure. In [4], an importance sampling approach is proposed for chance-constrained stochastic programs where the permitted probabilities of constraint violation are very small.

There is an interesting connection between problem-driven scenario generation and distributionally robust optimization [11,38,39]. In distributionally robust optimization, the distribution of the random variables in a stochastic program is itself uncertain, and one must optimize for the worst-case distribution. Solving a distributionally robust optimization problem thus involves finding, at least implicitly, the worst-case distribution or scenario set for given objective and constraints. In this sense, distributionally robust optimization could be considered as a problem-driven scenario generation method. Of particular relevance for this work, the paper [9] solves a distributionally robust portfolio selection problem involving the CVaR risk measure where the distribution of asset returns has specified discrete marginals, but unknown joint distribution.

The idea that in stochastic programs with tail risk measures some scenarios do not contribute to the calculation of the tail-risk measure was also exploited in [17]. However, they propose a solution algorithm rather than a method of scenario generation. Their approach is to iteratively solve the problem with a subset of scenarios, identify the scenarios which have loss in the tail, update their scenario set appropriately and resolve, until the true solution has been found. Their method has the benefit that it is not distribution dependent. On the other hand, their method works for only the β-CVaR risk measure, while our approach works in principle for any tail risk measure.

3 Tail risk measures and risk regions

In this section we present the core theory to our scenario generation methodology. Specifically, in Sect. 3.1 we formally define tail-risk measures of random variables and in Sect. 3.2 we define risk regions and present some key results related to these.

3.1 Tail risk of random variables

In our set-up we suppose we have some random variable representing an uncertain loss. For our purposes, we take a risk measure to be any function of a random variable. The following formal definition is adapted from [35].

Definition 1 (*Risk measure*) Let $(\Omega, \mathbb{P})$ be a probability space, and Θ be the set of measurable real-valued random variables on $(\Omega, \mathbb{P})$. Then, a risk measure is some function $\rho : \Theta \to \mathbb{R} \cup \{\infty\}$.

For a risk measure to be useful, it should in some way penalize potential large losses. For example, in the classical Markowitz problem [26], one aims to minimize the variance of the return of a portfolio. By choosing a portfolio with a low variance, we reduce the probability of larges losses as a direct consequence of Chebyshev's inequality (see for instance [6]). Various criteria for risk measures have been proposed; in [3] a *coherent risk measure* is defined to be a risk measure which satisfies axioms such as positive homogeneity and subadditivity; another perhaps desirable criterion for risk measures is that the risk measure is consistent with respect to first and second order stochastic dominance, see [27] for instance.

Besides not satisfying some of the above criteria, a major drawback with using variance as a measure is that it penalizes all large deviations from the mean, that is, it penalizes large profits as well as large losses. This motivates the idea of using risk measures which depend only on the upper tail of the loss distribution. To formalize this idea, we first recall the definition of quantile function.

Definition 2 (*Quantile function*) Suppose $\boldsymbol{\theta}$ is a random variable with distribution function $F_{\boldsymbol{\theta}}$. Then the generalized inverse distribution function, or *quantile function* is defined as follows:

$$\begin{aligned} F_{\boldsymbol{\theta}}^{-1} : (0, 1] &\to \mathbb{R} \cup \{\infty\} \\ \beta &\mapsto \inf\{x \in \mathbb{R} : F_{\boldsymbol{\theta}}(x) \geq \beta\}. \end{aligned}$$

We refer to the quantile function evaluated at β, $F_{\boldsymbol{\theta}}^{-1}(\beta)$, as the β-*quantile*.

The β-quantile can be interpreted as the smallest value for which the distribution function is greater than or equal to β. The β-tail of a distribution is the restriction of the distribution function to values equal to or above the β-quantile. In the context of risk management, we typically have $0.9 \leq \beta < 1.0$. The following definition says that a tail risk measure is a risk measure that only depends on the β-tail of a distribution.

Definition 3 (*Tail risk measure*) Let $\rho_\beta : \Theta \to \mathbb{R} \cup \{\infty\}$ be a risk measure per Definition 1. Then ρ_β is a β-tail risk measure if $\rho_\beta(\boldsymbol{\theta})$ depends only on the restriction of quantile function of $\boldsymbol{\theta}$ above β, in the sense that if $\boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}$ are random variables with $F_{\boldsymbol{\theta}}^{-1}|_{[\beta,1]} = F_{\tilde{\boldsymbol{\theta}}}^{-1}|_{[\beta,1]}$ then $\rho_\beta(\boldsymbol{\theta}) = \rho_\beta(\tilde{\boldsymbol{\theta}})$.

To show that ρ_β is a β-tail risk measure, we must show that $\rho_\beta(\boldsymbol{\theta})$ can be written as a function of the quantile function above or equal to β. Two very popular tail risk measures are the value-at-risk [21] and the conditional value-at-risk [30]:

Example 1 (*Value at risk*) Let $\boldsymbol{\theta}$ be a random variable, and $0 < \beta < 1$. Then, the $\beta-$VaR for $\boldsymbol{\theta}$ is defined to be the β-quantile of $\boldsymbol{\theta}$:

$$\beta\text{-VaR}(\boldsymbol{\theta}) := F_{\boldsymbol{\theta}}^{-1}(\beta).$$

Example 2 (*Conditional value at risk*) Let $\boldsymbol{\theta}$ be a random variable, and $0 < \beta < 1$. The following alternative characterization of β-CVaR [2] shows directly that it is a β-tail risk measure.

$$\beta\text{-CVaR}(\boldsymbol{\theta}) = \frac{1}{1-\beta}\int_{\beta}^{1} F_{\boldsymbol{\theta}}^{-1}(u)\, du.$$

Note that in the case that $\boldsymbol{\theta}$ is a continuous random variable, the β-CVaR is the conditional expectation of the random variable above its β-quantile (e.g. see [30]).

The observation that we exploit for this work is that very different random variables will have the same β-tail risk measure as long as their β-tails are the same.

When showing that two distributions have the same β-tails, it is convenient to use distribution functions rather than quantile functions. The following result gives conditions which ensure that the β-tails of two distributions are the same. We will make use of these in proofs later in this paper.

Lemma 1 *Suppose that $\boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}$ are random variables such that one of the two following conditions hold:*

(i) $F_{\tilde{\boldsymbol{\theta}}}(\theta) = F_{\boldsymbol{\theta}}(\theta)$ *for all* $\theta \geq F_{\boldsymbol{\theta}}^{-1}(\beta)$ *and* $F_{\tilde{\boldsymbol{\theta}}}(\theta) < \beta$ *for all* $\theta < F_{\boldsymbol{\theta}}^{-1}(\beta)$.
(ii) $F_{\tilde{\boldsymbol{\theta}}}(\theta) = F_{\boldsymbol{\theta}}(\theta)$ *for all* $\theta \geq L$ *for some* $L < F_{\boldsymbol{\theta}}^{-1}(\beta)$.

Then, $F_{\tilde{\boldsymbol{\theta}}}^{-1}(u) = F_{\boldsymbol{\theta}}^{-1}(u)$ *for all* $u \geq \beta$.

Proof We first prove that condition (i) implies that the β-tails are the same. Since $F_{\tilde{\boldsymbol{\theta}}}(\theta) = F_{\boldsymbol{\theta}}(\theta) \geq \beta$ for all $\theta \geq F_{\boldsymbol{\theta}}^{-1}(\beta)$, we must have $F_{\tilde{\boldsymbol{\theta}}}^{-1}(\beta) \leq F_{\boldsymbol{\theta}}^{-1}(\beta)$. Also, given $F_{\tilde{\boldsymbol{\theta}}}(\theta) < \beta$ for all $\theta < F_{\boldsymbol{\theta}}^{-1}(\beta)$ we must have $F_{\tilde{\boldsymbol{\theta}}}^{-1}(\beta) \geq F_{\boldsymbol{\theta}}^{-1}(\beta)$ and so $F_{\tilde{\boldsymbol{\theta}}}^{-1}(\beta) = F_{\boldsymbol{\theta}}^{-1}(\beta)$.

Now suppose $u \geq \beta$. Then,

$$\begin{aligned}
F_{\tilde{\boldsymbol{\theta}}}^{-1}(u) &= \inf\{\theta \in \mathbb{R} :\ F_{\tilde{\boldsymbol{\theta}}}(\theta) \geq u\} \\
&= \inf\{\theta \geq F_{\tilde{\boldsymbol{\theta}}}^{-1}(\beta) : F_{\tilde{\boldsymbol{\theta}}}(\theta) \geq u\} \\
&= \inf\{\theta \geq F_{\boldsymbol{\theta}}^{-1}(\beta) : F_{\boldsymbol{\theta}}(\theta) \geq u\} \\
&= \inf\{\theta \in \mathbb{R} :\ F_{\boldsymbol{\theta}}(\theta) \geq u\} \\
&= F_{\boldsymbol{\theta}}^{-1}(u)
\end{aligned}$$

where the second and fourth lines follow from the fact that quantile functions are non-decreasing.

In the case condition (ii) holds, we have for $L < \theta < F_{\boldsymbol{\theta}}^{-1}(\beta)$ that $F_{\tilde{\boldsymbol{\theta}}}(\theta) = F_{\boldsymbol{\theta}}(\theta) < \beta$, and since distribution functions are non-decreasing this means that $F_{\tilde{\boldsymbol{\theta}}}(\theta) < \beta$ for all $\theta < F_{\boldsymbol{\theta}}^{-1}(\beta)$. The result now follows by application of condition (i). □

3.2 Risk regions

In this paper we are primarily interested in problems of the following form:

$$\underset{x \in \mathcal{X}}{\text{minimize}}\ \rho_\beta(f(x, \boldsymbol{\xi})) \tag{1}$$

where $\mathcal{X} \subseteq \mathbb{R}^k$ is a deterministic set of feasible decisions, $\boldsymbol{\xi} \in \Xi \subseteq \mathbb{R}^d$ is a random vector defined on a probability space $(\Omega, \mathbb{P})$, the set Ξ is convex, $f : \mathcal{X} \times \Xi \to \mathbb{R}$ is a *loss function*, and ρ_β is a *tail risk measure*.

In order to solve these problems accurately, we need to be able to approximate well the tail risk measure of our the loss function $f(x, \boldsymbol{\xi})$ for all feasible decisions $x \in \mathcal{X}$.

To avoid repeated use of cumbersome notation we introduce the following shorthand for distribution and quantile functions:

$$\begin{aligned} F_x(\theta) &:= F_{f(x,\boldsymbol{\xi})}(\theta) = \mathbb{P}\left(f(x, \boldsymbol{\xi}) \leq \theta\right), \\ F_x^{-1}(\beta) &:= F_{f(x,\boldsymbol{\xi})}^{-1}(\beta) = \inf\{\theta \in \mathbb{R} :\ F_x(\theta) \geq \beta\}. \end{aligned}$$

In addition, since the loss function is only defined on Ξ, we frequently take complements of sets contained in Ξ. Again, to avoid repeated use of cumbersome notation, the standard notation for complements will apply with respect to Ξ. That is, for $\mathcal{R} \subseteq \Xi$ we write $\mathcal{R}^c$ in place of $\Xi \backslash \mathcal{R}$.

Since tail risk measures depend only on those outcomes which are in the β-tail, we aim to identify which outcomes lead to a loss in the β-tails for a feasible decision. This motivates the following definition.

Definition 4 (*Risk region*) For $0 < \beta < 1$ the β-risk region with respect to the decision $x \in \mathcal{X}$ is defined as follows:

$$\mathcal{R}_x(\beta) = \{\xi \in \Xi : F_x\left(f(x, \xi)\right) \geq \beta\},$$

or equivalently

$$\mathcal{R}_x(\beta) = \{\xi \in \Xi : f(x, \xi) \geq F_x^{-1}(\beta)\}. \tag{2}$$

The risk region with respect to the feasible region $\mathcal{X} \subset \mathbb{R}^k$ is defined to be:

$$\mathcal{R}_{\mathcal{X}}(\beta) = \bigcup_{x \in \mathcal{X}} \mathcal{R}_x(\beta). \tag{3}$$

The complement of this region is called the *non-risk region*. This can also be written

$$\mathcal{R}_{\mathcal{X}}(\beta)^c = \bigcap_{x \in \mathcal{X}} \mathcal{R}_x(\beta)^c. \tag{4}$$

The following basic properties of the risk region follow directly from the definition.

(i) $0 < \beta' < \beta < 1 \;\Rightarrow\; \mathcal{R}_{\mathcal{X}}(\beta) \subseteq \mathcal{R}_{\mathcal{X}}(\beta')$; (5)

(ii) $\mathcal{X}' \subset \mathcal{X} \;\Rightarrow \mathcal{R}_{\mathcal{X}'}(\beta) \subseteq \mathcal{R}_{\mathcal{X}}(\beta)$ for all $0 < \beta < 1$; (6)

(iii) If $\xi \mapsto f(x, \xi)$ is upper semi-continuous then $\mathcal{R}_x(\beta)$ is closed and $\mathcal{R}_x(\beta)^c$ is open. (7)

We now state a technical property and prove that this ensures the distribution of the random vector in a given region completely determines the value of a tail risk measure. In essence, this condition ensures that there is enough mass in the set to ensure that the β-quantile does not depend on the probability distribution outside of it.

Definition 5 (*Aggregation condition*) Suppose that $\mathcal{R}_{\mathcal{X}}(\beta) \subseteq \mathcal{R} \subset \Xi$ and that for all $x \in \mathcal{X}$, $\mathcal{R}$ satisfies the following condition:

$$\mathbb{P}\left(\boldsymbol{\xi} \in \{\xi \in \Xi : \theta' < f(x, \xi) < F_x^{-1}(\beta)\} \cap \mathcal{R}\right) > 0 \qquad \forall\, \theta' < F_x^{-1}(\beta)\,. \tag{8}$$

Then $\mathcal{R}$ is said to satisfy the *β-aggregation condition*.

The motivation for the term *aggregation condition* comes from Theorem 1 which follows. This result ensures that if a set satisfies the aggregation condition then we can transform the probability distribution of $\boldsymbol{\xi}$ so that all the mass in the complement of this set can be aggregated into a single point without affecting the value of the tail risk measure. This property is particularly relevant to scenario generation as if we have such a set, then all scenarios which it does not contain can be aggregated, reducing the size of the stochastic program. Note that the β-aggregation condition does not hold if $\boldsymbol{\xi}$ is a discrete random vector. However, in this case, the conclusion of the theorem holds without any extra conditions on $\mathcal{R}$.

Theorem 1 *Suppose that $\mathcal{R}_{\mathcal{X}}(\beta) \subseteq \mathcal{R} \subset \Xi$ and that $\tilde{\boldsymbol{\xi}}$ is a random vector for which*

$$\mathbb{P}(\boldsymbol{\xi} \in \mathcal{A}) = \mathbb{P}\left(\tilde{\boldsymbol{\xi}} \in \mathcal{A}\right) \qquad \text{for any measurable } \mathcal{A} \subseteq \mathcal{R}. \tag{9}$$

Then for any tail risk measure ρ_β we have $\rho_\beta(f(x, \boldsymbol{\xi})) = \rho_\beta\left(f(x, \tilde{\boldsymbol{\xi}})\right)$ for all $x \in \mathcal{X}$, if one of the following conditions hold:

(a) *$\mathcal{R}$ satisfies the β-aggregation condition,*
(b) *$\boldsymbol{\xi}$ is a discrete random vector.*

Proof Fix $x \in \mathcal{X}$. To show that $\rho_\beta\left(f(x,\boldsymbol{\xi})\right) = \rho_\beta\left(f(x,\tilde{\boldsymbol{\xi}})\right)$ we must show that the β-quantile and the β-tail distributions of $f(x,\boldsymbol{\xi})$ and $f(x,\tilde{\boldsymbol{\xi}})$ are the same. Using Lemma 1, the following two conditions are necessary and sufficient for this to occur:

$$F_x(\theta) = F_{f(x,\tilde{\boldsymbol{\xi}})}(\theta) \;\; \forall\, \theta \geq F_x^{-1}(\beta) \;\text{ and }\; F_{f(x,\tilde{\boldsymbol{\xi}})}(\theta) < \beta \;\; \forall\, \theta < F_x^{-1}(\beta)\,.$$

In the first case suppose that $\theta' \geq F_x^{-1}(\beta)$. Note that as a direct consequence of (9) we have

$$\mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{B}\right) = \mathbb{P}\left(\tilde{\boldsymbol{\xi}} \in \mathcal{B}\right) \qquad \text{for any } \mathcal{B} \supseteq \mathcal{R}^c. \tag{10}$$

Now,

$$\begin{aligned}
F_{f(x,\tilde{\boldsymbol{\xi}})}(\theta') &= \mathbb{P}\left(\tilde{\boldsymbol{\xi}} \in \{\xi \in \Xi :\; f(x,\xi) \leq \theta'\}\right) \\
&= \mathbb{P}\left(\tilde{\boldsymbol{\xi}} \in \underbrace{\mathcal{R}^c \cap \{\xi \in \Xi :\; f(x,\xi) \leq \theta'\}}_{=\mathcal{R}^c}\right) \\
&\quad + \mathbb{P}\left(\tilde{\boldsymbol{\xi}} \in \underbrace{\mathcal{R} \cap \{\xi \in \Xi :\; f(x,\xi) \leq \theta'\}}_{\subseteq \mathcal{R}}\right) \\
&= \mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{R}^c\right) + \mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{R} \cap \{\xi \in \Xi : f(x,\xi) \leq \theta'\}\right) \qquad \text{by (9) and (10)} \\
&= \mathbb{P}\left(\boldsymbol{\xi} \in \{\xi \in \Xi : f(x,\xi) \leq \theta'\}\right) = F_x(\theta') \qquad \text{as required.}
\end{aligned}$$

In the second case we suppose $\theta' < F_x^{-1}(\beta)$. We show that $F_{f(x,\tilde{\boldsymbol{\xi}})}(\theta') < \beta$ for each of the two conditions (a) and (b) separately. In the case where condition (a) holds, that is, when $\mathcal{R}$ satisfies the β-aggregation condition we have:

$$\begin{aligned}
F_{f(x,\tilde{\boldsymbol{\xi}})}(\theta') &= \mathbb{P}\left(\tilde{\boldsymbol{\xi}} \in \{\xi \in \Xi : f(x,\xi) \leq \theta'\}\right) \leq \mathbb{P}\left(\tilde{\boldsymbol{\xi}} \in \mathcal{R}^c \cup \{\xi \in \Xi : f(x,\xi) \leq \theta'\}\right) \\
&= \mathbb{P}\left(\tilde{\boldsymbol{\xi}} \in \underbrace{\{\xi \in \Xi : f(x,\xi) < F_x^{-1}(\beta)\}}_{\supseteq \mathcal{R}^c}\right) \\
&\quad - \mathbb{P}\left(\tilde{\boldsymbol{\xi}} \in \underbrace{\mathcal{R} \cap \{\xi \in \Xi : \theta' < f(x,\xi) < F_x^{-1}(\beta)\}}_{\subseteq \mathcal{R}}\right) \\
&= \mathbb{P}\left(\boldsymbol{\xi} \in \{\xi \in \Xi : f(x,\xi) < F_x^{-1}(\beta)\}\right) \\
&\quad - \mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{R} \cap \{\xi \in \Xi : \theta' < f(x,\xi) < F_x^{-1}(\beta)\}\right) \qquad \text{by (9) and (10)} \\
&< \mathbb{P}\left(\boldsymbol{\xi} \in \{\xi \in \Xi : f(x,\xi) < F_x^{-1}(\beta)\}\right) \qquad \text{by (8)} \\
&\leq \beta
\end{aligned}$$

as required. In the case condition (b) holds, that is when $\boldsymbol{\xi}$ is discrete, we have:

$$\begin{aligned} F_{f(x,\tilde{\boldsymbol{\xi}})}(\theta') &\leq \mathbb{P}\left(f(x,\tilde{\boldsymbol{\xi}}) < F_x^{-1}(\beta)\right) \\ &= \mathbb{P}\left(f(x,\boldsymbol{\xi}) < F_x^{-1}(\beta)\right) \\ &< \beta \qquad \text{since } \boldsymbol{\xi} \text{ is discrete} \end{aligned}$$

as required. □

It is difficult to verify that a set $\mathcal{R} \supseteq \mathcal{R}_{\mathcal{X}}(\beta)$ satisfies the β-aggregation condition by directly checking that the condition (8) holds. The following proposition gives conditions under which it holds immediately for $\mathcal{R}_{\mathcal{X}}(\beta')$ when $\beta' < \beta$.

Proposition 1 *Suppose $\beta' < \beta$ and F_x is continuous at $F_x^{-1}(\beta)$ for all $x \in \mathcal{X}$. Then, $\mathcal{R}_{\mathcal{X}}(\beta')$ satisfies the β-aggregation condition. That is, for all $x \in \mathcal{X}$*

$$\mathbb{P}\left(\boldsymbol{\xi} \in \{\xi \in \Xi : \theta' < f(x,\xi) < F_x^{-1}(\beta)\} \cap \mathcal{R}_{\mathcal{X}}\left(\beta'\right)\right) > 0 \qquad \forall\, \theta' < F_x^{-1}(\beta).$$

Proof Fix $x \in \mathcal{X}$. Since F_x is continuous at $F_x^{-1}(\beta)$ we must have that $F_x^{-1}\left(\beta'\right) < F_x^{-1}(\beta)$. Now, for all $F_x^{-1}\left(\beta'\right) < \theta' < F_x^{-1}(\beta)$, we have $\{\xi \in \Xi : \theta' < f(x,\xi) < F_x^{-1}(\beta)\} \subset \mathcal{R}_{\mathcal{X}}(\beta')$ and so

$$\begin{aligned} &\mathbb{P}\left(\boldsymbol{\xi} \in \{\xi \in \Xi : \theta' < f(x,\xi) < F_x^{-1}(\beta)\} \cap \mathcal{R}_{\mathcal{X}}\left(\beta'\right)\right) \\ &\quad = \mathbb{P}\left(\theta' < f(x,\boldsymbol{\xi}) < F_x^{-1}(\beta)\right) > 0. \end{aligned}$$

□

For convenience, we now drop β from our notation and terminology. Thus, we refer to the β-risk region and β-aggregation condition as simply the risk region and aggregation condition respectively, and write $\mathcal{R}_{\mathcal{X}}(\beta)$ as $\mathcal{R}_{\mathcal{X}}$.

All sets satisfying the aggregation condition must contain the risk region, however, the aggregation condition does not necessarily hold for the risk region itself.

We must impose extra conditions on the problem to avoid some degenerate cases where the aggregation condition and the conclusion of Theorem 1 do not hold. The following example demonstrates such a degenerate case.

Example 3 Let $\mathcal{X} = \mathbb{R}^+\backslash\{0\}$, $\Xi = [0,1]$, $\boldsymbol{\xi} \sim \text{Uniform}(0,1)$ and $f : (x,\xi) \mapsto x\xi$. Then $\mathcal{R}_x = [\beta, 1]$ for all $x \in \mathcal{X}$, and so $\mathcal{R}_{\mathcal{X}} = [\beta, 1]$. Now, consider the random variable $\phi(\boldsymbol{\xi})$ where $\phi : \mathbb{R} \to \mathbb{R}$ is defined as follows:

$$\phi(\xi) = \begin{cases} \xi & \text{if } \xi \geq \beta, \\ 0 & \text{othewise.} \end{cases}$$

Since $\phi(\boldsymbol{\xi}) = \boldsymbol{\xi}$ for all $\boldsymbol{\xi} \in \mathcal{R}_{\mathcal{X}}$ we have $\mathbb{P}\left(\phi(\boldsymbol{\xi}) \in A\right) = \mathbb{P}\left(\boldsymbol{\xi} \in A\right)$ for all $A \subseteq \mathcal{R}_{\mathcal{X}}$. On the other hand, we have that $F^{-1}_{f(x,\phi(\boldsymbol{\xi}))}(\beta) = 0 < \beta = F^{-1}_{f(x,\boldsymbol{\xi})}(\beta)$.

The following result provides extra conditions for continuous distributions which ensure that the aggregation condition holds for the risk region $\mathcal{R}_{\mathcal{X}}$.

Proposition 2 *Suppose that $\boldsymbol{\xi}$ is a continuous random vector whose support coincides with Ξ, and that the following conditions hold:*

(i) $\xi \mapsto f(x,\xi)$ *is continuous for all* $x \in \mathcal{X}$,
(ii) *For each* $x \in \mathcal{X}$ *there exists* $x' \in \mathcal{X}$ *such that*

$$\text{int}\,(\Xi) \cap \text{int}\,(\mathcal{R}_x \cap \mathcal{R}_{x'}) \neq \emptyset \text{ and } \text{int}\,(\Xi) \cap \text{int}\,(\mathcal{R}_{x'} \backslash \mathcal{R}_x) \neq \emptyset, \tag{11}$$

(iii) $\text{int}\,(\Xi) \cap \text{int}\,(\mathcal{R}_{\mathcal{X}})$ *is connected.*

Then the risk region $\mathcal{R}_{\mathcal{X}}$ satisfies the aggregation condition.

Proof Fix $x \in \mathcal{X}$ and $\theta' < F_x^{-1}(\beta)$. Pick $x' \in \mathcal{X}$ such that (11) holds. Also, let $\xi_0 \in \text{int}\,(\Xi) \cap \text{int}\,(\mathcal{R}_{x'} \backslash \mathcal{R}_x)$ and $\xi_1 \in \text{int}\,(\Xi) \cap \text{int}\,(\mathcal{R}_x \cap \mathcal{R}_{x'})$. Since $\text{int}\,(\Xi) \cap \text{int}\,(\mathcal{R}_{\mathcal{X}})$ is connected there exists a continuous path from ξ_0 to ξ_1. That is, there exists a continuous function $\gamma : [0,1] \to \text{int}\,(\Xi) \cap \text{int}\,(\mathcal{R}_{\mathcal{X}})$ such that $\gamma(0) = \xi_0$ and $\gamma(1) = \xi_1$. Now, $f(x,\xi_0) < F_x^{-1}(\beta)$ and $f(x,\xi_1) \geq F_x^{-1}(\beta)$ and so given that $t \mapsto f(x,\gamma(t))$ is continuous there must exist $0 < t < 1$ such that $\theta' < f(x,\gamma(t)) < F_x^{-1}(\beta)$. That is,

$$\text{int}\,(\Xi) \cap \text{int}\,(\mathcal{R}_{\mathcal{X}}) \cap \{\xi \in \Xi : \theta' < f(x,\xi) < F_x^{-1}(\beta)\} \neq \emptyset.$$

This is a non-empty open set contained in the support of $\boldsymbol{\xi}$ and so has positive probability, hence the aggregation condition holds for $\mathcal{R}_{\mathcal{X}}$. □

The following proposition gives a condition under which the non-risk region is convex.

Proposition 3 *Suppose that for each $x \in \mathcal{X}$ the function $\xi \mapsto f(x,\xi)$ is convex. Then, the non-risk region $\mathcal{R}_{\mathcal{X}}^c$ is convex.*

Proof For $x \in \mathcal{X}$, if $\xi \mapsto f(x,\xi)$ is convex then the set $\mathcal{R}_x^c = \{\xi \in \Xi : f(x,\xi) < F_x^{-1}(\beta)\}$ must be convex. The intersection of convex sets is convex, hence $\mathcal{R}_{\mathcal{X}}^c = \bigcap_{x\in\mathcal{X}} \mathcal{R}_x^c$ is convex. □

This convexity condition is held by a large class of stochastic programs. Two-stage stochastic linear programs have loss functions of the following general form:

$$Q(x,\boldsymbol{\xi}) = \min_y \{\boldsymbol{q}^T y | \boldsymbol{W} y = \boldsymbol{h} - \boldsymbol{T} x,\ y \geq 0\}$$

where $\boldsymbol{q}, y \in \mathbb{R}^r$, $\boldsymbol{h} \in \mathbb{R}^t$, $\boldsymbol{W} \in \mathbb{R}^{t\times r}$ and $\boldsymbol{T} \in \mathbb{R}^{t\times k}$, and $\boldsymbol{\xi}$ is the concatenation of all the stochastic components of the problem; that is, $\boldsymbol{\xi}^T = \left(\boldsymbol{q}^T, \boldsymbol{h}^T, \boldsymbol{T}_1, \ldots, \boldsymbol{T}_t, \boldsymbol{W}_1, \ldots, \boldsymbol{W}_t\right)$ where $\boldsymbol{T}_i$ and $\boldsymbol{W}_i$ denote the i-th rows of the matrices $\boldsymbol{T}$ and $\boldsymbol{W}$ respectively. Standard results in stochastic programming guarantee that $\xi \mapsto Q(x,\xi)$ is convex if the only random components of the problem are $\boldsymbol{h}$ and $\boldsymbol{T}$, that is if $\xi^T = (h^T, T_1, \ldots, T_t)$. See for instance [7, Chapter 3, Theorem 2].

The random vector in the following definition plays a special role in our theory.

Definition 6 (*Aggregated random vector*) For some set $\mathcal{R}_{\mathcal{X}} \subseteq \mathcal{R} \subset \Xi$ the *aggregated random vector* is defined as follows:

$$\psi_{\mathcal{R}}(\boldsymbol{\xi}) := \begin{cases} \boldsymbol{\xi} & \text{if } \boldsymbol{\xi} \in \mathcal{R}, \\ \mathbb{E}\left[\, \boldsymbol{\xi} | \boldsymbol{\xi} \in \mathcal{R}^c \,\right] & \text{otherwise.} \end{cases}$$

If $\mathcal{R}$ satisfies the aggregation condition and $\mathbb{E}\left[\boldsymbol{\xi} | \boldsymbol{\xi} \in \mathcal{R}^c\right] \in \mathcal{R}^c_{\mathcal{X}}$ then Theorem 1 guarantees that $\rho_\beta\left(f\left(x, \psi_{\mathcal{R}}(\boldsymbol{\xi})\right)\right) = \rho_\beta\left(f\left(x, \boldsymbol{\xi}\right)\right)$ for all $x \in \mathcal{X}$. The latter condition holds, for example, if $\xi \mapsto f(x, \xi)$ is convex for all $x \in \mathcal{X}$, since by Proposition 3 we have that $\mathcal{R}^c_{\mathcal{X}}$ is convex and also $\mathcal{R}^c \subseteq \mathcal{R}^c_{\mathcal{X}}$. Under these conditions, as well as preserving the value of the tail risk measure, the function $\psi_{\mathcal{R}}$ will also preserve the expectation for affine loss functions.

Corollary 1 *Suppose for each $x \in \mathcal{X}$ the function $\xi \mapsto f(x, \xi)$ is affine and for a set $\mathcal{R} \subset \Xi$ satisfying the aggregation condition we have that $\mathbb{E}\left[\boldsymbol{\xi} | \boldsymbol{\xi} \in \mathcal{R}^c\right] \in \mathcal{R}^c$. Then,*

$$\begin{aligned} \rho_\beta\left(f\left(x, \psi_{\mathcal{R}}(\boldsymbol{\xi})\right)\right) = \rho_\beta\left(f\left(x, \boldsymbol{\xi}\right)\right) \text{ and } & \mathbb{E}\left[f\left(x, \psi_{\mathcal{R}}\left(\boldsymbol{\xi}\right)\right)\right] \\ = & \mathbb{E}\left[f(x, \boldsymbol{\xi})\right] \text{ for all } x \in \mathcal{X}. \end{aligned}$$

Proof The equality of the tail-risk measures follows immediately from Theorem 1. For the expectation function we have

$$\begin{aligned} \mathbb{E}\left[\psi_{\mathcal{R}}(\boldsymbol{\xi})\right] &= \mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{R}\right) \mathbb{E}\left[\psi_{\mathcal{R}}(\boldsymbol{\xi}) | \boldsymbol{\xi} \in \mathcal{R}\right] + \mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{R}^c\right) \mathbb{E}\left[\psi_{\mathcal{R}}(\boldsymbol{\xi}) | \boldsymbol{\xi} \in \mathcal{R}^c\right] \\ &= \mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{R}\right) \mathbb{E}\left[\boldsymbol{\xi} | \boldsymbol{\xi} \in \mathcal{R}\right] + \mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{R}^c\right) \mathbb{E}\left[\boldsymbol{\xi} | \boldsymbol{\xi} \in \mathcal{R}^c\right] = \mathbb{E}\left[\boldsymbol{\xi}\right]. \end{aligned}$$

Since $\xi \mapsto f(x, \xi)$ is affine this means that

$$\mathbb{E}\left[f\left(x, \psi_{\mathcal{R}}(\boldsymbol{\xi})\right)\right] = f\left(x, \mathbb{E}\left[\psi_{\mathcal{R}}(\boldsymbol{\xi})\right]\right) = f\left(x, \mathbb{E}\left[\boldsymbol{\xi}\right]\right) = \mathbb{E}\left[f(x, \boldsymbol{\xi})\right].$$

□

4 Scenario generation

In the previous section, we showed that under mild conditions the value of a tail risk measure only depends on the distribution of outcomes in the risk region. In this section we demonstrate how this feature may be exploited for the purposes of scenario generation.

We assume throughout this section that our scenario sets are constructed from some underlying probabilistic model from which we can draw independent identically distributed samples. We also assume we have a set $\mathcal{R}_{\mathcal{X}} \subseteq \mathcal{R} \subset \Xi$ which satisfies the aggregation condition for the problem under consideration, and for which we can easily test membership. The set $\mathcal{R}$ may be an *exact risk region*, that is $\mathcal{R} = \mathcal{R}_{\mathcal{X}}$, or it could a *conservative risk region*, that is $\mathcal{R} \supset \mathcal{R}_{\mathcal{X}}$. To avoid repeating cumbersome terminology, we simply refer to $\mathcal{R}$ as a risk region, differentiating between the conservative and exact cases only where necessary. The complement $\mathcal{R}^c$ will be referred to

as the *aggregation region* for reasons which will become clear. Our general approach to scenario generation is to prioritize the construction of scenarios in the risk region $\mathcal{R}$.

In Sect. 4.1 we present and analyse a scenario generation method which we call aggregation sampling. In Sect. 4.2 we briefly discuss alternative ways of exploiting risk regions for scenario generation.

4.1 Aggregation sampling

In *aggregation sampling* the user specifies a number of *risk scenarios*, that is, the number of scenarios to represent the risk region. The algorithm then draws samples from the distribution, storing those samples which lie in the risk region and aggregating those in the aggregation region into a single point. In particular, the samples in the aggregation region are aggregated into their mean. The algorithm terminates when the specified number of risk scenarios has been reached. This is detailed in Algorithm 1.

input : $\mathcal{R} \subset \Xi$ set satisfying aggregation condition, $N_{\mathcal{R}}$ number of required risk scenarios
output: $\{(\xi_s, p_s)\}_{s=1}^{N_{\mathcal{R}}+1}$ scenario set
$n_{\mathcal{R}^c} \leftarrow 0, n_{\mathcal{R}} \leftarrow 0, \xi_{\mathcal{R}^c} = \mathbf{0}$;
while $n_{\mathcal{R}} < N_{\mathcal{R}}$ **do**
  Sample new point ξ;
  if $\xi \in \mathcal{R}$ **then**
    $\xi_{n_{\mathcal{R}}+1} \leftarrow \xi$; $n_{\mathcal{R}} \leftarrow n_{\mathcal{R}} + 1$;
  else
    $\xi_{\mathcal{R}^c} \leftarrow \frac{1}{n_{\mathcal{R}^c}+1}\left(n_{\mathcal{R}^c}\xi_{\mathcal{R}^c} + \xi\right)$; $n_{\mathcal{R}^c} \leftarrow n_{\mathcal{R}^c} + 1$
if $n_{\mathcal{R}^c} > 0$ **then**
  $\xi_{N_{\mathcal{R}}+1} \leftarrow \xi_{\mathcal{R}^c}$;
else
  Sample new point ξ;
  $n_{\mathcal{R}^c} \leftarrow 1$; $\xi_{N_{\mathcal{R}}+1} \leftarrow \xi$;
foreach i *in* $1, \ldots, N_{\mathcal{R}}$ **do** $p_i \leftarrow \frac{1}{n_{\mathcal{R}^c}+N_{\mathcal{R}}}$ $p_{N_{\mathcal{R}}+1} \leftarrow \frac{n_{\mathcal{R}^c}}{n_{\mathcal{R}^c}+N_{\mathcal{R}}}$

Algorithm 1: Aggregation sampling

Aggregation sampling can be thought of as equivalent to sampling from the aggregated random vector from Definition 6 for large sample sizes. Aggregation sampling is thus consistent with standard Monte Carlo sampling only if $\mathcal{R}$ satisfies the aggregation condition and $\mathbb{E}\left[\boldsymbol{\xi} | \boldsymbol{\xi} \in \mathcal{R}^c\right] \in \mathcal{R}^c$. In Sect. 5, we provide conditions under which we can prove consistency. Note that it is possible that the algorithm could terminate without sampling any scenario in the aggregation region. This could happen in cases where $\mathbb{P}(\boldsymbol{\xi} \in \mathcal{R}^c)$ is very small, and the number of specified risk scenarios n is relatively small. In this case, to ensure that the algorithm terminates in a reasonable amount of time and that the scenario set which the algorithm outputs always has a consistent number of scenarios, we sample an arbitrary scenario in place of a scenario representing the aggregated scenarios. This situation is irrelevant for the asymptotic analysis of the algorithm.

We now study the performance of our aggregation sampling algorithm. Let $a = \mathbb{P}(\boldsymbol{\xi} \in \mathcal{R}^c)$ be the probability of the aggregation region, and n the desired number of risk scenarios. Let $N(n)$ denote the *effective sample size* for aggregation sampling, that is, the number of samples drawn until the algorithm terminates.[1] The aggregation sampling algorithm can be viewed as a sequence of Bernoulli trials where a trial is a success if the corresponding sample lies in the aggregation region, and which terminates once we have reached n failures, that is, once we have sampled n scenarios from the risk region. We can therefore write down the distribution of $N(n)$:

$$N(n) \sim n + \mathcal{NB}(n, a),$$

where $\mathcal{NB}(n, a)$ denotes a *negative binomial* random variable whose probability mass function is as follows:

$$\binom{k+n-1}{k}(1-a)^n a^k, \qquad k \geq 0.$$

The expected effective sample size of aggregation sampling is thus:

$$\mathbb{E}[N(n)] = n + n\frac{a}{1-a}. \tag{12}$$

The expected effective sample size $N(n)$ can be thought of as the required sample size to construct a scenario set via Monte Carlo sampling with n scenarios in the risk region $\mathcal{R}$. Thus, the greater the expected effective sample size, the greater the benefit of using aggregation sampling over standard Monte Carlo sampling. From (12) we can see that the expected effective sample size increases as the probability a of the aggregation region increases. Therefore, when constructing a risk region $\mathcal{R} \supseteq \mathcal{R}_{\mathcal{X}}$ for the purposes of scenario generation, it is important that $\mathcal{R}$ is as tight an approximation of the exact risk region $\mathcal{R}_{\mathcal{X}}$ as possible in order that $a = \mathbb{P}(\boldsymbol{\xi} \in \mathcal{R}^c)$ is as large as possible. Also, the fact that the advantage of using aggregation sampling over standard Monte Carlo sample improves as the probability of the risk region increases, also tells us that this methodology will potentially work better for problems with higher values of β and which are more constrained due to the relations (5) and (6).

4.2 Alternative approaches

Aggregation reduction In aggregation reduction one draws a fixed number of samples n from the distribution and then aggregates all those in the aggregation region. As opposed to aggregation sampling, this method uses a fixed number of samples, but constructs a scenario set with a random number of scenarios. Let $R(n)$ denote the number of scenarios which are aggregated in the aggregation reduction method. Aggregation reduction can similarly be viewed as a sequence of n Bernoulli trials, where success

[1] For simplicity of exposition we discount the event that the while loop of the algorithm terminates with $n_{\mathcal{R}^c} = 0$ which occurs with probability $(1-a)^n$.

and failure are defined in the same way as described above. The number of aggregated scenarios in aggregation reduction is therefore distributed as follows:

$$R(n) \sim \mathcal{B}(n, a)$$

where $\mathcal{B}(n, a)$ denotes a binomial random variable and so we have

$$\mathbb{E}[R(n)] = na. \tag{13}$$

Again, the performance of this method, in terms of the expected number of aggregated scenarios, can be seen to improve as the probability of the aggregation region increases.
Alternative sampling methods The above algorithms and analyses assume that the samples of $\boldsymbol{\xi}$ were identically, independently distributed. However, in principle the algorithms will work for any unbiased sequence of samples. This opens up the possibility of enhancing the scenario aggregation and reduction algorithms by using them in conjunction with variance reduction techniques such as importance sampling, or antithetic sampling [20].[2] The formulae (12) and (13) will still hold, but a will be the probability of a *sample* occuring in the aggregation region rather than the actual probability of the aggregation region itself.
Alternative representations of the aggregation region The above algorithms can also be generalized in how they represent the non-risk region. Because aggregation sampling and aggregating reduction only represent the non-risk region with a single scenario, they do not in general preserve the overall expectation of the loss function, or any other statistics of the loss function except for the value of a tail risk measure. These algorithms should therefore generally only be used for problems which only involve tail risk measures. However, if the loss function is affine (in the sense of Corollary 1), then collapsing all points in the non-risk region to the conditional expectation preserves the overall expectation.

If expectation or any other statistic of the cost function is used in the optimization problem then one could represent the non-risk region region with many scenarios. For example, instead of aggregating all scenarios in the non-risk region into a single point we could apply a clustering algorithm to them such as k-means. The ideal allocation of points between the risk and non-risk regions will be problem dependent and is beyond the scope of this paper.

5 Consistency of aggregation sampling

The reason that aggregation sampling and aggregation reduction work is that, for large sample sizes, they are equivalent to sampling from the aggregated random vector, and if the aggregation condition holds then the aggregated random vector yields the same optimization problem as the original random vector. We only prove consistency for aggregation sampling and not aggregation reduction as the proofs are very similar.

[2] Batch sampling methods such as stratified sampling will not work with aggregation sampling which requires samples to be drawn sequentially.

Essentially, the only difference is that aggregation sampling has the additional complication of terminating after a random number of samples.

We suppose in this section that we have a sequence of independently identically distributed (i.i.d.) random vectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots$ with the same distribution as $\boldsymbol{\xi}$, and which are defined on the product probability space Ω^∞.

5.1 Uniform convergence of empirical β-quantiles

The i.i.d. sequence of random vectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots$ can be used to estimate the distribution and quantile functions of $\boldsymbol{\xi}$. We introduce the additional short-hand for the empirical distribution and quantile functions:

$$F_{n,x}(\theta) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\xi \in \Xi : f(x,\xi) \leq \theta\}}(\boldsymbol{\xi}_i) \quad \text{and} \quad F^{-1}_{n,x}(u) := \inf\{\theta \in \mathbb{R} : \ F_{n,x}(\theta) \geq u\}.$$

Note that these are random-valued functions on the probability space Ω^∞. It is immediate from the strong law of large numbers that for all $\bar{x} \in \mathcal{X}$ and $\theta \in \mathbb{R}$, we have $F_{n,\bar{x}}(\theta) \overset{\text{w.p.1}}{\to} F_{\bar{x}}(\theta)$ as $n \to \infty$. In addition, if $F_{\bar{x}}$ is strictly increasing at $\theta = F^{-1}_{\bar{x}}(\beta)$, that is for all $\epsilon > 0$

$$F_{\bar{x}}\left(F^{-1}_{\bar{x}}(\beta) - \epsilon\right) < \beta < F_{\bar{x}}\left(F^{-1}_{\bar{x}}(\beta) + \epsilon\right).$$

then we also have $F^{-1}_{n,\bar{x}}(\beta) \overset{\text{w.p.1}}{\to} F^{-1}_{\bar{x}}(\beta)$ as $n \to \infty$; see for instance [32, Chapter 2]. The following result extends this pointwise convergence to a convergence which is uniform with respect to $x \in \mathcal{X}$.

Theorem 2 *Suppose the following hold:*

(i) *For each $x \in \mathcal{X}$, F_x is strictly increasing and continuous at $F^{-1}_x(\beta)$,*
(ii) *For all $\bar{x} \in \mathcal{X}$ with probability 1 the mapping $x \mapsto f(x, \boldsymbol{\xi})$ is continuous at $\bar{x}$,*
(iii) *$\mathcal{X} \subset \mathbb{R}^k$ is compact.*

Then, with probability 1

$$\lim_{n\to\infty} \sup_{x \in \mathcal{X}} \left| F^{-1}_{n,x}(\beta) - F^{-1}_x(\beta) \right| = 0.$$

The proof of this result relies on various continuity properties of the distribution and quantile functions which are provided in "Appendix A". Some elements of the proof below have been adapted from [34, Theorem 7.48], a result which concerns the uniform convergence of expectation functions.

Proof Fix $\epsilon_0 > 0$ and $\bar{x} \in \mathcal{X}$. Since $F_{\bar{x}}$ is right-continuous with left limits, it has only countably many discontinuities, and so there exists $0 < \epsilon < \epsilon_0$ such that $F_{\bar{x}}$ is continuous at $F^{-1}_{\bar{x}}(\beta) \pm \epsilon$. Since $F_{\bar{x}}$ is strictly increasing at $F^{-1}_{\bar{x}}(\beta)$,

$$\delta := \min\left\{\beta - F_{\bar{x}}\left(F^{-1}_{\bar{x}}(\beta) - \epsilon\right),\ F_{\bar{x}}\left(F^{-1}_{\bar{x}}(\beta) + \epsilon\right) - \beta\right\} > 0. \tag{14}$$

By Corollary 5 in "Appendix A" the mapping $x \mapsto F_x\left(F_{\bar{x}}^{-1}(\beta)-\epsilon\right)$ is continuous at $\bar{x}$. Applying Lemma 3 in "Appendix A", there exists a neighborhood W of $\bar{x}$ such that with probability 1

$$\limsup_{n\to\infty}\sup_{x\in W\cap\mathcal{X}}\left|F_{n,x}(F_{\bar{x}}^{-1}(\beta)-\epsilon)-F_{n,\bar{x}}(F_{\bar{x}}^{-1}(\beta)-\epsilon)\right|<\delta.$$

In addition, by the strong law of large numbers, with probability 1

$$\lim_{n\to\infty}\left|F_{n,\bar{x}}\left(F_{\bar{x}}^{-1}(\beta)-\epsilon\right)-F_{\bar{x}}\left(F_{\bar{x}}^{-1}(\beta)-\epsilon\right)\right|=0. \tag{15}$$

Note that for all $x\in W\cap\mathcal{X}$

$$\begin{aligned}&\left|F_{n,x}\left(F_{\bar{x}}^{-1}(\beta)-\epsilon\right)-F_{\bar{x}}\left(F_{\bar{x}}^{-1}(\beta)-\epsilon\right)\right|\\&\quad\le\left|F_{n,x}(F_{\bar{x}}^{-1}(\beta)-\epsilon)-F_{n,\bar{x}}(F_{\bar{x}}^{-1}(\beta)-\epsilon)\right|\\&\qquad+\left|F_{n,\bar{x}}\left(F_{\bar{x}}^{-1}(\beta)-\epsilon\right)-F_{\bar{x}}\left(F_{\bar{x}}^{-1}(\beta)-\epsilon\right)\right|.\end{aligned}$$

Thus, with probability 1

$$\begin{aligned}&\limsup_{n\to\infty}\sup_{x\in W\cap\mathcal{X}}\left|F_{n,x}\left(F_{\bar{x}}^{-1}(\beta)-\epsilon\right)-F_{\bar{x}}\left(F_{\bar{x}}^{-1}(\beta)-\epsilon\right)\right|\\&\le\limsup_{n\to\infty}\sup_{x\in W\cap\mathcal{X}}\left|F_{n,x}(F_{\bar{x}}^{-1}(\beta)-\epsilon)-F_{n,\bar{x}}(F_{\bar{x}}^{-1}(\beta)-\epsilon)\right|\\&\quad+\limsup_{n\to\infty}\left|F_{n,\bar{x}}\left(F_{\bar{x}}^{-1}(\beta)-\epsilon\right)-F_{\bar{x}}\left(F_{\bar{x}}^{-1}(\beta)-\epsilon\right)\right|\\&<\delta+0=\delta.\end{aligned} \tag{16}$$

Similarly, we can choose W such that with probability 1

$$\limsup_{n\to\infty}\sup_{x\in W\cap\mathcal{X}}\left|F_{n,x}\left(F_{\bar{x}}^{-1}(\beta)+\epsilon\right)-F_{\bar{x}}\left(F_{\bar{x}}^{-1}(\beta)+\epsilon\right)\right|<\delta. \tag{17}$$

Using (14), (16) and (17) we can conclude that for all $x\in W\cap\mathcal{X}$ with probability 1

$$\limsup_{n\to\infty}F_{n,x}\left(F_{\bar{x}}^{-1}(\beta)-\epsilon\right)<\beta<\liminf_{n\to\infty}F_{n,x}\left(F_{\bar{x}}^{-1}(\beta)+\epsilon\right).$$

Hence, we have that for all $x\in W\cap\mathcal{X}$, with probability 1, there exists N such that for all $n>N$

$$F_{\bar{x}}^{-1}(\beta)-\epsilon<F_{n,x}^{-1}(\beta)\le F_{\bar{x}}^{-1}(\beta)+\epsilon,$$

and so we can conclude that

$$\limsup_{n\to\infty} \sup_{x\in W\cap\mathcal{X}} \left|F_{n,x}^{-1}(\beta) - F_{\tilde{x}}^{-1}(\beta)\right| \leq \epsilon < \epsilon_0. \tag{18}$$

Also, by Proposition 6 in "Appendix A" the function $x \mapsto F_x^{-1}(\beta)$ is continuous and so the neighborhood W can also be chosen so that

$$\sup_{x\in W\cap\mathcal{X}} \left|F_{\tilde{x}}^{-1}(\beta) - F_x^{-1}(\beta)\right| < \epsilon_0, \tag{19}$$

and so combining (18) and (19) we have that with probability 1

$$\limsup_{n\to\infty} \sup_{x\in W\cap\mathcal{X}} \left|F_{n,x}^{-1}(\beta) - F_x^{-1}(\beta)\right| < 2\epsilon_0.$$

Finally, since $\mathcal{X}$ is compact, there exists a finite number of points $x_1, \ldots, x_m \in \mathcal{X}$ with corresponding neighborhoods $W_1, \ldots, W_m$ covering $\mathcal{X}$, such that with probability 1, the following holds:

$$\limsup_{n\to\infty} \sup_{x\in W_j\cap\mathcal{X}} \left|F_{n,x}^{-1}(\beta) - F_x^{-1}(\beta)\right| < 2\epsilon_0 \qquad \text{for } i = 1, \ldots, m$$

that is, with probability 1,

$$\limsup_{n\to\infty} \sup_{x\in\mathcal{X}} \left|F_{n,x}^{-1}(\beta) - F_x^{-1}(\beta)\right| < 2\epsilon_0.$$

Since the choice of ϵ_0 was arbitrary the result follows. □

To facilitate the statement and proofs of the following results we introduce the following index sets which keep track of the indices of the samples which are in the risk and aggregation regions.

$$\mathcal{I}_{\mathcal{R}}(n) = \{1 \leq j \leq n : \boldsymbol{\xi}_j \in \mathcal{R}\},$$
$$\mathcal{I}_{\mathcal{R}^c}(n) = \{1 \leq j \leq n : \boldsymbol{\xi}_j \in \mathcal{R}^c\}.$$

The following corollary shows that we have uniform convergence of the β-quantiles when sampling from the aggregated random vector $\psi_{\mathcal{R}}(\boldsymbol{\xi})$. In order to state and prove this result, we introduce the following additional notation for the distribution and quantile functions for $f(x, \psi_{\mathcal{R}}(\boldsymbol{\xi}))$, and their empirical counterparts for the sample $\psi_{\mathcal{R}}(\boldsymbol{\xi}_1), \psi_{\mathcal{R}}(\boldsymbol{\xi}_2), \ldots$:

$$\tilde{F}_x(\theta) = \mathbb{P}\left(f(x, \psi_{\mathcal{R}}(\boldsymbol{\xi})) \leq \theta\right)$$
$$\tilde{F}_x^{-1}(u) = \inf\{\theta \in \mathbb{R} : \tilde{F}_x(\theta) \geq u\}$$

$$\tilde{F}_{n,x}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\xi\in\Xi:\ f(x,\xi)\le\theta\}}\left(\psi_{\mathcal{R}}(\boldsymbol{\xi}_i)\right)$$
$$= \frac{|\mathcal{I}_{\mathcal{R}^c}(n)|}{n}\mathbb{1}_{\{\xi\in\Xi:\ f(x,\xi)\le\theta\}}\left(\mathbb{E}\left[\boldsymbol{\xi}|\boldsymbol{\xi}\in\mathcal{R}^c\right]\right)+\frac{1}{n}\sum_{i\in\mathcal{I}_{\mathcal{R}}(n)}\mathbb{1}_{\{\xi\in\Xi:\ f(x,\xi)\le\theta\}}(\boldsymbol{\xi}_i)$$

$$\tilde{F}^{-1}_{n,x}(u) = \inf\{\theta\in\mathbb{R}:\tilde{F}_{n,x}(\theta)\ge u\}$$

Like $F_{n,x}$ and $F^{-1}_{n,x}$, the final two functions are random-valued functions on the probability space Ω^∞.

Corollary 2 *Let $\mathcal{R}_\mathcal{X}\subseteq\mathcal{R}\subset\mathbb{R}^d$ be a set satisfying the aggregation condition, and suppose that conditions (i)–(iii) from Theorem 2 hold and in addition:*

(iv) $\mathbb{E}\left[\boldsymbol{\xi}|\boldsymbol{\xi}\in\mathcal{R}^c\right]\in\text{int}\left(\mathcal{R}^c_\mathcal{X}\right)$.
(v) *The mapping $x\mapsto f\left(x,\mathbb{E}\left[\boldsymbol{\xi}|\boldsymbol{\xi}\in\mathcal{R}^c\right]\right)$ is continuous.*

Then with probability 1

$$\lim_{n\to\infty}\sup_{x\in\mathcal{X}}|\tilde{F}^{-1}_{n,x}(\beta)-F^{-1}_x(\beta)| = 0.$$

Proof Since $\mathcal{R}$ satisfies the aggregation condition, and condition (a) holds, by Theorem 1, we have that $\tilde{F}^{-1}_x(\beta)=F^{-1}_x(\beta)$ for all $x\in\mathcal{X}$. Therefore, to prove this result, we will apply Theorem 2 to $f(x,\psi_\mathcal{R}(\boldsymbol{\xi}))$ and so must show that conditions (i)–(iii) from Theorem 2 also hold for $f(x,\psi_\mathcal{R}(\boldsymbol{\xi}))$. Condition (iii) holds immediately, and condition (ii) holds for $f(x,\psi_\mathcal{R}(\boldsymbol{\xi}))$ since $x\mapsto f(x,\boldsymbol{\xi})$ is continuous with probability 1, and $x\mapsto f\left(x,\mathbb{E}\left[\boldsymbol{\xi}|\boldsymbol{\xi}\in\mathcal{R}^c\right]\right)$ is continuous.

It remains to show that $\tilde{F}_x$ is continuous and strictly increasing at $F^{-1}_x(\beta)$ for all $x\in\mathcal{X}$. Fix $x\in\mathcal{X}$. Since $F_x(\theta)$ and $\tilde{F}_x(\theta)$ coincide for $\theta\ge F^{-1}_x(\beta)$ and F_x is strictly increasing at $F^{-1}_x(\beta)$, we have

$$\begin{aligned}\tilde{F}_x\left(F^{-1}_x(\beta)+\epsilon\right) &= F_x\left(F^{-1}_x(\beta)+\epsilon\right)\\ &> F_x(F^{-1}_x(\beta))\\ &= \tilde{F}_x\left(F^{-1}_x(\beta)\right)\end{aligned}$$

and so $\tilde{F}_x$ is also strictly increasing at $F^{-1}_x(\beta)$. Finally, to show that $\tilde{F}_x$ is continuous at $F^{-1}_x(\beta)$, it suffices to show that it is left continuous, since all distribution functions are right continuous. For $\epsilon>0$ sufficiently small we have that $f(x,\mathbb{E}\left[\boldsymbol{\xi}|\boldsymbol{\xi}\in\mathcal{R}^c\right])<F^{-1}_x(\beta)$, and so

$$\begin{aligned}&\tilde{F}_x(F^{-1}_x(\beta))-\tilde{F}_x(F^{-1}_x(\beta)-\epsilon)\\ &\quad=\mathbb{P}\left(F^{-1}_x(\beta)-\epsilon<f(x,\psi_\mathcal{R}(\boldsymbol{\xi}))\le F^{-1}_x(\beta)\right)\\ &\quad=\mathbb{P}\left(\boldsymbol{\xi}\in\{\xi\in\Xi:F^{-1}_x(\beta)-\epsilon<f(x,\xi)\le F^{-1}_x(\beta)\}\cap\mathcal{R}\right)\end{aligned}$$

$$\leq \mathbb{P}\left(F_x^{-1}(\beta) - \epsilon < f(x, \boldsymbol{\xi}) \leq F_x^{-1}(\beta)\right)$$
$$= F_x(F_x^{-1}(\beta)) - F_x(F_x^{-1}(\beta) - \epsilon).$$

Now, since by assumption F_x is continuous at $F_x^{-1}(\beta)$, we have that $\lim_{\epsilon \downarrow 0} \left(F_x(F_x^{-1}(\beta)) - F_x(F_x^{-1}(\beta) - \epsilon\right) = 0$, and so must also have $\lim_{\epsilon \downarrow 0} \left(\tilde{F}_x(F_x^{-1}(\beta)) - \tilde{F}_x(F_x^{-1}(\beta) - \epsilon\right) = 0$ as required. □

In the next subsection this result will be used to show that any point in the interior of the non-risk region $\mathcal{R}^c$ will, with probability 1, be in the non-risk region of the sampled scenario set as the sample size grows large.

5.2 Equivalence of aggregation sampling with sampling from aggregated random vector

The main obstacle in showing that aggregation sampling is equivalent to sampling from the aggregated random vector is to show that the aggregated scenario in the non-risk region converges almost surely to the conditional expectation of the non-risk region as the number of specified risk scenarios tends to infinity. Recall from Sect. 4 that $N(n)$ denotes the effective sample size in aggregation sampling when we require n risk scenarios and is distributed as $n + \mathcal{NB}(n, a)$ where a is the probability of the non-risk region. The purpose of the next lemma is to show that as $n \to \infty$ the number of samples drawn from the non-risk region almost surely tends to infinity.

Lemma 2 *Suppose $M(n) \sim \mathcal{NB}(n, p)$ where $0 < p < 1$. Then with probability 1 we have that $\lim_{n\to\infty} M(n) = \infty$.*

Proof First note that,

$$\{\lim_{n\to\infty} M(n) = \infty\}^c = \bigcup_{k\in\mathbb{N}} \left(\bigcap_{n\in\mathbb{N}} \bigcup_{t>n} \{M(t) > k\}^c \right) = \bigcup_{k\in\mathbb{N}} \limsup_{n\to\infty} \{M(n) \leq k\}.$$

Hence, to show that $\mathbb{P}(\{\lim_{n\to\infty} M(n) = \infty\}) = 1$ it is enough to show for each $k \in \mathbb{N}$ we have that

$$\mathbb{P}\left(\limsup_{n\to\infty} \{M(n) \leq k\} \right) = 0. \tag{20}$$

Now, fix $k \in \mathbb{N}$. Then for all $n \in \mathbb{N}$ we have that

$$\mathbb{P}(M(n) = k) = \binom{k+n-1}{k} (1-p)^n \, p^k,$$

and in particular,

$$\mathbb{P}(M(n+1) = k) = \binom{k+n}{k} (1-p)^{n+1} p^k = \frac{k+n}{n} (1-p)\, \mathbb{P}(M(n) = k).$$

For large enough n we have that $\frac{k+n}{n}(1-p) \leq c < 1$ for some constant c, hence $\sum_{n=1}^{\infty} \mathbb{P}(M(n)=k) < +\infty$ and so

$$\sum_{n=1}^{\infty} \mathbb{P}(M(n) \leq k) = \sum_{n=1}^{\infty} \sum_{j=1}^{k} \mathbb{P}(M(n)=j) = \sum_{j=1}^{k} \sum_{n=1}^{\infty} \mathbb{P}(M(n)=j) < \infty.$$

The result (20) now holds by the first Borel–Cantelli Lemma [6, Section 4]. □

The next Corollary shows that the strong law of large numbers still applies for the conditional expectation of the non-risk region in aggregation sampling despite the sample size being a random quantity.

Corollary 3 *Suppose* $\mathbb{E}\left[\|\boldsymbol{\xi}\|\right] < +\infty$ *and* $\mathbb{P}(\boldsymbol{\xi} \in \mathcal{R}^c) > 0$. *Then with probability 1*

$$\lim_{n\to\infty} \left\| \frac{1}{N(n)-n} \sum_{i \in \mathcal{I}_{\mathcal{R}^c}(N(n))} \boldsymbol{\xi}_i - \mathbb{E}\left[\, \boldsymbol{\xi} | \boldsymbol{\xi} \in \mathcal{R}^c \,\right] \right\| = 0.$$

Proof Define the following measurable subsets of Ω^{∞}:

$$\Omega_1 = \left\{ \omega \in \Omega^{\infty} : \lim_{n\to\infty} (N(n)(\omega) - n) = \infty \right\},$$
$$\Omega_2 = \left\{ \omega \in \Omega^{\infty} : \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\mathcal{R}^c}(\boldsymbol{\xi}_i(\omega)) \boldsymbol{\xi}_i(\omega) = \mathbb{E}\left[\mathbb{1}_{\mathcal{R}^c}(\boldsymbol{\xi})\boldsymbol{\xi}\right] \right\},$$
$$\Omega_3 = \left\{ \omega \in \Omega^{\infty} : \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\mathcal{R}^c}(\boldsymbol{\xi}_i) = \mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{R}^c\right) \right\}.$$

By the strong law of large numbers Ω_2 and Ω_3 have probability one. Since $N(n) - n \sim \mathcal{NB}(n, a)$, where $a = \mathbb{P}(\boldsymbol{\xi} \in \mathcal{R}^c)$, Ω_1 has probability 1 by Lemma 2. Therefore, $\Omega_1 \cap \Omega_2 \cap \Omega_3$ has probability 1 and so it is enough to show that for any $\omega \in \Omega_1 \cap \Omega_2 \cap \Omega_3$ we have that

$$\frac{1}{N(n)(\omega)-n} \sum_{i \in \mathcal{I}_{\mathcal{R}^c}(N(n))} \boldsymbol{\xi}_i(\omega) \to \mathbb{E}\left[\, \boldsymbol{\xi} | \boldsymbol{\xi} \in \mathcal{R}^c \,\right] \text{ as } n \to \infty.$$

Let $\omega \in \Omega_1 \cap \Omega_2 \cap \Omega_3$. Since $\omega \in \Omega_2 \cap \Omega_3$, we have that as $m \to \infty$:

$$\frac{1}{\frac{1}{m}\sum_{i=1}^{m} \mathbb{1}_{\mathcal{R}^c}(\boldsymbol{\xi}_i(\omega))} \frac{1}{m} \sum_{i=1}^{m} \mathbb{1}_{\mathcal{R}^c}(\boldsymbol{\xi}_i(\omega)) \boldsymbol{\xi}_i(\omega)$$
$$\to \frac{1}{\mathbb{P}(\boldsymbol{\xi} \in \mathcal{R}^c)} \mathbb{E}\left[\mathbb{1}_{\mathcal{R}^c}(\boldsymbol{\xi})\boldsymbol{\xi}\right] = \mathbb{E}\left[\boldsymbol{\xi} | \boldsymbol{\xi} \in \mathcal{R}^c\right].$$

Now, fix $\epsilon > 0$. Then there exists $N_1(\omega) \in \mathbb{N}$ such

$$m > N_1(\omega) \implies \left\| \frac{1}{\frac{1}{m}\sum_{i=1}^{m} \mathbb{1}_{\mathcal{R}^c}(\boldsymbol{\xi}_i(\omega))} \frac{1}{m} \sum_{i=1}^{m} \mathbb{1}_{\mathcal{R}^c}\left(\boldsymbol{\xi}_i(\omega)\right) \boldsymbol{\xi}_i(\omega) - \mathbb{E}\left[\boldsymbol{\xi} | \boldsymbol{\xi} \in \mathcal{R}^c\right] \right\| < \epsilon.$$

Since $\omega \in \Omega_1$ there exists $N_2(\omega)$ such that

$$n > N_2(\omega) \implies N(n)(\omega) > N_1(\omega).$$

Noting that

$$\begin{aligned} &\frac{1}{\frac{1}{N(n)(\omega)}\sum_{i=1}^{N(n)(\omega)} \mathbb{1}_{\mathcal{R}^c}(\boldsymbol{\xi}_i(\omega))} \frac{1}{N(n)(\omega)} \sum_{i=1}^{N(n)(\omega)} \mathbb{1}_{\mathcal{R}^c}(\boldsymbol{\xi}_i(\omega))\boldsymbol{\xi}_i(\omega) \\ &\quad = \frac{1}{\frac{N(n)(\omega)-n}{N(n)(\omega)}} \frac{1}{N(n)(\omega)} \sum_{i=1}^{N(n)(\omega)} \mathbb{1}_{\mathcal{R}^c}(\boldsymbol{\xi}_i(\omega))\boldsymbol{\xi}_i(\omega) \\ &\quad = \frac{1}{N(n)(\omega) - n} \sum_{i \in \mathcal{I}_{\mathcal{R}^c}(N(n))} \boldsymbol{\xi}_i, \end{aligned}$$

we have that

$$n > N_2(\omega) \implies \left\| \frac{1}{N(n)(\omega) - n} \sum_{i \in \mathcal{I}_{\mathcal{R}^c}(N(n))} \boldsymbol{\xi}_i(\omega) - \mathbb{E}\left[\boldsymbol{\xi} | \boldsymbol{\xi} \in \mathcal{R}^c\right] \right\| < \epsilon$$

and so $\frac{1}{N(n)(\omega)-n} \sum_{i \in \mathcal{I}_{\mathcal{R}^c}(N(n))} \boldsymbol{\xi}_i(\omega) \to \mathbb{E}\left[\boldsymbol{\xi} | \boldsymbol{\xi} \in \mathcal{R}^c\right]$ as $n \to \infty$. □

To show that aggregation sampling yields solutions consistent with the underlying random vector $\boldsymbol{\xi}$, we show that with probability 1, for n large enough, it is equivalent to sampling from the aggregated random vector $\psi_{\mathcal{R}}(\boldsymbol{\xi})$, as defined in Definition 6. If the region $\mathcal{R}$ satisfies the aggregation condition, and $\mathbb{E}\left[\boldsymbol{\xi} | \boldsymbol{\xi} \in \mathcal{R}^c\right] \in \mathcal{R}^c_{\mathcal{X}}$, Theorem 1 tells us that $\rho_\beta\left(f(x, \psi_{\mathcal{R}}(\boldsymbol{\xi}))\right) = \rho_\beta\left(f(x, \boldsymbol{\xi})\right)$ for all $x \in \mathcal{X}$. Hence, if sampling is consistent for the risk measure ρ_β, then aggregation sampling is also consistent.

Noting that $|\mathcal{I}_{\mathcal{R}^c}(N(n))| = N(n) - n$, we introduce the following notation for the empirical distribution and quantile functions for loss function with scenario set constructed by aggregation sampling with n risk scenarios.

$$\hat{F}_{n,x}(\theta) = \frac{1}{N(n)} \left((N(n) - n) \mathbb{1}_{\{\xi \in \Xi:\, f(x,\xi) \le \theta\}} \left(\frac{1}{N(n) - n} \sum_{i \in \mathcal{I}_{\mathcal{R}^c}(N(n))} \boldsymbol{\xi}_i \right) + \sum_{i \in \mathcal{I}_{\mathcal{R}}(n)} \mathbb{1}_{\{\xi \in \Xi:\, f(x,\xi) \le \theta\}}(\boldsymbol{\xi}_i) \right)$$

$$\hat{F}^{-1}_{n,x}(u) = \inf\{\theta \in \mathbb{R} : \ \hat{F}_{n,x}(\theta) \geq u\}$$

Note that these latter functions will depend on the sample $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{N(n)}$.

Theorem 3 *Let $\mathcal{R}_{\mathcal{X}} \subseteq \mathcal{R} \subset \mathbb{R}^d$ be a set satisfying the aggregation condition. Suppose that that conditions (i)–(v) from Theorem 2 and Corollary 2 hold, and in addition that*

(vi) *For each $x \in \mathcal{X}$, $\xi \mapsto f(x, \xi)$ is continuous at $\mathbb{E}\left[\ \boldsymbol{\xi} | \boldsymbol{\xi} \in \mathcal{R}^c\ \right]$*

Then, with probability 1, for all $u \geq \beta$

$$\lim_{n\to\infty} \sup_{x\in\mathcal{X}} |\hat{F}^{-1}_{n,x}(u) - \tilde{F}^{-1}_{n,x}(u)| = 0. \tag{21}$$

Proof We actually prove a slightly stronger result, that is, with probability 1, there exists $N > 0$ such that for all $n > N$, $x \in \mathcal{X}$ and $u \geq \beta$ we have that $\hat{F}^{-1}_{n,x}(u) = \tilde{F}^{-1}_{n,x}(u)$. First, note that if

$$\theta \geq \max\left\{ f\left(x, \frac{1}{N(n)-n} \sum_{i\in\mathcal{I}_{\mathcal{R}^c}(N(n))} \boldsymbol{\xi}_i\right), f\left(x, \mathbb{E}\left[\ \boldsymbol{\xi} | \boldsymbol{\xi} \in \mathcal{R}^c\ \right]\right)\right\}$$

then

$$\begin{aligned}\hat{F}_{n,x}(\theta) &= \frac{N(n)-n}{N(n)} + \frac{1}{N(n)} \sum_{i\in\mathcal{I}_{\mathcal{R}}(N(n))} \mathbb{1}_{\{\xi\in\Xi:\ f(x,\xi)\leq\theta\}}(\boldsymbol{\xi}_i) \\ &= \tilde{F}_{N(n),x}(\theta).\end{aligned}$$

So if the following holds with probability 1

$$\begin{aligned}\liminf_{n\to\infty} \inf_{x\in\mathcal{X}} \Bigg(\tilde{F}^{-1}_{n,x}(\beta) - \max\Bigg(f\left(x, \frac{1}{N(n)-n} \sum_{i\in\mathcal{I}_{\mathcal{R}^c}(N(n))} \boldsymbol{\xi}_i\right), \\ f\left(x, \mathbb{E}\left[\ \boldsymbol{\xi} | \boldsymbol{\xi} \in \mathcal{R}^c\ \right]\right)\Bigg)\Bigg) > 0\end{aligned} \tag{22}$$

then, by application of Lemma 1, this implies that with probability 1, there exists $N > 0$ such that for all $n > N$ and for all $u \geq \beta$ and $x \in \mathcal{X}$ we have $\hat{F}^{-1}_{n,x}(u) = \tilde{F}^{-1}_{N(n),x}(u)$ as required. Since $\mathbb{E}\left[\ \boldsymbol{\xi} | \boldsymbol{\xi} \in \mathcal{R}^c\ \right] \in \mathcal{R}^c_{\mathcal{X}}$ we have that $f(x, \mathbb{E}\left[\ \boldsymbol{\xi} | \boldsymbol{\xi} \in \mathcal{R}^c\ \right]) < F^{-1}_x(\beta)$ for all $x \in \mathcal{X}$, and since $\mathcal{X}$ is compact there exists $\delta > 0$ such that

$$\inf_{x\in\mathcal{X}} \left(F^{-1}_x(\beta) - f\left(x, \mathbb{E}\left[\ \boldsymbol{\xi} | \boldsymbol{\xi} \in \mathcal{R}^c\ \right]\right)\right) > \delta. \tag{23}$$

By Corollary 3, the compactness of $\mathcal{X}$ and the continuity of $\xi \mapsto f(x,\xi)$ at $\mathbb{E}\left[\boldsymbol{\xi}|\boldsymbol{\xi} \in \mathcal{R}^c\right]$, we have with probability 1

$$\limsup_{n\to\infty} \sup_{x\in\mathcal{X}} \left| f\left(x, \frac{1}{N(n)-n} \sum_{i\in\mathcal{I}_{\mathcal{R}^c}(N(n))} \boldsymbol{\xi}_i\right) - f\left(x, \mathbb{E}\left[\boldsymbol{\xi}|\boldsymbol{\xi} \in \mathcal{R}^c\right]\right) \right| = 0. \quad (24)$$

Also, by Corollary 2, with probability 1

$$\limsup_{n\to\infty} \sup_{x\in\mathcal{X}} \left| F_x^{-1}(\beta) - \tilde{F}_{N(n),x}^{-1}(\beta) \right| = 0. \quad (25)$$

Thus, letting $z(x) = F_x^{-1}(\beta) - f\left(x, \mathbb{E}\left[\boldsymbol{\xi}|\boldsymbol{\xi} \in \mathcal{R}^c\right]\right)$, we also have with probability 1 that

$$\lim_{n\to\infty} \sup_{x\in\mathcal{X}} \left| \left(\tilde{F}_{N(n),x}^{-1}(\beta) - f\left(x, \frac{1}{N(n)-n} \sum_{i\in\mathcal{I}_{\mathcal{R}^c}(N(n))} \boldsymbol{\xi}_i\right)\right) - z(x) \right| = 0.$$

In particular, with probability 1 there exists N such that for $n > N$

$$\sup_{x\in\mathcal{X}} \left| \left(\tilde{F}_{N(n),x}^{-1}(\beta) - f\left(x, \frac{1}{N(n)-n} \sum_{i\in\mathcal{I}_{\mathcal{R}^c}(N(n))} \boldsymbol{\xi}_i\right)\right) - z(x) \right| < \frac{\delta}{2}. \quad (26)$$

In which case, for $n > N$

$$\begin{aligned}
&\inf_{x\in\mathcal{X}} \left(\tilde{F}_{N(n),x}^{-1}(\beta) - f\left(x, \frac{1}{N(n)-n} \sum_{i\in\mathcal{I}_{\mathcal{R}^c}(N(n))} \boldsymbol{\xi}_i\right)\right) \\
&= \inf_{x\in\mathcal{X}} \left(z(x) + \tilde{F}_{N(n),x}^{-1}(\beta) - f\left(x, \frac{1}{N(n)-n} \sum_{i\in\mathcal{I}_{\mathcal{R}^c}(N(n))} \boldsymbol{\xi}_i\right) - z(x)\right) \\
&\geq \inf_{x\in\mathcal{X}} z(x) - \sup_{x\in\mathcal{X}} \left| \left(\tilde{F}_{N(n),x}^{-1}(\beta) - f\left(x, \frac{1}{N(n)-n} \sum_{i\in\mathcal{I}_{\mathcal{R}^c}(N(n))} \boldsymbol{\xi}_i\right)\right) - z(x) \right| \\
&> \delta - \frac{\delta}{2} = \frac{\delta}{2} \qquad \text{by (23) and (26).}
\end{aligned}$$

We can similarly show that

$$\limsup_{n\to\infty} \inf_{x\in\mathcal{X}} \left(\tilde{F}_{N(n),x}^{-1}(\beta) - f\left(x, \mathbb{E}\left[\boldsymbol{\xi}|\boldsymbol{\xi} \in \mathcal{R}^c\right]\right)\right) > 0$$

holds with probability 1. Hence (22) holds with probability 1 and the proof is complete. □

Note that although the continuity conditions (ii), (v) and (vi) look complicated, the loss function $f : \mathcal{X} \times \Xi \to \mathbb{R}$ will typically be continuous everywhere, and so these will be satisfied automatically.

6 A conservative risk region for monotonic loss functions

In order to use risk regions for scenario generation, we need to have a characterization of the risk region which conveniently allows us to test membership. In general this is a difficult as the risk region depends on the loss function, the distribution and the problem constraints. Therefore, as a proof-of-concept, in the following two sections we derive risk regions for two classes of problems. In this section we propose a conservative risk region for problems which have monotonic loss functions.

Definition 7 (*Monotonic loss function*) A loss function $f : \mathcal{X} \times \Xi \to \mathbb{R}$ is monotonic increasing if for all $x \in \mathcal{X}$ and $\xi, \tilde{\xi} \in \Xi$ such that $\xi < \tilde{\xi}$ we have $f(x, \xi) < f(x, \tilde{\xi})$. Similarly, we say it is monotonic decreasing if for all $x \in \mathcal{X}$ and $\xi, \tilde{\xi} \in \Xi$ such that $\xi < \tilde{\xi}$ we have $f(x, \xi) > f(x, \tilde{\xi})$.

Monotonic loss functions occur naturally in stochastic linear programming. The following result presents a class of loss functions which arise in the context of network design, and gives conditions under which they are monotonic.

Proposition 4 *Suppose $\mathcal{X} \subseteq \mathbb{R}^k_+$, $\Xi \subseteq \mathbb{R}^d_+$ and the loss function $Q(x, \xi)$ is defined to be the optimal value to the following linear program:*

$$\min_{y,z} q^T y + u^T z \tag{27}$$

$$\text{such that } Wy + z \geq \xi \tag{28}$$

$$By \leq b \tag{29}$$

$$Ty \leq Vx \tag{30}$$

$$\mathcal{N}y = 0 \tag{31}$$

$$y, z \geq 0, \tag{32}$$

where $W, B, T, V, \mathcal{N}$ are matrices and q, u, b are vectors of compatible dimensions. Then, $Q(x, \xi)$ is monotonic increasing under the following conditions:

1. $q, u > 0$,
2. $b \geq 0$,
3. $W, B, T, V \geq 0$.

Proof Fix $x \in \mathcal{X}$. The problem is always feasible since $y = 0$ and $z = \xi$ is a feasible solution. Since $x \geq 0$ and $q, u > 0$ the problem (27)–(32) is bounded below by zero. In addition, when $\xi \geq 0$ with at least one component strictly greater than zero, the optimal solution (y^*, z^*) must contain at least one strictly positive element due to constraint (28) and the fact that $W \geq 0$, and so in this case the optimal value is strictly positive. Because the problem is both bounded below and feasible, strong duality applies and so $Q(x, \xi)$ is also equal to the optimal solution to the dual problem:

$$\max_{\pi,\nu,\eta,\lambda} \xi^T \pi - x^T V^T \nu - b^T \eta \tag{33}$$

$$\text{such that } W^T \pi - T^T \nu - B^T \eta + \mathcal{N}^T \lambda \leq q \tag{34}$$

$$\pi \leq u \tag{35}$$

$$\pi, \nu, \eta \geq 0. \tag{36}$$

Let $\bar{\xi}, \tilde{\xi} \in \Xi$ be such that $\bar{\xi} < \tilde{\xi}$. In the first case suppose that $\bar{\xi} \neq 0$, and let $(\bar{\pi}, \bar{\nu}, \bar{\eta}, \bar{\lambda})$ be the optimal dual variables for (33)–(36) for $\xi = \bar{\xi}$. As discussed above, this means that $Q(x, \bar{\xi}) > 0$, and given that $x^T V^T \bar{\nu} + b^T \bar{\eta} \geq 0$, at least one component of $\bar{\pi}$ will be greater than zero in order for the objective of the dual to be strictly positive. Now, $(\bar{\pi}, \bar{\nu}, \bar{\eta}, \bar{\lambda})$ is also a feasible solution to the dual problem with $\xi = \tilde{\xi}$ and so

$$\begin{aligned} Q(x, \bar{\xi}) &= \bar{\xi}^T \bar{\pi} - x^T V^T \bar{\nu} - b^T \bar{\eta} \\ &< \tilde{\xi}^T \bar{\pi} - x^T V^T \bar{\nu} - b^T \bar{\eta} \\ &\leq Q(x, \tilde{\xi}). \end{aligned}$$

In the second case suppose that $\bar{\xi} = 0$. In this case $y = 0,\ z = 0$ is a feasible solution to the primal problem (27)–(32) with $\xi = \bar{\xi}$ and this solution has an objective value of zero. Since the objective is bounded below by zero, this means this solution is also optimal and so $Q(x, \bar{\xi}) = 0$. Since $\tilde{\xi} > 0$ we have that $Q(x, \tilde{\xi}) > 0$, and so $Q(x, \bar{\xi}) < Q(x, \tilde{\xi})$. Hence $Q(x, \xi)$ is monotonic as required. □

This recourse function arises in stochastic network design, and the problem formulation in the previous proposition was adapted from a model in the paper [31]. In this type of problem, we have a network consisting of suppliers, processing units, and customers, and decisions must be made relating to opening facilities and the capacities of nodes and arcs. The problem which defines the recourse function $Q(x, \xi)$ depends on the capacity and opening decisions x of the first stage, and the demand of the customers ξ. The aim of the problem is construct of flow of products y which minimize transportation costs for satisfying customers demand, plus penalties for any unsatisfied demand z.

For a problem with a monotonic loss function, the following result defines a conservative risk region.

Theorem 4 *Suppose the loss function $f : \mathcal{X} \times \Xi \to \mathbb{R}$ is monotonic increasing. Then the following set is a conservative risk region:*

$$\mathcal{R}_1 = \{\xi \in \Xi : \mathbb{P}(\boldsymbol{\xi} > \xi) \leq 1 - \beta\}. \tag{37}$$

Similarly, if the loss function is monotonic decreasing then the following set is a conservative risk region:

$$\mathcal{R}_2 = \{\xi \in \Xi : \mathbb{P}(\boldsymbol{\xi} < \xi) \leq 1 - \beta\}. \tag{38}$$

Proof Suppose $f(x, \xi)$ is monotonic increasing and let $\xi \in \mathcal{R}_{\mathcal{X}}$, then

$$\mathbb{P}(\boldsymbol{\xi} > \xi) \leq \mathbb{P}(f(x, \boldsymbol{\xi}) > f(x, \xi)) \qquad \text{by monotonicity}$$

$$= 1 - \underbrace{\mathbb{P}\left(f(x, \boldsymbol{\xi}) \leq f(x, \xi)\right)}_{\geq \beta}$$
$$\leq 1 - \beta$$

and so $\xi \in \mathcal{R}_1$ as required. The set $\mathcal{R}_2$ can similarly be shown to be a conservative risk region when $f(x, \xi)$ is monotonic decreasing. □

7 An exact risk region for the portfolio selection problem

In this section, we characterize exactly the risk region of the portfolio selection problem when the distribution of asset returns belongs to a certain class of distributions.

In the portfolio selection problem one aims to choose a portfolio of financial assets with uncertain returns. For $i = 1, \ldots, d$, let x_i denote the amount to invest in asset i, and $\boldsymbol{\xi}_i$ the random return of asset i. The loss function in this problem is the negative total return, that is $f(x, \boldsymbol{\xi}) = \sum_{i=1}^{d} -x_i \boldsymbol{\xi}_i = -x^T \boldsymbol{\xi}$, and $\Xi = \mathbb{R}^d$. The set $\mathcal{X} \subset \mathbb{R}^d$ of feasible portfolios may encompass constraints like no short-selling ($x \geq 0$), total investment ($\sum_{i=1}^{d} x_i = 1$) and quotas on certain stocks ($x \leq c$).

The following corollary gives sufficient conditions for the risk region to satisfy the aggregation condition, and for aggregation sampling to be consistent.

Corollary 4 *Suppose that $\mathcal{R} \supseteq \mathcal{R}_{\mathcal{X}}$ and that the following conditions hold:*

1. *$\boldsymbol{\xi}$ is continuous with support $\mathbb{R}^d$,*
2. *There exists $x_1, x_2 \in \mathcal{X}$ which are linearly independent,*
3. *$0 \notin \mathcal{X}$,*
4. *$\mathcal{X}$ is compact.*

Then $\mathcal{R}$ satisfies the aggregation condition, and aggregation sampling with respect to $\mathcal{R}$ is consistent in the sense of Theorem 3.

Proof To prove that $\mathcal{R}$ satisfies the aggregation condition, it is enough to show that $\mathcal{R}_{\mathcal{X}}$ satisfies the aggregation condition. We prove this by showing that all the conditions of Proposition 2 hold. Note that $x \mapsto -x^T \xi$ is continuous so condition (i) of Proposition 2 holds immediately.

For each $x \in \mathcal{X}$ the interior of the corresponding risk region and non-risk region are open half-spaces:

$$\text{int}\left(\mathcal{R}_x\right) = \{\xi \in \mathbb{R}^d : -x^T \xi > F_x^{-1}(\beta)\} \text{ and}$$
$$\text{int}\left(\mathcal{R}_x^c\right) = \{\xi \in \mathbb{R}^d : -x^T \xi < F_x^{-1}(\beta)\}.$$

Fix $\bar{x} \in \mathcal{X}$. Then either $\bar{x}$ is linearly independent to x_1 or it is linearly independent to x_2. Assume it is linearly independent to x_1. Now, $\text{int}\left(\mathcal{R}_{\bar{x}}\right)$ and $\text{int}\left(\mathcal{R}_{x_1}\right)$ are non-parallel half-spaces and so both $\text{int}\left(\mathcal{R}_{\bar{x}} \cap \mathcal{R}_{x_1}\right)$ and $\text{int}\left(\mathcal{R}_{x_1} \backslash \mathcal{R}_{\bar{x}}\right) = \text{int}\left(\mathcal{R}_{x_1}\right) \cap \text{int}\left(\mathcal{R}_{\bar{x}}^c\right)$ are non-empty, and since we also have $\Xi = \mathbb{R}^d$, condition (ii) of Proposition 2 is satisfied.

Since $\mathcal{R}_{x_1}$ and $\mathcal{R}_{x_2}$ are non-parallel half-spaces, their union $\mathcal{R}_{x_1} \cup \mathcal{R}_{x_2}$ is connected. Similarly, for any $x \in \mathcal{X}$, we must have $\mathcal{R}_x$ being non-parallel with either

$\mathcal{R}_{x_1}$ or $\mathcal{R}_{x_2}$ and so $\mathcal{R}_x \cup \mathcal{R}_{x_1} \cup \mathcal{R}_{x_2}$ must also be connected. Hence, $\mathcal{R}_{\mathcal{X}} = \bigcup_{x \in \mathcal{X}} \left(\mathcal{R}_x \cup \mathcal{R}_{x_1} \cup \mathcal{R}_{x_2}\right)$ is connected so condition (iii) of Proposition 2 is also satisfied. Hence $\mathcal{R}$ satisfies the aggregation condition.

We show that aggregation sampling is consistent in the sense of Theorem 3 by showing that the conditions of this theorem hold. We have already shown that condition (i) of Theorem 3 holds. The loss function is continuous, and so condition (ii) of Theorem 3 holds. Let $\epsilon > 0$, then

$$\begin{aligned} &F_x\left(F_x^{-1}(\beta) + \epsilon\right) - F_x\left(F_x^{-1}(\beta)\right) \\ &\quad = \mathbb{P}\left(\boldsymbol{\xi} \in \{\xi \in \mathbb{R}^d : F_x^{-1}(\beta) < -x^T\xi \leq F_x^{-1}(\beta) + \epsilon\}\right). \end{aligned}$$

Since $x \neq 0$, the set defining this event has a non-empty interior, and since the support of $\boldsymbol{\xi}$ is $\mathbb{R}^d$, this probability is greater than zero. Hence, F_x is increasing at $F_x^{-1}(\beta)$. Since $\boldsymbol{\xi}$ is continuous, we also have that F_x is continuous and so condition (iii) of Theorem 3 holds.

By Proposition 3 $\mathcal{R}_{\mathcal{X}}^c$ is convex, and since $\mathcal{R}^c \subseteq \mathcal{R}_{\mathcal{X}}^c$ and $\mathcal{R}_{\mathcal{X}}$ is open we have $\mathbb{E}\left[\boldsymbol{\xi} | \boldsymbol{\xi} \in \mathcal{R}^c\right] \in \text{int}\left(\mathcal{R}_{\mathcal{X}}^c\right)$, and so condition (iv) of Theorem 3 holds. Finally, condition (v) of Theorem 3 holds by assumption and so aggregation sampling with the set $\mathcal{R}$ is consistent in sense of Theorem 3. □

Elliptical distributions are a general class of distributions which include among others the multivariate Normal and multivariate t-distributions. See [16] for a full overview of the subject.

Definition 8 (*Spherical and elliptical distributions*) Let $\boldsymbol{\zeta}$ be a random vector in $\mathbb{R}^d$. Then $\boldsymbol{\zeta}$ is said to be *spherical* if its distribution is invariant under orthonormal transformations; that is, if

$$\boldsymbol{\zeta} \sim U\boldsymbol{\zeta} \qquad \text{for all } U \in \mathbb{R}^{d \times d} \text{ orthonormal.}$$

Let $\boldsymbol{\xi}$ be a random vector in $\mathbb{R}^d$. Then $\boldsymbol{\xi}$ is said to be *elliptical* if it can be written $\boldsymbol{\xi} = P\boldsymbol{\zeta} + \mu$ where $P \in \mathbb{R}^{d \times d}$ is non-singular, $\mu \in \mathbb{R}^d$, and $\boldsymbol{\zeta}$ is random vector with spherical distribution. We will denote this $\boldsymbol{\xi} \sim \text{Elliptical}(\boldsymbol{\zeta}, P, \mu)$.

An important property of elliptical distributions is that for any $x \in \mathbb{R}^d$ we can characterize exactly the distribution of $x^T\boldsymbol{\xi}$. If $\boldsymbol{\xi} \sim \text{Elliptical}(\boldsymbol{\zeta}, P, \mu)$ then:

$$-x^T\boldsymbol{\xi} \sim \|Px\|\,\boldsymbol{\zeta}_1 - x^T\mu, \tag{39}$$

where $\boldsymbol{\zeta}_1$ is the first component of the random vector $\boldsymbol{\zeta}$, and $\|\cdot\|$ denotes the standard Euclidean norm. By (39) the β-quantile of the loss of a portfolio is as follows:

$$F_x^{-1}(\beta) = \|Px\|\,F_{\boldsymbol{\zeta}_1}^{-1}(\beta) - x^T\mu.$$

Therefore, the exact risk region for $\boldsymbol{\xi} \sim \text{Elliptical}(\boldsymbol{\zeta}, P, \mu)$, is as follows:

$$\bigcup_{x \in \mathcal{X}} \{\xi \in \mathbb{R}^d : -x^T \xi \geq \| Px \| F_{\zeta_1}^{-1}(\beta) - x^T \mu\}. \tag{40}$$

This characterization is not practical for testing whether or not a point belongs to the risk region, which is required for our scenario generation algorithms. However, a more convenient form is available in the case where $\mathcal{X} \subset \mathbb{R}^d$ is convex. Before stating the result, we recall the concept of a projection onto a convex set.

Definition 9 (*Projection*) Let $C \subset \mathbb{R}^d$ be a closed convex set. Then for any point $\xi \in \mathbb{R}^d$, we define the projection of ξ onto C to be the unique point $p_C(\xi) \in C$ such that $\inf_{x \in C} \| x - \xi \| = \| p_C(\xi) - \xi \|$.

By a slight abuse of notation, for a set $\mathcal{A} \subset \mathbb{R}^d$ and a matrix $T \in \mathbb{R}^{d \times d}$, we write $T(\mathcal{A}) := \{T\xi : \xi \in \mathcal{A}\}$. Finally, recall that the conic hull of a set $\mathcal{A} \subset \mathbb{R}^d$, which we denote $\text{conic}(\mathcal{A})$, is the smallest convex cone containing $\mathcal{A}$.

Theorem 5 *Suppose* $\boldsymbol{\xi} \sim \text{Elliptical}(\boldsymbol{\zeta}, P, \mu)$ *and* $\mathcal{X} \subset \mathbb{R}^d$ *is a convex set. Then the exact non-risk region in* (40) *can be written as follows:*

$$\mathcal{R}_{\mathcal{X}}^c = P^T \left(\{\tilde{\xi} \in \mathbb{R}^d : \left\| p_{K'}(\tilde{\xi} - \tilde{\mu}) \right\| < F_{\zeta_1}^{-1}(\beta)\} \right) \tag{41}$$

where $\tilde{\mu} = (P^T)^{-1}\mu$, $K' = -PK$ *and* $K = \text{conic}(\mathcal{X})$.

Proof

$$\begin{aligned}
\mathcal{R}_{\mathcal{X}}^c &= \{\xi \in \mathbb{R}^d : -x^T \xi < \| Px \| F_{\zeta_1}^{-1}(\beta) - x^T \mu \ \forall x \in \mathcal{X}\} \\
&= \{\xi \in \mathbb{R}^d : \tilde{x}^T \xi < \| P\tilde{x} \| F_{\zeta_1}^{-1}(\beta) + \tilde{x}^T \mu \ \forall \tilde{x} \in -\mathcal{X}\} \\
&= \{\xi \in \mathbb{R}^d : \tilde{x}^T (\xi - \mu) < \| P\tilde{x} \| F_{\zeta_1}^{-1}(\beta) \ \forall \tilde{x} \in -\mathcal{X}\} \\
&= P^T \left(\{\tilde{\xi} \in \mathbb{R}^d : \left\| p_{K'}(\tilde{\xi} - \tilde{\mu}) \right\| < F_{\zeta_1}^{-1}(\beta)\} \right) \\
&\quad \text{by application of Corollary 6 in Appendix B}
\end{aligned}$$

□

8 Numerical tests

In this section, we test the performance of the methodology developed in this paper. For the portfolio selection problem, when $\mathcal{X} \subseteq \mathbb{R}_+^d \backslash \{0\}$ the loss function $f(x, \xi) = -x^T \xi$ is monotonic decreasing. We therefore use this problem throughout this section to test both the conservative risk region presented in Sect. 6, and the exact risk region presented in Sect. 7.

In order to test whether a point belongs to the exact non-risk region in (41) requires the projection of a point onto a convex cone. This can be done by solving a small linear complementarity problem. See [36] or our follow-up paper [15] for more details. We solve linear complementarity problems using code from the Siconos numerics library [1]. To test whether a point $\xi \in \mathcal{Z}$ belongs to the conservative risk region in (38), involves the evaluation of the probability $\mathbb{P}(\boldsymbol{\xi} < \xi)$. Since calculating this probability exactly involves evaluating a multidimensional integral we approximate the probability by taking a large sample from $\boldsymbol{\xi}$, and using the empirical distribution function of this sample. Repeatedly testing membership of both types of risk region is therefore computationally intensive. Ways of mitigating this issue are discussed in our follow-up paper [15]. These membership tests, and the aggregation sampling algorithm have been implemented and made available as a package for the Julia programming language [14]. All experiments were conducted on a laptop with an Intel Core i7-720QM CPU at 1.6 GHz.

8.1 Probability of risk regions

As discussed in Sect. 4.1, the performance of the aggregation sampling algorithm with respect to standard Monte Carlo sampling improves as the probability of the aggregation region increases. In this first experiment we observe the behavior of this probability over a range of dimensions.

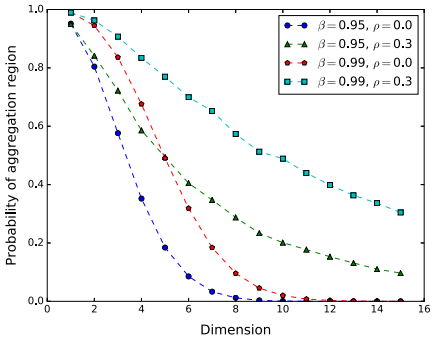
For this experiment, we suppose that $K = \text{conic}(\mathcal{X}) = \mathbb{R}^d_+$, and that the random vector follows a multivariate Normal distribution $\mathcal{N}(0, \Lambda(\rho))$, where the covariance matrix $\Lambda(\rho)$, for $0 \leq \rho < 1$, is defined as follows:

$$\Lambda_{ij}(\rho) = \begin{cases} \rho & \text{if } i \neq j, \\ 1 & \text{otherwise.} \end{cases}$$

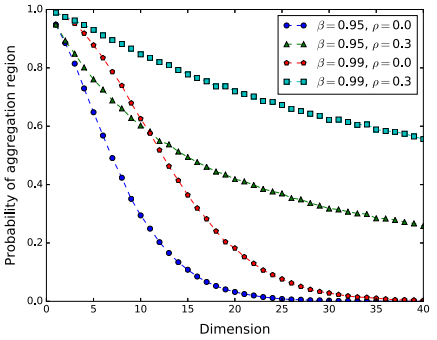
In particular, we calculate the probability for the case $\rho = 0$, that is where the asset returns are independently distributed, and the case $\rho = 0.3$, that is where the asset returns are positively correlated. The probabilities of the non-risk regions are estimated by sampling and testing membership for 20,000 points.

The results of this experiment are plotted in Fig. 1. In Fig. 1a, b are plotted the probabilities of the conservative and exact aggregation regions. To aid the readers' intuition we have also plotted a reduced scenario set in two dimensions using conservative and exact risk regions in Fig. 1c, d for $\rho = 0.3$ and $\beta = 0.95$.

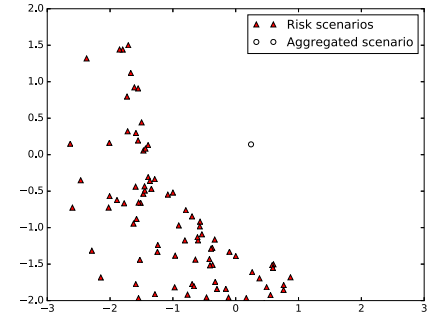
The figures show that not only is the probability of the conservative aggregation region smaller than that of the exact aggregation region but also it decays much more quickly. This emphasizes the importance of using an exact risk region for aggregation sampling if possible. Interestingly the probability of the aggregation regions for the correlated asset returns is greater and decays more slows than that of the independent asset returns. This tells us that, in addition to the loss function, the performance of our methodology depends strongly on the distribution of the random vector. Although the probability of the conservative aggregation region decays fairly rapidly, it remains non-negligible for random vectors of a moderate dimension, around 15, for the corre-



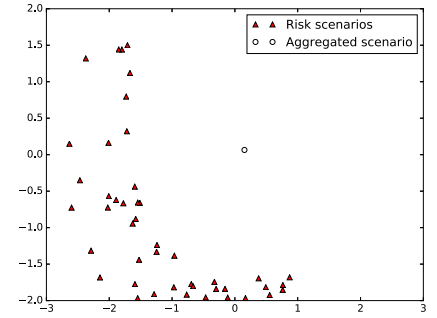
(a) Probability of conservative aggregation region



(b) Probability of exact risk region



(c) Scenario set constructed by aggregation reduction using conservative aggregation region



(d) Scenario set constructed by aggregation reduction using exact aggregation region

Fig. 1 Probabilities of conservative and exact aggregation regions

lated asset returns. For exact aggregation regions, the probability remains high for the correlated asset returns for up to a dimension of 40.

8.2 Performance of aggregation sampling

We now test the performance of the aggregation sampling algorithm using conservative and exact risk regions against standard Monte Carlo sampling in terms of the quality of the solutions each method yields.

Experimental Set-up We use the following problem:

$$\begin{aligned} \underset{x \geq 0}{\text{minimize}} \;\; & \beta\text{-CVaR}(-x^T \boldsymbol{\xi}) \\ \text{subject to} \; & x^T \mu \geq t \\ & \sum_{i=1}^{d} x_i = 1 \\ & x \geq 0. \end{aligned} \tag{P}$$

where the asset returns follow a multivariate Normal distribution $\mathcal{N}(\mu, \Sigma)$. We use two distributions: one of dimension 5 and another of dimension 10. These distributions have been fitted from monthly return data for randomly selected companies in the FTSE 100 index. The problem is thus to select a portfolio which minimizes the conditional value-at-risk of the one-month return, subject to a minimimum expected return of t, and no short-selling. These distributions have been made available online [13] in an HDF5 file, and can be accessed using the keys "normal/dim = 5/dist 1" and "normal/dim = 10/dist 1". We use the target expected one-month return $t = 0.005$ which is feasible for the constructed problems.

This problem has been chosen so that we can solve the problem exactly for Normally distributed returns, and so calculate the optimality gap for solutions found from solving scenario-based approximations. The following formula is easily verified by recalling that for continuous probability distributions, the β-CVaR is just the conditional expectation of the random variable above the β-quantile (see [29] for instance):

$$\beta\text{-CVaR}(-x^T\boldsymbol{\xi}) = (1-\beta)\mu^T x + \sqrt{x^T\Sigma x}\int_{\Phi^{-1}(\beta)}^{\infty} z\, d\Phi(z) \tag{42}$$

where Φ denotes the distribution function of the standard Normal distribution. The problem (P) can therefore be solved exactly using an interior point algorithm and in our experiments we use the software package IPOPT [37] to do this.

Denote by $\{(\xi_s, p_s)\}_{s=1}^n$ a scenario set of size n, where ξ_s denotes the vector of asset returns in scenario s, and p_s the corresponding probability. Then, the scenario-based approximation to (P) using this scenario set, is the following linear program:

$$\begin{aligned}
\underset{x,y,\alpha}{\text{minimize}}\ & \alpha + \frac{1}{1-\beta}\sum_{s=1}^{n} p_s y_s \\
\text{subject to}\ & y_s \geq -x^T\xi_s - \alpha \\
& x^T\mu \geq t \\
& \sum_{i=1}^{d} x_i = 1 \\
& x, y \geq 0.
\end{aligned}$$

See [29] for more details on how β-CVaR is linearized for discrete random vectors in this way. These scenario-based problems are modelled using JuMP [10] and solved using Gurobi 7.5 [18].

We are interested in the quality and stability of the solutions that are yielded by our scenario generation method as compared to standard Monte Carlo sampling for a given scenario set size. To this end, in each experiment, for a range of scenario set sizes, we construct 100 scenario sets using sampling and aggregation sampling with conservative and exact risk regions, solve the resulting problems, and calculate the optimality gaps for the solutions that these yield.

Denote by z^* the optimal solution value for problem (P), and by $\tilde{x}$ a solution found by solving a scenario-based approximation. Then the optimality gap of $\tilde{x}$ is given by

$$\beta\text{-CVaR}(-\tilde{x}^T\boldsymbol{\xi}) - z^*$$

where β-CVaR$(-\tilde{x}^T\boldsymbol{\xi})$ calculated using (42).

Results In Fig. 2 are presented the results of these stability tests for two different problems. In the first problem we have $d = 5$ and $\beta = 0.95$. In the second problem we have $d = 10$ and $\beta = 0.99$. For each scenario set size and scenario generation method we have drawn a box plot of the optimality gap of the 100 constructed scenario sets. In the legend of each plot we have given the estimated probability of the aggregation regions, a, and the true optimal value z^* is included in the title. Note that Cons. Agg. sampling and Exact Agg. Sampling are abbrieviations for, respectively, aggregation sampling using the conservative risk region, and aggregation sampling using the exact risk region.

In both cases, both aggregation sampling methods outperform standard Monte Carlo sampling for all scenario set sizes in terms of the size and variability of the calculated optimality gaps. This is because for aggregation sampling we are effectively sampling more scenarios compared with standard Monte Carlo sampling. Aggregation sampling with exact risk regions also significantly outperforms aggregation sampling with conservative risk regions. The improved performance can be expected given that its probability is greater than that of the conservative risk region which gives a greater effective sample size.

9 Conclusions

In this paper we have demonstrated that for stochastic programs which use a tail risk measure, a significant portion of the support of the random vector in the problem may not participate in the calculation of that tail risk measure, whatever feasible decision is used. As a consequence, for scenario-based problems, if we concentrate our scenarios in the region of the distribution which is important to the problem, the risk region, we can represent the uncertainty in our problem in a more parsimonious way, thus reducing the computational burden of solving it.

We have proposed and analyzed two specific methods of scenario generation using risk regions: aggregation sampling and aggregation reduction. Both of these methods were shown to be more effective, in comparison to standard Monte Carlo sampling, as the probability of the non-risk region increases: in essence the higher this probability the more redundancy there is in the original distribution. The application of our methodology relies on having a convenient characterization of a risk region. For portfolio selection problems we derived the exact risk region when returns have an elliptical distribution. However, a characterization of the exact risk region will generally not be possible. Nevertheless, it is sufficient to have a conservative risk region. For stochastic programs with monotonic loss functions, a wide problem class which includes some network design problems, we were able to derive such a region.

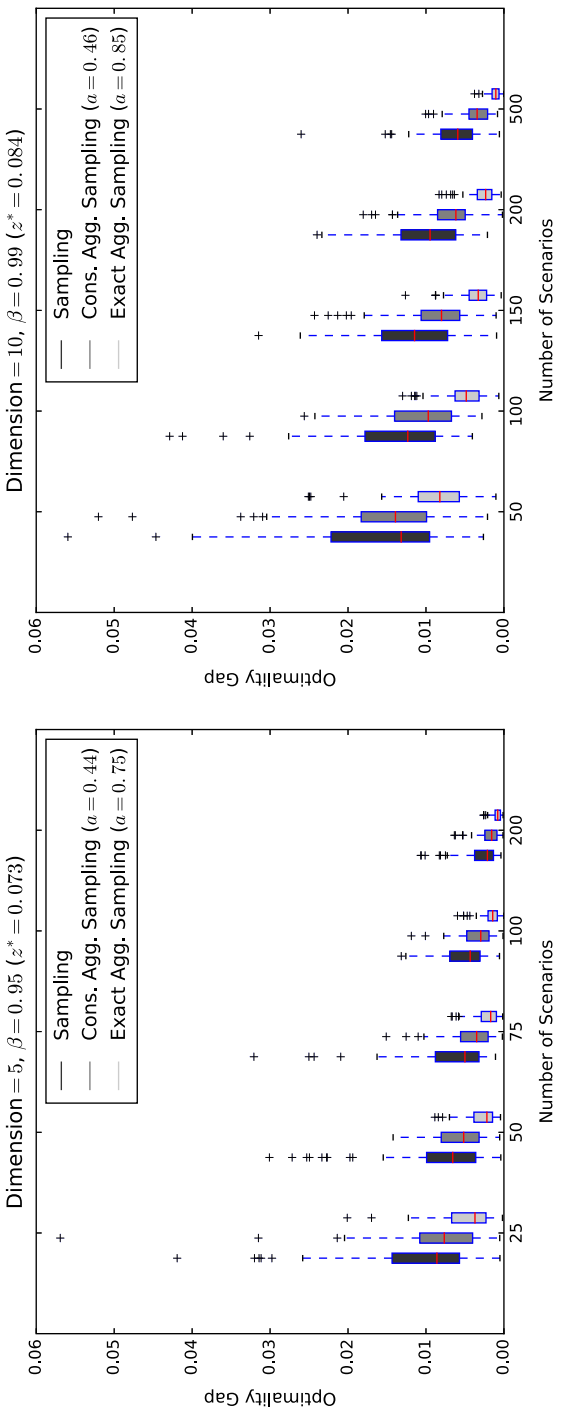


Fig. 2 Error bar plot of optimality gaps yielded by sampling and aggregation sampling

The effectiveness of our methodology depends on the probability of the aggregation region, that is the exact or conservative non-risk region used in our scenario generation algorithms. We observed that for both the stochastic programs with monotonic loss function and portfolio selection problems that this probability tends to zero as the dimension of the random vector in the problem increases. However, in some circumstances this effect is mitigated. We observed that small positive correlations slowed down this convergence for the portfolio selection problem.

We tested the performance of our aggregation sampling algorithm for portfolio selection problems using both the exact non-risk region and the conservative risk region for monotonic loss functions. This demonstrated a significant improvement on the performance of standard Monte Carlo sampling, particularly when an exact non-risk region was used.

The methodology has much potential. For some small to moderately-sized network design problems this methodology could yield much better solutions. In particular the methodology is agnostic to the presence of integer variables, and so could be used to solve difficult mixed integer programs.

In our follow-up paper [15] we demonstrate that our methodology may be applied to more difficult and realistic portfolio selection problems such as those involving integer variables, and for which the asset returns are no longer elliptically distributed. In the same paper we also discuss some of the technical issues involved in applying the method, such as finding the conic hull of the feasible region, and methods of projecting points onto this. We also investigate the use of artificial constraints as a way of making our methodology more effective.

Acknowledgements We would like to thank the reviewers and guest editor for their very thorough feedback which has allowed us to much improve this paper. Thanks also to Burak Buke and David Leslie who also gave feedback on an earlier version of the paper. Finally, we gratefully acknowledge the support of the EPSRC funded EP/H023151/1 STOR-i Centre for Doctoral Training.

A Continuity of distribution and quantile functions

Throughout we use the following set-up: $\mathcal{X} \subset \mathbb{R}^k$ a decision space, $\boldsymbol{\xi}$ a random vector with support $\Xi \subset \mathbb{R}^d$ defined on a probability space $(\Omega, \mathcal{B}, \mathbb{P})$, and a cost function $f : \mathcal{X} \times \mathbb{R}^d \to \mathbb{R}$. The quantity is $f(x, \boldsymbol{\xi})$ is assumed to be measurable for all $x \in \mathcal{X}$. In this appendix we prove a series of technical results related to the continuity of the distribution and quantile functions for $f(x, \boldsymbol{\xi})$. These are required for the proofs in Sect. 5.

The following elementary result concerns the continuity of an expectation function.

Proposition 5 *Suppose for $g : \mathcal{X} \times \Xi \to \mathbb{R}$, and a given $\bar{x} \in \mathcal{X}$ the following holds:*

(i) *$x \mapsto g(x, \boldsymbol{\xi})$ is continuous at $\bar{x}$ with probability 1,*

(ii) *There exists a neighborhood W of $\bar{x}$ and integrable $h : \Xi \to \mathbb{R}$ such that, for all $x \in W$ we have $g(x, \boldsymbol{\xi}) \leq h(\boldsymbol{\xi})$ with probability 1.*

Then, $x \mapsto \mathbb{E}\left[g(x, \boldsymbol{\xi})\right]$ is continuous at $\bar{x}$.

Proof Let $(x_k)_{k=1}^{\infty}$ be some sequence in $\mathcal{X}$ such that $x_k \to \bar{x}$ as $k \to \infty$. Without loss of generality $x_k \in W$ for all $k \in \mathbb{N}$. By assumption (i), almost surely we have $g(x_k, \boldsymbol{\xi}) \to g(\bar{x}, \boldsymbol{\xi})$ as $k \to \infty$. Using assumption (ii) we can apply the Lebesgue theorem of dominated convergence so that:

$$\begin{aligned}\lim_{k\to\infty} \mathbb{E}\left[g(x_k, \boldsymbol{\xi})\right] &= \mathbb{E}\left[\lim_{k\to\infty} g(x_k, \boldsymbol{\xi})\right] \\ &= \mathbb{E}\left[g(\bar{x}, \boldsymbol{\xi})\right]\end{aligned}$$

and hence $x \mapsto \mathbb{E}\left[g(x, \boldsymbol{\xi})\right]$ is continuous at $\bar{x}$. □

The continuity of the distribution function immediately follows from the above proposition.

Corollary 5 *Suppose for a given $\bar{x} \in \mathcal{X}$ that $x \mapsto f(x, \boldsymbol{\xi})$ is continuous with probability 1 at $\bar{x}$, and for $z \in \mathbb{R}$ the distribution function $F_{\bar{x}}$ is continuous at z. Then, $x \mapsto F_x(z)$ is continuous at $\bar{x}$.*

Proof Let $g(x, \boldsymbol{\xi}) = \mathbb{1}_{\{f(x,\boldsymbol{\xi}) \leq z\}}$ so that $F_x(z) = \mathbb{E}\left[g(x, \boldsymbol{\xi})\right]$. The function $g(x, \boldsymbol{\xi})$ is clearly dominated by the integrable function $h(\boldsymbol{\xi}) = 1$. It is therefore enough to show that $x \mapsto g(x, \boldsymbol{\xi})$ is almost surely continuous at $\bar{x}$ as the result will then follow from Proposition 5.

Since $F_{\bar{x}}$ is continuous at z, we must have $\mathbb{P}\left(f(\bar{x}, \boldsymbol{\xi}) = z\right) = 0$. Almost surely, we have that for $\omega \in \Omega$ that $x \mapsto f(x, \boldsymbol{\xi}(\omega))$ is continuous at $\bar{x}$. Let's first assume that $f(\bar{x}, \boldsymbol{\xi}(\omega)) > z$. In this case, there exist some neighborhood V of $\bar{x}$ such that $x \in V \Rightarrow f(x, \boldsymbol{\xi}(\omega)) > z$, which in turn implies $|g(x, \boldsymbol{\xi}) - g(\bar{x}, \boldsymbol{\xi})| = 0$. Hence $x \mapsto g(x, \boldsymbol{\xi}(\omega))$ is continuous at $\bar{x}$. The same argument holds if $f(\bar{x}, \boldsymbol{\xi}(\omega)) < z$. Hence, with probability 1, $x \mapsto g(x, \boldsymbol{\xi})$ is continuous at $\bar{x}$. □

Continuity of the quantile function follows from the continuity of the distribution function but requires that the distribution function is strictly increasing at the required quantile.

Proposition 6 *Suppose for some $\bar{x} \in \mathcal{X}$, and $z = F_{\bar{x}}^{-1}(\beta)$ that the conditions of Corollary 5 hold, and in addition that $F_{\bar{x}}$ is strictly increasing at $F_{\bar{x}}^{-1}(\beta)$, that is for all $\epsilon > 0$*

$$F_{\bar{x}}\left(F_{\bar{x}}^{-1}(\beta) - \epsilon\right) < \beta < F_{\bar{x}}\left(F_{\bar{x}}^{-1}(\beta) + \epsilon\right).$$

Then $x \mapsto F_x^{-1}(\beta)$ is continuous at $\bar{x}$.

Proof Assume $x \mapsto F_x^{-1}(\beta)$ is not continuous at $\bar{x}$. This means there exists $\epsilon > 0$ such that for all neighborhoods W of $\bar{x}$

$$\text{there exists } x' \in W \text{ such that } \left| F_{\bar{x}}^{-1}(\beta) - F_{x'}^{-1}(\beta) \right| > \epsilon.$$

Now set,

$$\gamma := \min\{\beta - F_{\bar{x}}\left(F_{\bar{x}}^{-1}(\beta) - \epsilon\right), F_{\bar{x}}\left(F_{\bar{x}}^{-1}(\beta) + \epsilon\right) - \beta\} > 0$$
$$\text{since } F_{\bar{x}} \text{ strictly increasing at } F_{\bar{x}}^{-1}(\beta).$$

By the continuity of $x \mapsto F_x\left(F_{\bar{x}}^{-1}(\beta)\right)$ at $\bar{x}$ there exists W a neighborhood of $\bar{x}$, such that:

$$x \in W \Longrightarrow \left| F_x\left(F_{\bar{x}}^{-1}(\beta)\right) - F_{\bar{x}}\left(F_{\bar{x}}^{-1}(\beta)\right) \right| < \gamma. \tag{43}$$

But for the x' identified above we have

$$F_{x'}^{-1}(\beta) < F_{\bar{x}}^{-1}(\beta) - \epsilon \quad \text{or} \quad F_{x'}^{-1}(\beta) > F_{\bar{x}}^{-1}(\beta) + \epsilon$$

and so given that $F_{\bar{x}}$ is non-decreasing, and by the definition of γ we must have:

$$\left| F_{\bar{x}}\left(F_{\bar{x}}^{-1}(\beta)\right) - F_{\bar{x}}\left(F_{x'}^{-1}(\beta)\right) \right| \geq \gamma$$

which contradicts (43). □

Recall, that for a sequence of i.i.d. random vectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots$ with the same distribution as $\boldsymbol{\xi}$, we define the sampled distribution function as follows:

$$F_{n,x}(z) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{f(x,\boldsymbol{\xi}_i) \leq z\}}.$$

The final result concerns the continuity of the sampled distribution function.

Lemma 3 *Suppose for $g : \mathcal{X} \times \Xi \to \mathbb{R}$, and $\bar{x} \in \mathcal{X}$ the conditions from Proposition 5 hold. Then for all $\epsilon > 0$ there exists a neighborhood W, of $\bar{x}$, such that with probability 1*

$$\limsup_{n \to \infty} \sup_{x \in W \cap \mathcal{X}} \left| \frac{1}{n} \sum_{i=1}^{n} g(x, \boldsymbol{\xi}_i) - \frac{1}{n} \sum_{i=1}^{n} g(\bar{x}, \boldsymbol{\xi}_i) \right| < \epsilon.$$

In particular, if $x \mapsto f(x, \boldsymbol{\xi})$ is continuous at $\bar{x}$ with probability 1 and $F_{\bar{x}}$ is continuous at $z \in \mathbb{R}$ then for all $\epsilon > 0$ there exists a neighborhood W, of $\bar{x}$ such that with probability 1

$$\limsup_{n\to\infty} \sup_{x\in W\cap\mathcal{X}} \left| F_{n,x}(z) - F_{n,\bar{x}}(z) \right| < \epsilon. \tag{44}$$

Proof Fix $\bar{x} \in \mathcal{X}$, and $\epsilon > 0$. Let $(\gamma_k)_{k=1}^{\infty}$ be any sequence of positive numbers converging to zero and define

$$V_k := \{x \in \mathcal{X} : \|x - \bar{x}\| \leq \gamma_k\},$$
$$\delta_k(\boldsymbol{\xi}) := \sup_{x\in V_k} |g(x,\boldsymbol{\xi}) - g(\bar{x},\boldsymbol{\xi})|.$$

Note first that the quantity $\delta_k(\boldsymbol{\xi})$ is Lebesgue measurable (see [34, Theorem 7.37] for instance). By assumption (i) of Proposition 5 the mapping $x \mapsto g(x,\boldsymbol{\xi})$ is continuous at $\bar{x}$ with probability 1, hence $\delta_k(\boldsymbol{\xi}) \to 0$ almost surely as $k \to \infty$. Now, since $|g(x,\boldsymbol{\xi})| \leq h(\boldsymbol{\xi})$ we must have $|\delta_k(\boldsymbol{\xi})| \leq 2h(\boldsymbol{\xi})$, therefore, by the Lebesgue dominated convergence theorem, we have that

$$\lim_{k\to\infty} \mathbb{E}\left[\delta_k(\boldsymbol{\xi})\right] = \mathbb{E}\left[\lim_{k\to\infty} \delta_k(\boldsymbol{\xi})\right] = 0. \tag{45}$$

Note also that

$$\sup_{x\in V_k} \left| \frac{1}{n}\sum_{i=1}^{n} g(x,\boldsymbol{\xi}_i) - \frac{1}{n}\sum_{i=1}^{n} g(\bar{x},\boldsymbol{\xi}_i) \right| \leq \frac{1}{n}\sum_{i=1}^{n} \sup_{x\in V_k} \left| g(x,\boldsymbol{\xi}_i) - g(\bar{x},\boldsymbol{\xi}_i) \right|$$

and so

$$\sup_{x\in V_k} \left| \frac{1}{n}\sum_{i=1}^{n} g(x,\boldsymbol{\xi}_i) - \frac{1}{n}\sum_{i=1}^{n} g(\bar{x},\boldsymbol{\xi}_i) \right| \leq \frac{1}{n}\sum_{i=1}^{n} \delta_k(\boldsymbol{\xi}_i).$$

Since the sequence of random vectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots$ is i.i.d. we have by the strong law of large numbers that the right-hand side of (46) converges with probability 1 to $\mathbb{E}\left[\delta_k(\boldsymbol{\xi})\right]$ as $n \to \infty$. Hence, with probability 1

$$\limsup_{n\to\infty} \sup_{x\in V_k} \left| \frac{1}{n}\sum_{i=1}^{n} g(x,\boldsymbol{\xi}_i) - \frac{1}{n}\sum_{i=1}^{n} g(\bar{x},\boldsymbol{\xi}_i) \right| \leq \mathbb{E}\left[\delta_k(\boldsymbol{\xi})\right]. \tag{46}$$

By (45) we can pick $k \in \mathbb{N}$ such that $\mathbb{E}\left[\delta_k(\boldsymbol{\xi})\right] < \epsilon$ and so setting $W = V_k$ we have by (46) with probability 1

$$\limsup_{n\to\infty} \sup_{x\in W\cap\mathcal{X}} \left| \frac{1}{n}\sum_{i=1}^{n} g(x,\boldsymbol{\xi}_i) - \frac{1}{n}\sum_{i=1}^{n} g(\bar{x},\boldsymbol{\xi}_i) \right| < \epsilon.$$

The result (44) follows immediately as the special case $g(x,\boldsymbol{\xi}) = \mathbb{1}_{\{f(x,\boldsymbol{\xi})\leq z\}}$. □

B Convex cone results

The results in this appendix relate to the characterization of the non-risk region for the portfolio selection problem with elliptically distributed returns.

The following two propositions give properties about projections onto convex cones which are required in the proof of the main results of this appendix.

Proposition 7 *Suppose $K \subset \mathbb{R}^d$ is a convex cone. Then, for all $\xi \in \mathbb{R}^d$:*

$$p_K(\xi)^T (\xi - p_K(\xi)) = 0.$$

Proof First note that we must have $p_K(\xi)^T \xi \geq 0$. If this is not the the case then

$$\|\xi - p_K(\xi)\|^2 = \|p_K(\xi)\|^2 - 2p_K(\xi)^T \xi + \|\xi\|^2 > \|\xi\|^2 = \|\xi - 0\|^2$$

which contradicts the definition of $p_K(\xi)$ since $0 \in K$. Now assume that $p_K(\xi)^T (\xi - p_K(\xi)) \neq 0$, and set $\tilde{x} = \frac{p_K(\xi)^T \xi}{\|p_K(\xi)\|^2} p_K(\xi) \in K$. Now,

$$p_K(\xi)^T (\tilde{x} - \xi) = p_K^T \xi - p_K^T \xi = 0.$$

By assumption $p_k^T \xi \neq \|p_K(\xi))\|^2$, and so $\tilde{x} \neq p_K(\xi)$, hence

$$\begin{aligned}
&\|p_K(\xi) - \xi\|^2 \\
&\quad = \|(p_K(\xi) - \tilde{x}) + (\tilde{x} - \xi)\|^2 \\
&\quad = \|(p_K(\xi) - \tilde{x})\|^2 - 2\underbrace{(p_K(\xi) - \tilde{x})^T (\tilde{x} - \xi)}_{=0} + \|(\tilde{x} - \xi)\|^2 > \|(\tilde{x} - \xi)\|^2
\end{aligned}$$

which, again, contradictions the definition of $p_K(\xi)$ since $\tilde{x} \in K$. □

Proposition 8 *Suppose $K \subset \mathbb{R}^d$ be a convex cone and $x \in K$. Then for any $\xi \in \mathbb{R}^d$*

$$x^T \xi \leq x^T p_K(\xi).$$

Proof The result holds trivially if $\xi \in K$ so we assume $\xi \notin K$. Assume there exists $\tilde{x} \in K$ such that $\tilde{x}^T \xi > \tilde{x}^T p_K(\xi)$. For all $0 \leq \lambda \leq 1$ we have $\lambda x + (1-\lambda) p_K(\xi) \in K$. Now,

$$\begin{aligned}
&\|(\lambda \tilde{x} + (1 - \lambda) p_K(\xi)) - \xi\|^2 - \|\xi - p_K(\xi)\|^2 \\
&\quad = \lambda^2 \|\tilde{x} - p_K(\xi)\|^2 + 2\lambda (\tilde{x} - p_K(\xi))^T (p_K(\xi) - \xi) \\
&\quad = \lambda^2 \|\tilde{x} - p_K(\xi)\|^2 - 2\lambda \underbrace{\tilde{x}^T (\xi - p_K(\xi))}_{>0 \text{ by assumption}}.
\end{aligned}$$

That is, for $0 < \lambda < \frac{\tilde{x}^T (\xi - p_K(\xi))}{2\|p_K(\xi) - \tilde{x}\|}$ we have $\|\lambda \tilde{x} + (1 - \lambda) p_K(\xi) - \xi\| < \|\xi - p_K(\xi)\|$ which contradicts the definition of $p_K(\xi)$. □

The next two results describe the non-risk region for the portfolio selection problem with elliptically distributed returns when $\mathcal{X}$ is a convex set. The first describes the exact non-risk region for elliptically distributed returns in the case $P = I$, and the second generalizes the result to any non-singular matrix.

Theorem 6 *Suppose* $\mathcal{X} \subset \mathbb{R}^d$ *is convex and* $\mu \in \mathbb{R}^d$, *and let* $\mathcal{A} := \{\xi \in \mathbb{R}^d : x^T(\xi - \mu) < \|x\|\ \alpha\ \forall x \in \mathcal{X}\}$ *and* $\mathcal{B} := \{\xi \in \mathbb{R}^d : \|p_K(\xi - \mu)\| < \alpha\}$ *where* $K = \text{conic}\,(\mathcal{X})$. *Then,* $\mathcal{A} = \mathcal{B}$.

Proof $(\mathcal{B} \subseteq \mathcal{A})$
Suppose $\xi \in \mathcal{B}$ and let $x \in \mathcal{X}$, then $x \in K$ and so

$$\begin{aligned} x^T(\xi - \mu) &\leq x^T p_K(\xi - \mu) \qquad \text{by Proposition 8} \\ &\leq \|x\|\ \|p_K(\xi) - \mu\| \qquad \text{by the Cauchy–Schwartz inequality} \\ &< \|x\|\ \alpha \qquad \text{since } \xi \in \mathcal{B}. \end{aligned}$$

Hence $\xi \in \mathcal{A}$.
$(\mathcal{A} \subseteq \mathcal{B})$
Suppose $\xi \notin \mathcal{B}$ and set $x = p_K(\xi - \mu) \in K$. Now,

$$\begin{aligned} x^T(\xi - \mu) &= p_K(\xi - \mu)^T(\xi - \mu) \\ &= p_K(\xi - \mu)^T p_K(\xi - \mu) + p_K(\xi - \mu)^T\left((\xi - \mu) - p_K(\xi - \mu)\right) \\ &= p_K(\xi - \mu)^T p_K(\xi - \mu) \qquad \text{by Proposition 7} \\ &\geq \|x\|\ \alpha \qquad \text{since } \xi \notin \mathcal{B}. \end{aligned}$$

Since $\mathcal{X}$ is convex we have $x = \lambda\bar{x}$ for some $\bar{x} \in \mathcal{X}$ and so we must also have $\bar{x}^T\xi \geq \|\bar{x}\|\,\alpha$, hence $\xi \notin \mathcal{A}$. □

Corollary 6 *Suppose* $\mathcal{X}$ *is convex, and* $P \in \mathbb{R}^{d\times d}$ *is a non-singular matrix. Let,* $\mathcal{A} := \{\xi \in \mathbb{R}^d : x^T(\xi - \mu) < \|Px\|\ \alpha\ \forall x \in \mathcal{X}\}$ *and* $\mathcal{B} := P^T\left(\{\tilde{\xi} \in \mathbb{R}^d : \left\|p_{K'}(\tilde{\xi} - \tilde{\mu})\right\| < \alpha\}\right)$ *where* $\tilde{\mu} = (P^T)^{-1}\mu$, $K' = PK$, *and* $K = \text{conic}\,(\mathcal{X})$. *Then,* $\mathcal{A} = \mathcal{B}$.

Proof First note that $K' = PK = P\,\text{conic}\,(\mathcal{X}) = \text{conic}\,(P\mathcal{X})$. Now,

$$\begin{aligned} \mathcal{B} &= P^T\left(\{\tilde{\xi} \in \mathbb{R}^d : \left\|p_{K'}(\tilde{\xi} - \tilde{\mu})\right\| < \alpha\}\right) \\ &= P^T\left(\{\tilde{\xi} \in \mathbb{R}^d : \tilde{x}^T(\tilde{\xi} - \tilde{\mu}) < \|\tilde{x}\|\ \alpha\ \forall \tilde{x} \in P\mathcal{X}\}\right) \qquad \text{by Theorem 6} \\ &= \{\xi \in \mathbb{R}^d : \tilde{x}^T\left((P^T)^{-1}\xi - \tilde{\mu}\right) < \sqrt{\tilde{x}^T\tilde{x}}\alpha\ \forall \tilde{x} \in P\mathcal{X}\} \\ &= \{\xi \in \mathbb{R}^d : x^T P^T\left((P^T)^{-1}\xi - (P^T)^{-1}\mu\right) < \|Px\|\ \alpha\ \forall x \in \mathcal{X}\} \\ &= \{\xi \in \mathbb{R}^d : x^T(\xi - \mu) < \|Px\|\ \alpha\ \forall x \in K\} = \mathcal{A} \end{aligned}$$

□

References

1. Acary, Vincent, Pérignon, Franck: Siconos: a software platform for modeling, simulation, analysis and control of nonsmooth dynamical systems. Simul. News Eur. **17**(3/4), 19–26 (2007)
2. Acerbi, C., Tasche, D.: On the coherence of expected shortfall. J. Bank. Finance **26**(7), 1487–1503 (2002)
3. Artzner, P., Delbaen, F., Eber, J., Heath, D.: Coherent measures of risk. Math. Finance **9**(3), 203–228 (1999)
4. Barrera, J., Homem-de Mello, T., Moreno, E., Pagnoncelli, B.K., Canessa, G.: Chance-constrained problems and rare events: an importance sampling approach. Math. Program. **157**(1), 153–189 (2016)
5. Bieniek, M.: A note on the facility location problem with stochastic demands. Omega **55**, 53–60 (2015)
6. Billingsley, P.: Probability and Measure, 3rd edn. Wiley, New York (1995)
7. Birge, J.R., Louveaux, F.: Introduction to Stochastic Programming. Springer, New York (1997)
8. Dantzig, G.B., Glynn, P.W.: Parallel processors for planning under uncertainty. Ann. Oper. Res. **22**(1), 1–21 (1990)
9. Doan, X.V., Li, X., Natarajan, K.: Robustness to dependency in portfolio optimization using overlapping marginals. Oper. Res. **63**(6), 1468–1488 (2015)
10. Dunning, I., Huchette, J., Lubin, M.: Jump: a modeling language for mathematical optimization. SIAM Rev. **59**(2), 295–320 (2017)
11. Dupačová, J.: Uncertainties in minimax stochastic programs. Optimization **60**(10–11), 1235–1250 (2011)
12. Dupačová, J., Gröwe-Kuska, N., Römisch, W.: Scenario reduction in stochastic programming: an approach using probability metrics. Math. Program. **95**(3), 493–511 (2003)
13. Fairbrother, J: Distributions modelling FTSE100 stock returns (2017). https://dx.doi.org/10.17635/lancaster/researchdata/158. Accessed 24 Nov 2019
14. Fairbrother, J: TailRiskScenGen.jl: a julia package for scenario generation for stochastic programs with tail risk measure (2017). https://github.com/STOR-i/TailRiskScenGen.jl. Accessed 24 Nov 2019
15. Fairbrother, J., Turner, A., Wallace, S.W.: Scenario generation for single-period portfolio selection problems with tail risk measures: coping with high dimensions and integer variables. INFORMS J. Comput. **30**(3), 472–491 (2018)
16. Fang, K.T., Kotz, S., Ng, K.W.: Symmetric Multivariate and Related Distributions (Chapman & Hall/CRC Monographs on Statistics & Applied Probability), vol. 11. Chapman and Hall, London (1989)
17. García-Bertrand, R., Mínguez, R.: Iterative scenario based reduction technique for stochastic optimization using conditional value-at-risk. Optim. Eng. **15**(2), 355–380 (2014)
18. Gurobi Optimization Inc. Gurobi optimizer reference manual (2016)
19. Heitsch, H., Römisch, W.: Scenario tree reduction for multistage stochastic programs. CMS **6**(2), 117–133 (2009)
20. Higle, J.L.: Variance reduction and objective function evaluation in stochastic linear programs. INFORMS J. Comput. **10**(2), 236–247 (1998)
21. Jorion, P.: Value at Risk: The New Benchmark for Controlling Market Risk. Irwin Professional, Norman (1996)
22. King, A.J., Rockafellar, R.T.: Asymptotic theory for solutions in statistical estimation and stochastic programming. Math. Oper. Res. **18**(1), 148–162 (1993)
23. Kozmík, V., Morton, D.P.: Evaluating policies in risk-averse multi-stage stochastic programming. Math. Program. **152**(1), 275–300 (2015)
24. Linderoth, J., Shapiro, A., Wright, S.: The empirical behavior of sampling methods for stochastic programming. Ann. Oper. Res. **142**(1), 215–241 (2006)
25. Mak, W.K., Morton, D.P., Wood, R.K.: Monte Carlo bounding techniques for determining solution quality in stochastic programs. Oper. Res. Lett. **24**, 47–56 (1999)
26. Markowitz, H.M.: Portfolio selection. J. Finance **7**, 77–91 (1952)
27. Ogryczak, W., Ruszczyński, A.: Dual stochastic dominance and related mean-risk models. SIAM J. Optim. **13**(1), 60–78 (2002). (electronic)
28. Pflug, G.C.: Scenario tree generation for multiperiod financial optimization by optimal discretization. Math. Program. **89**(2), 251–271 (2001)
29. Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. J. Risk **2**(3), 21–41 (2000)

30. Rockafellar, R.T., Uryasev, S.: Conditional value-at-risk for general loss distributions. J. Bank. Finance **26**(7), 1443–1471 (2002)
31. Santoso, T., Ahmed, S., Goetschalckx, M., Shapiro, A.: A stochastic programming approach for supply chain network design under uncertainty. Eur. J. Oper. Res. **167**(1), 96–115 (2005)
32. Serfling, R.J.: Approximation Theorems of Mathematical Statistics. Wiley Series in Probability and Statistics—Applied Probability and Statistics Section Series. Wiley, Hoboken (1980)
33. Shapiro, A.: Monte Carlo sampling methods. In Ruszczyński, A., Shapiro, A. (eds.) Stochastic Programming, volume 10 of Handbooks in Operations Research and Management Science, chapter 6. Elsevier Science B.V., Amsterdam, pp. 353–425 (2003)
34. Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming: Modeling and Theory, volume 9 of MPS-SIAM Series on Optimization. SIAM, Philadelphia (2009)
35. Tasche, D.: Expected shortfall and beyond. J. Bank. Finance **26**(7), 1519–1533 (2002)
36. Ujvari, M.: On the projection onto a finitely generated cone. Acta Cybern. **22**. https://doi.org/10.14232/actacyb.22.3.2016.7
37. Wächter, A., Biegler, L.T.: On the implementation of a primal–dual interior point filter line search algorithm for large-scale nonlinear programming. Math. Program. **106**(1), 25–57 (2006)
38. Wiesemann, W., Kuhn, D., Sim, M.: Distributionally robust convex optimization. Oper. Res. **62**(6), 1358–1376 (2014)
39. Žáčková, J.: On minimax solutions of stochastic linear programming problems. Časopis pro pěstování matematiky **91**(4), 423–430 (1966)