Stochastics and Statistics

Tight tail probability bounds for distribution-free decision making

Ernst Roos¹,², Ruud Brekelmans¹, Wouter van Eekelen¹, Dick den Hertog¹, Johan S. H. van Leeuwaarden¹

¹Tilburg University, Department of Econometrics and Operations Research, Warandelaan 2, 5037 AB Tilburg, the Netherlands
²Amsterdam Business School, University of Amsterdam, Plantage Muidergracht 12, 1018 TV Amsterdam, the Netherlands

A R T I C L E  I N F O

Keywords:
Applied probability
Chebyshev inequality
Distributionally robust optimization
Chance constraints

A B S T R A C T

Chebyshev’s inequality provides an upper bound on the tail probability of a random variable based on its mean and variance. While tight, the inequality has been criticized for only being attained by pathological distributions that abuse the unboundedness of the underlying support and are not considered realistic in many applications. We provide alternative tight lower and upper bounds on the tail probability given a bounded support, mean and mean absolute deviation of the random variable. We obtain these bounds as exact solutions to semi-infinite linear programs. We apply the bounds for distribution-free analysis of the newsvendor model, stop-loss reinsurance and a problem from radiotherapy optimization with an ambiguous chance constraint.

© 2021 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

1. Introduction

Chebyshev’s inequality provides an upper bound on the tail probability of a random variable using only its first two moments (Bienaymé, 1853; Chebyshev, 1867). Let the ambiguity set \( \mathcal{P}_{\mu, \sigma} \) contain all distributions with a given mean \( \mu \) and variance \( \sigma^2 \), and let the random variable \( X \) follow some distribution \( P \in \mathcal{P}_{\mu, \sigma} \). Chebyshev's inequality (the one-sided version also known as Cantelli’s inequality) then follows from the worst-case distribution that solves the optimization problem

\[
\mathbb{P}(X \geq t) \leq \sup_{X \sim P \in \mathcal{P}_{\mu, \sigma}} \mathbb{E}[1[X \geq t]] = \frac{\sigma^2}{\sigma^2 + (t - \mu)^2}.
\]

This inequality is tight, meaning it cannot be improved in general. However, Chebyshev’s inequality can be criticized for only being attained by pathological distributions that abuse the unboundedness of the underlying support. Indeed, the worst-case distribution takes only values on the points \( \mu - \sigma^2/(t - \mu) \) and \( t \) (with probabilities \( (t - \mu)^2/(\sigma^2 + (t - \mu)^2) \) and \( \sigma^2/(\sigma^2 + (t - \mu)^2) \), resp.), which can be regarded unrealistic (Van Parys, Goulat, & Kuhn, 2016). A variant of the Chebyshev inequality that was already considered in Gauss (1821) restricts the distributions it considers to be unimodal. This yields an improvement by a factor \( \frac{4}{9} \) over the classical Chebyshev inequality. This idea of including unimodality has been extended to the multivariate case recently as well (Van Parys et al., 2016).

Multivariate generalizations of Chebyshev’s inequality have also been studied. In Bertsimas & Popescu (2005) and Vandenberghe, Boyd, & Comanor (2007) generalizations are studied through formulating a convex optimization problem, given that the prescribed confidence region can be described by polynomial or linear and quadratic inequalities, respectively. In Grechuk, Molyboha, & Zabaranke (2010) on the other hand, closed-form variants of Chebyshev's inequality are provided for different dispersion measures than the variance. Generalized versions of Chebyshev's inequality for products of random variables that focus on a one-sided inequality have also received some attention recently (Rujeerapaiboon, Kuhn, & Wiesemann, 2018).

All the above mentioned inequalities, however, still assume an unbounded support. In many practical applications some information on the minimum and maximum of uncertain parameters is known. This is particularly true for OR applications that consider uncertain parameters that are known to be nonnegative, such as inventory management, service operations, appointment scheduling and pricing mechanisms. De Schepper & Heijnen (1995) derive tail probability bounds that incorporate the upper bound of the random variable's range. These bounds are attained by discrete distributions, supported on two or three atoms.

Next to restricting the support, a second potential improvement of Chebychev's inequality concerns robustness for outliers.
Whereas the (sample) variance is greatly influenced by outliers, the mean absolute deviation (MAD) is less sensitive for large deviations from the mean, and hence a potentially more robust measure of statistical dispersion in data. We therefore propose to replace variance with MAD. Using the MAD comes with additional advantages. We show that the set of extremal distributions for which the derived tail bounds are tight is more varied than a single pathological distribution: it consists of an infinite number of mixed distributions instead. Second, because the MAD is a linear function, it allows for elegant closed-form bounds, a feature we shall leverage when applying the bounds to domain-specific OR questions.

In obtaining the robust tail bounds, we need to solve

$$\sup_{X \sim P_{(\mu,b,d)}} \mathbb{E}_P[1 \{X \geq t\}]$$

(2)

with $P_{(\mu,b,d)}$ the ambiguity set that contains all distributions with a given mean $\mu$, support $[0,b]$ and mean absolute deviation $d$. Optimization problem (2) is a semi-infinite linear optimization problem (LP) that is reminiscent of those arising in moment problems, and typically does not allow for an analytic (closed-form) solution. Using the MAD-based ambiguity set $P_{(\mu,b,d)}$, the dual program to (2) can be solved explicitly. While comparable dual programs are often solvable as semidefinite or second-order conic programs (see, e.g., Das, Dhara, & Natarajan, 2021; Natarajan, Sim, & Uichanco, 2018; Natarajan & Zhou, 2007; Perakis & Roels, 2008; Xin & Goldberg, 2021), analytic solutions as in our case are typically hard to attain.

The solution of (2) gives a tight upper bound on the tail probability of all random variables with a given bounded support, mean and MAD. This new robust bound is of a similar simplicity and generality as the original Chebyshev inequality, and can therefore be used in various applications. The worst-case distribution that solves (2) is, however, more complicated than the two-point distribution of the Chebyshev inequality, and is a mixed distribution with up to three discrete parts and one continuous part. We also derive a number of additional tail probability bounds: the tight lower bound under $P_{(\mu,b,d)}$ ambiguity, the tight upper and lower bounds under $P_{(\mu,b,d)}$ ambiguity, where we also condition on the skewness $\mathbb{P}(X \geq \mu) = \beta$, and the tight upper bound under $P_{(\mu,b,d)}$ ambiguity, where we condition on the median $m$ and the mean absolute deviation from the median $d_m$.

Recent advances in Distributionally Robust Optimization (DRO) also exploit MAD-based ambiguity sets to obtain closed-form expressions for stochastic quantities such as the minimum and maximum expectation of a convex function (Ghosal & Wiesemann, 2020; Postek, Ben-Tal, Den Hertog, & Meloen, 2018). These closed-form expressions are then used to solve minmax and maxmin optimization problems that arise naturally in decision making under uncertainty. Postek et al. (2018) specifically use results from Ben-Tal & Hochman (1972) on tight upper and lower bounds on the expectation of a convex function of a random variable. This paper presents the first closed-form solution for the combination of $P_{(\mu,b,d)}$ (or $P_{(\mu,b,d,\beta)}$) constraints and a non-convex objective function. This proof method is not restricted to the indicator function as in (2), and could potentially work for a much larger class of (measurable) functions.

We apply the robust bounds for distribution-free analysis of three applications that can be subjected to minmax or maxmin optimization. We start with the newsvendor model, the basic single-period inventory model that searches for the optimal order quantity in view of overage and underage costs. The second application is stop-loss reinsurance, in which an insurance company faces a claim which it passes up to a predefined level, while the reinsurance company covers the remainder. We study this problem from both the insurer’s and reinsurer’s perspective, the latter of which requires an extension of our tail probability bound. Last, we study a continuous optimization problem from radiotherapy optimization with an ambiguous chance constraint. Application of the derived tail probability bound yields a computationally tractable convex reformulation that can be solved with traditional methods. The three applications illustrate different aspects of the derived tail probability bound and the primal-dual proof used to obtain it. First, the newsvendor example is a direct application of the bound to a classical OR problem. Second, the stop-loss reinsurance application illustrates how the primal-dual proof technique can be extended to more complex functionals than the tail probability. Third, the radiotherapy optimization example highlights the connection of our result to the field of distributionally robust optimization. We should say that the models have been chosen somewhat arbitrarily, and there are many other OR questions where tail probability bounds under mean-MAD constraints can prove useful.

The paper is organized as follows. We present the tail probability bounds in Section 2 and the three applications in Section 3. Section 4 presents some conclusions and several directions for future research.

### 2. Tail probability bounds

In this section we derive novel bounds for the probability $\mathbb{P}(X \geq t)$, where $X$ is a random variable with given support, mean and MAD. We obtain the bounds by solving the semi-infinite LP

$$\sup_{P \in P_{(\mu,b,d)}} \int \chi_{[t,\infty)} \ dP(x),$$

(3)

where we maximize over a set of probability measures with the stated characteristics, i.e.,

$$P_{(\mu,b,d)} = \{P : \mathcal{B} \rightarrow [0,1] \mid \mathbb{P}(X \in [0,b]) = 1, \mathbb{E}_P[X] = \mu, \mathbb{E}_P[|X-\mu|] = d\}$$

(4)

with $\mathcal{B}$ the Borel $\sigma$-algebra of the closed set $[0,b]$, and $\mu, b, d \in \mathbb{R}_+$ are parameters that describe all known properties of the distribution. We solve the semi-infinite LP in Section 2.1. In Section 2.2 we derive more bounds, based on different ambiguity sets. We compare the novel bounds with some existing bounds in Section 2.3.

#### 2.1. Tight lower and upper bounds

Since $\mathbb{P}$ is a probability measure it should satisfy the constraint $\int_{[t,\infty)} \ d\mathbb{P}(x) = 1$. Moreover, this probability measure should satisfy the mean and MAD constraints $\int_{[t,\infty)} x \ d\mathbb{P}(x) = \mu$ and $\int_{[t,\infty)} |x-\mu| \ d\mathbb{P}(x) = d$. Under these constraints, we solve the semi-infinite linear program (3), which gives our first main result.

**Theorem 1.** Consider a random variable $X$ with a distribution $\mathbb{P}$ in $P_{(\mu,b,d)}$. Then,

$$\sup_{P \in P_{(\mu,b,d)}} \mathbb{P}(X \geq t) = \sup_{P \in P_{(\mu,b,d)}} \mathbb{P}(X > t)$$

$$= \begin{cases} 1 - \frac{d}{2d - d_0}, & t \in [0, \tau_1], \\ 1 - \frac{d}{2d - d_0}, & t \in [\tau_1, \mu], \\ \frac{d_0}{2d - d_0}, & t \in [\mu, \tau_2], \\ \frac{d_0}{2d - d_0}, & t \in [\tau_2, b], \end{cases}$$

(5)

with $\tau_1$ and $\tau_2$ given by

$$\tau_1 = \mu - \frac{d(b - \mu)}{2d - d_0}, \quad \tau_2 = \mu + \frac{d\mu}{2d - d_0}.$$

**Proof.** Let $\mathcal{M}^+$ be the set of non-negative measures defined on the measurable space $([0,b], \mathcal{B})$. We need to solve
Due to dual feasibility, we must have that $\lambda_0^* + \lambda_1^* \mu + \lambda_2^* \xi_0 d - 1[x \geq t] \geq 0$ pointwise for each $x \in [0, b]$. This inequality combined with Eq. (8) is also known as the complementary slackness relation in (semi-infinite) linear programming. Complementary slackness implies that the worst-case probability distribution is supported on the points where the dual function $F^* (x) = \lambda_0^* + \lambda_1^* x + \lambda_2^* [x - \mu]$ coincides with the indicator function $\mathbb{1} [x \geq t]$. For scenario 1a we have one unique option, a three-point distribution on $[0, t, b]$. The corresponding optimal probabilities of (6) follow from solving $p_0 + p_1 + p_2 = 1$, $p_t + p_0 b = \mu$, $\mu - p_t - p_2 (b - \mu) = d$. This gives

$$p_0 = 1 - p_t - p_b, \quad p_t = \frac{\mu}{t} - \frac{bd}{2 (b - \mu)}, \quad p_b = \frac{d}{2 (b - \mu)}.$$  

(9)

Since strong duality holds as the primal and dual optimal values agree, (9) is the optimal solution.

Scenario 1b implies $F (0) = F (t) = F (b) = 1$ and hence $\lambda_0 = 1 = \lambda_2$ with objective value 1. One feasible primal solution is e.g. $p_1 = \frac{d}{2 (t - \mu)}$, $p_b = \frac{d}{2 (b - \mu)}$, $p_2 = 1 - p_t - p_b$, with objective 1. Note that this solution is not a unique optimum, as the dual function $F_{1b}^* (x)$ coincides with $\mathbb{1} [x \geq t]$ on the entire interval $[t, b]$. Therefore, one can construct an arbitrary (discrete, continuous or mixed) probability distribution with support on the interval $[t, b]$, which then serves as the worst-case distribution, as long as the mean and MAD constraints are satisfied.

Scenario 2a implies $F (0) = F (\mu) = 0$, $F (t) = 1$, which gives

$$\lambda_1 = \lambda_2 = \frac{1}{2 (t - \mu)}, \quad \lambda_0 = -\frac{\mu}{2 (t - \mu)}.$$  

and objective value

$$\lambda_0 + \lambda_1 \mu + \lambda_2 d = \frac{d}{2 (t - \mu)}.$$  

Solving the optimal probabilities of (6), where we take $[0, \mu, t]$ for the support of the worst-case distribution, indeed confirms that $p_1 = \frac{d}{2 (t - \mu)}$.

Scenario 2b gives $F (0) = 0$, $F (\mu) = F (b) = 1$, which results in

$$\lambda_0 = \frac{1}{2}, \quad \lambda_1 = \frac{1}{2 \mu}, \quad \lambda_2 = \frac{1}{2 \mu}.$$  

and dual objective value

$$\lambda_0 + \lambda_1 \mu + \lambda_2 d = 1 - \frac{d}{2 \mu}.$$  

Solving (6) with support $[0, t, b]$ confirms that $p_0 = \frac{d}{2 \mu}$.

The proof is then completed by finding the minimum for each scenario and determining the values of $t_1$ and $t_2$ for scenarios 1 and 2, respectively. We remark that the proof is identical for the strict inequality. Because the majorant is a continuous function,
it is irrelevant whether the indicator function that is majorized is lower or upper semi-continuous.

We mention some noteworthy characteristics of the bound in Theorem 1. The bound is continuous in \( t = \mu \). If the support is symmetric around \( \mu \), then the worst-case probability is at least \( 1/2 \) for \( t \in [0, \mu] \). The upper bound for \( t \in [\mu, b] \) is increasing for \( d \leq 2\mu(t - \mu)/t \) and decreasing for larger values of \( d \). This last observation is interesting as one might anticipate the bound to increase with \( \text{MAD} \). This also implies that when \( \text{MAD} \) is unknown, the worst-case probability bound on the upper support and mean is given by the result of Theorem 1 for \( d = 2\mu(t - \mu)/t \). This indeed returns Markov’s inequality. We also mention that the support information \( [0, b] \) can easily be extended to \([a, b]\) with \( a \in \mathbb{R} \) by shifting the distribution accordingly. The tail bounds for the second and third interval then change into

\[
\begin{align*}
\frac{\mu - a}{t - a} - \frac{d(b - t)}{2(t - a)(b - \mu)} \quad \text{and} \quad 1 - \frac{d}{2(\mu - a)},
\end{align*}
\]

respectively. Similarly, the result can be adapted to a support that is only bounded from below or above. For such supports, one of the cases in (5) disappears. Specifically, when the support of \( X \) is given by \([0, \infty)\), it follows that \( t_1 = \mu \), while for the support \((-\infty, b]\), it follows that \( t_2 = \mu \).

For a tight lower bound on \( \mathbb{P}(X > t) \), we can use the results and the remark above on a slightly altered version of the input. The idea is formalized in the following result:

**Corollary 1.** For a random variable \( X \) with a distribution \( \mathbb{P} \) in \( \mathcal{P}_{\mu}(\mu, b, d) \),

\[
\inf_{P_{\mu}(\mu, b, d)} \mathbb{P}(X \geq t) = \inf_{P_{\mu}(\mu, b, d)} \mathbb{P}(X > t) = \begin{cases} 
1 - \frac{d}{2\mu - t}, & t \in [0, t_1], \\
\frac{d}{2\mu - t}, & t \in [t_1, b], \\
\frac{d}{2\mu - t} + \frac{dt}{2\mu - b - \mu}, & t \in [\mu, b], \\
0, & t \in [t_2, b]
\end{cases}
\]

(10)

with

\[
\tau_1 = \mu - \frac{d(b - \mu)}{2(b - \mu) - d}, \quad \tau_2 = \mu + \frac{d\mu}{2\mu - d}.
\]

**Proof.** We reformulate the infimum as follows:

\[
\inf_{P_{\mu}(\mu, b, d)} \mathbb{P}(X > t) = 1 - \sup_{P_{\mu}(\mu, b, d)} \mathbb{P}(X \leq t)
\]

(11)

where

\[
\mathcal{P}_{\mu}(\mu, b, d) = \{ P : \mathbb{B} \to [0, 1] | \mathbb{P}(X \in [a, b]) = 1, E_\mu[X] = \mu, E_\mu[|X - \mu|] = d \},
\]

\( \mathbb{B} = \{ 0, b \} \), and \( \mathbb{F} = \{ [0, b] \} \). This transformation essentially flips the support around \( \mu \) such that \( \mathbb{P}(X = b - \mu) = \mathbb{P}(X = \mu - b) \). Therefore, the maximum probability below \( t \) is equal to the maximum probability above the flipped threshold \( \bar{t} \). Plugging in the results from Theorem 1 for \( t \in [a, b] \) then yields (11). Similarly, the result for \( \inf_{P_{\mu}(\mu, b, d)} \mathbb{P}(X \geq t) \) can be obtained.

We now describe in more detail the worst-case distributions that are revealed in the proof of Theorem 1.

**Corollary 2.** Consider the set of worst-case distributions

\[
P^* = \text{arg} \sup_{P \in \mathcal{P}_{\mu}(\mu, b, d)} E_\mu[X | X \geq t].
\]

Then,

(i) If \( t \in [0, \tau_1] \), \( P^* = \{ P \in \mathcal{P}_{\mu}(\mu, b, d) \mid \mathbb{P}(X \in [t, b]) = 1 \} \), all distributions in \( \mathcal{P}_{\mu}(\mu, b, d) \) that are supported on the interval \([t, b]\).

(ii) If \( t \in [\tau_1, \tau_2] \), \( P^* = \{ P \in \mathcal{P}_{\mu}(\mu, b, d) \mid \mathbb{P}(X = 0) = \frac{\mu - \tau_1}{\mu - \mu}, \mathbb{P}(X > \tau_2) = 1 - \frac{d}{2\mu - \tau_2} \} \), the three-point distribution as derived in scenario 1a in the proof of Theorem 1.

(iii) If \( t \in \mu, \tau_2, b \), \( P^* = \{ P \in \mathcal{P}_{\mu}(\mu, b, d) \mid \mathbb{P}(X = \mu) = \frac{d}{2\mu - \tau_2} \}, \mathbb{P}(X > t) = 1 - \frac{d}{2\mu - \tau_2} \), all discrete/mixed distributions with probability mass \( \frac{d}{2\mu - \tau_2} \) on \( 0 \) and the remainder of its probability mass supported on \([t, b]\).

(iv) If \( t \in [\tau_2, b] \), \( P^* = \{ P \in \mathcal{P}_{\mu}(\mu, b, d) \mid \mathbb{P}(X = t) = \frac{d}{2\mu - \tau_2} \}, \mathbb{P}(X > t) = 1 - \frac{d}{2\mu - \tau_2} \), all discrete/mixed distributions with probability mass \( \frac{d}{2\mu - \tau_2} \) on \( t \) and the remainder of its probability mass supported on \([0, \mu]\).

**Proof.** The proof follows almost directly from the complementarity slackness relationship explained in the proof of Theorem 1. For \( t \in [0, \tau_1] \) the dual solution function coincides with \( 1 \{ x \geq t \} \) on the interval \([t, b]\). Hence, all distributions that are supported on this interval and comply with the mean and MAD requirements are possible candidates for the worst-case distribution. Next, one can apply a similar reasoning for \( t \in [\mu, \tau_2] \) and \( t \in [\tau_2, b] \). The worst-case distribution can exist on the range where the dual solution function \( P^*(x) \) and the indicator function coincide. To attain the same optimal value, the probability mass on the singletons is chosen accordingly. Finally, note that the second case is already shown in the proof of Theorem 1.

Observe that when \( t \) equals \( \tau_1, \mu, \) or \( \tau_2 \), there is only a single discrete extremal distribution. Figure 2 provides examples of the worst-case distributions for several different parameter settings and values of \( t \). For the sake of exposition, all depicted examples have a continuous part that is uniform over its supported interval. **Corollary 2** shows that the ambiguity set \( \mathcal{P}_{\mu}(\mu, b, d) \) results in a non-trivial collection of worst-case distributions; that is, the mean-MAD approach results in a set that does not solely include discrete distributions with a small number of atoms for \( t \notin \{ \tau_1, \mu \} \cup \{ \tau_2 \} \).

2.2. Sharp bounds for other types of ambiguity

The primal-dual method is a general approach, often used in DRO, with a much wider range of possible applications. This subsection demonstrates that the semi-infinite programming problems can be adapted to different ambiguity sets, thus incorporating other types of information. We first derive alternative (tight) bounds for the tail probability, where we use a different measure of central tendency: the median. We then turn back to mean-MAD information and derive sharper bounds with a commonly used skewness measure that complements the mean-MAD ambiguity set, i.e., \( \mathbb{P}(X \geq \mu) \).

For the first adjustment, assume we know the following parameters: the support \([0, b]\), the median \( m \) and the mean absolute deviation (from the median) \( d_m \). We now obtain a different set of distributions, namely

\[
P_{\mu}(b, d_m) = \{ P : \mathbb{B} \to [0, 1] | \mathbb{P}(X \in [0, b]) = 1, \mathbb{P}(X \geq m) \geq \frac{1}{2}, \mathbb{P}(X \leq m) \geq \frac{1}{2}, E_m[|X - m|] = d_m \},
\]

If the distribution of \( X \) resides in this ambiguity set, the tight bounds are given by the optimal value of

\[
\sup_{P \in \mathcal{P}_{\mu}} \int_{x} 1_{x \in [x]} d\mathbb{P}(x)
\]

s.t.

\[
\int_{x} d\mathbb{P}(x) = 1, \quad \int_{x} 1_{x \in [x]} d\mathbb{P}(x) \geq \frac{1}{2}, \quad \int_{x} |x - m| d\mathbb{P}(x) = d_m.
\]

(12)
This then gives the dual problem

\[
\inf_{\lambda_0, \lambda_1, \lambda_2 \geq 0} \lambda_0 + (\lambda_1 + \lambda_2) \frac{1}{2} + \lambda_2 d_0
\]

s.t. \(1_{[x \leq t]} \leq \lambda_0 + \lambda_1 1_{[x \leq m]} + \lambda_1 1_{[x \leq m]} + \lambda_2 |x - m| =: F_m(x), \forall x \in [0, b].\)

\[
\text{This then gives the dual problem} \\
\inf_{\lambda_0, \lambda_1, \lambda_2 \geq 0} \lambda_0 + (\lambda_1 + \lambda_2) \frac{1}{2} + \lambda_2 d_0
\]

\[
\text{s.t. } 1_{[x \leq t]} \leq \lambda_0 + \lambda_1 1_{[x \leq m]} + \lambda_1 1_{[x \leq m]} + \lambda_2 |x - m| =: F_m(x), \forall x \in [0, b].
\]

(13)

We can solve the dual problem by exploiting the structure of \(F_m(x)\), as with the proof of Theorem 1. The different scenarios are depicted in Fig. 3. The following theorem presents the median-MAD bounds. The details of the proof are relegated to the online supplement.

**Theorem 2.** For a random variable \(X\) with a distribution \(P \in \mathcal{P}_{(m, b, d_0)}\),

\[
\sup_{P \in \mathcal{P}_{(m, b, d_0)}} P(X \geq t) = \begin{cases} \inf \left\{ 1, \frac{b - 2d_0}{d_0} + \frac{1}{2} \right\}, & t \in [0, m), \\ \inf \left\{ \frac{b}{\mu - m}, \frac{m}{\mu - m} \right\}, & t \in [m, b]. \end{cases}
\]

(14)

In robust statistics the median is widely considered as a more suitable location parameter than the mean when the distribution is estimated from historical data and contaminated with outliers through another (possibly fat-tailed) distribution (Casella & Berger, 2002). As the former puts less importance on the tail of the distribution, it is barely affected by those outliers. The median and MAD around the median are the robust variants of, respectively, the mean and standard deviation. Next to describing the underlying distribution more accurately, the median might also provide a better measure of central tendency for the quantity that we are estimating from historical data. For example, the median wealth of a population is a better measure of 'typical' wealth than the expected value since the distribution of wealth is often skewed. To illustrate the median-MAD tail bound (14), we plot a small example in which the ground truth follows a Pareto distribution, see Fig. 4. In particular, when the shape parameter equals 2, the variance is infinite, rendering Chebyshev's inequality useless, but the median-MAD bound can still be computed. In the remainder of this paper, we will focus on the mean-MAD ambiguity set now that we demonstrated our approach is also applicable to other types of information.

We next consider the tail bounds when also a specific measure of skewness is known: \(\beta = P(X \geq \mu)\). Since this statistic contains information about the distribution of \(X\) relative to its mean \(\mu\), it is often combined with mean-MAD information (Ben-Tal & Hochman, 1972; Postek et al., 2018). Define the restricted ambiguity set

\[
P_{(\mu, b, d, \beta)} = \{ P : P \in P_{(m, b, d)}, P(X \geq \mu) = \beta \}.
\]

(15)

Using this ambiguity set results in new tight bounds. These results are stated in the following two results for which the primal-dual proofs are also provided in the online supplement.
Theorem 3. For a random variable $X$ with a distribution $\mathbb{P} \in \mathcal{P}(\mu, b, \beta)$,

$$
\begin{align*}
\sup_{\mathbb{P} \in \mathcal{P}(\mu, b, \beta)} \mathbb{P}(X \geq t) &= \begin{cases} 
1, & t \in [0, \tau_1], \\
\frac{1 - \beta}{\beta t} - \frac{d}{\beta}, & t \in [\tau_1, \mu], \\
\frac{1 - \beta}{2(1 - \beta)} - \frac{d}{2\beta}, & t \in [\mu, \tau_2], \\
0, & t \in [\tau_2, b].
\end{cases} \\
\end{align*}
$$

with

$$
\tau_1 = \mu - \frac{d}{2(1 - \beta)}, \quad \tau_2 = \mu + \frac{d}{2\beta}.
$$

Theorem 4. For a random variable $X$ with a distribution $\mathbb{P} \in \mathcal{P}(\mu, b, \beta)$,

$$
\begin{align*}
\inf_{\mathbb{P} \in \mathcal{P}(\mu, b, \beta)} \mathbb{P}(X > t) &= \begin{cases} 
\beta \frac{d}{2(1 - \beta)} - \frac{d}{2\beta}, & t \in [0, \tau_1], \\
\frac{1 - \beta}{(1 - \beta)} - \frac{d}{2\beta}, & t \in [\tau_1, \mu], \\
\frac{\beta(\mu - t)}{2(1 - \beta)} + \frac{d}{2\beta}, & t \in [\mu, \tau_2], \\
0, & t \in [\tau_2, b].
\end{cases} \\
\end{align*}
$$

with

$$
\tau_1 = \mu - \frac{d}{2(1 - \beta)}, \quad \tau_2 = \mu + \frac{d}{2\beta}.
$$

2.3. Comparison with other bounds

Closely related to our results is the discussion in section 4.1 of Ghosal & Wiesemann (2020). They consider, among others, the ambiguity set

$$
\mathcal{P}(\mu, b, d) = \{\mathbb{P} : \mathbb{E} X \in [0, b] = 1, \mathbb{E}\mathbb{P}[X] = \mu, \mathbb{E}\mathbb{P}[|X - \mu|] \leq d\}.
$$

The only difference with the ambiguity set we use is the inclusion of all distributions with a lower mean absolute deviation. This has major implications for the maximum and minimum probability that $X$ exceeds $t$, however. First of all, it should be noted that the distribution with all its probability mass on $\mu$ is an element of $\mathcal{P}(\mu, b, d)$ for any value of $d$. This means that for any $t \leq \mu$ it holds that

$$
\sup_{\mathbb{P} \in \mathcal{P}(\mu, b, d)} \mathbb{P}(X \geq t) = 1.
$$

Moreover, for any $t > \mu$ and $d > 2\mu(t - \mu)/t$, the maximum probability of $X$ exceeding $t$ is attained by a distribution with a mean absolute deviation equal to $2\mu(t - \mu)/t$, which is explained by the observation that the bound we obtain is decreasing in $d$ for $d > 2\mu(t - \mu)/t$.

Clearly, because of the above observations, the theoretical maximum of $\mathbb{P}(X > t)$ has a much simpler closed-form solution than (5) for the ambiguity set $\mathcal{P}(\mu, b, d)$. A big downside is that many of the distributions contained in $\mathcal{P}(\mu, b, d)$ but not in $\mathcal{P}(\mu, b, \beta)$ are unrealistic. Especially when the mean absolute deviation is known or can be accurately estimated, there is little reason to consider distributions with a different (in this case lower) mean absolute deviation. For large values of $d$ relative to $t$ in particular, using $\mathcal{P}(\mu, b, d)$ can lead to an overestimation of the maximum value of $\mathbb{P}(X > t)$.

The observation that the maximum value of $\mathbb{P}(X > t)$ is decreasing in $d$ for large values of $d$ also means that considering distributions with a lower mean absolute deviation can lead to a higher bound on $\mathbb{P}(X > t)$.

Comparing the result of Theorem 1 to Cantelli’s inequality (1) is harder, since we assume the mean absolute deviation to be known, but not the variance. Hence, some relation between these two quantities is needed to be able to make a comparison. In particular, we will use that

$$
d^2 \leq \sigma^2 \leq \frac{d(b - a)}{2}.
$$

Throughout the comparison below we assume that $d$ is given and compare the bound obtained in Theorem 1 with Cantelli’s bound.
for different values of $\sigma$ satisfying (18). Figure 6 illustrates this comparison for a simple numerical example with the following parameters: $a = -1$, $\mu = 0$, $b = 1$, $d = \frac{1}{2}$. We consider three values for $\sigma$: $\sigma = d = \frac{1}{4}$, $\sigma = \frac{1}{2}$ and $\sigma = \sqrt{d(b-a)/2} = \frac{1}{2}$.

Figure 6 gives rise to a number of interesting observations. First of all, we note that since Cantelli’s bound is 1 for any $t \leq \mu$, the bound from Theorem 1 is at most Cantelli’s bound as it includes an interval for which it is not 1. Furthermore, the flat area in the blue line corresponds to the values of $t$ such that

$$\min \left\{ \frac{d}{2(t - \mu)}, 1 - \frac{d}{2(\mu - a)} \right\} = 1 - \frac{d}{2(\mu - a)},$$

which corresponds to all $\mu \leq t \leq t_2 := \mu + \frac{d(\mu - a)}{2(\mu - a) - d}$. Moreover, we note that for $\sigma = d$, Cantelli’s bound is lower than (5) for all $t_2 \leq t \leq b$. This is true for all parameters as:

$$\frac{d^2}{d^2 + (t - \mu)^2} \leq \frac{d^2}{2d(t - \mu)} = \frac{d}{2(t - \mu)}.$$

In particular, for $\sigma = d$ Cantelli’s bound and (5) always coincide at $t = \mu + d$, since

$$\frac{d^2}{d^2 + d^2} = \frac{1}{2} = \frac{d}{2d}.$$

If, on the other hand, we choose $\sigma = \sqrt{d(b-a)/2}$, its highest possible value, Cantelli’s bound is higher than (5). This is true for all parameter values as well, as Cantelli’s bound is increasing in $\sigma$ and must thus be at least (5) for its highest possible value.

For intermediate values of $\sigma$, we observe behavior similar to the line corresponding to $\sigma = \frac{1}{2}$ in Fig. 6. More specifically, we find that (5) is lower than Cantelli’s bound for all $t$ in the two intervals $[0, t]$ and $[t, T]$, with the three boundaries given by

$$t = \mu + \frac{\sqrt{d\sigma^2}}{2(\mu - a) - d},$$

$$T = \mu + \frac{d\sigma^2}{\sigma^2 - 1},$$

$$T = \min \left\{ b, \frac{\sigma^2}{d} + \frac{\sigma^2}{\sigma^2-1} \right\}.$$
Note that for some $\sigma$, such as $\sigma = \sqrt{d(b-a)/2}$ in Fig. 6, it holds that $\hat{\sigma} \geq \sigma$, that is, (5) is lower than Cantelli’s bound for all $t \in [\mu, \tau]$. To visually clarify all boundaries discussed above, Fig. 7 only shows Cantelli’s bound for $\sigma = 0.27$ and marks $t_2$, $\hat{\sigma}$ and $\sigma$. It should be noted that this value of $\sigma$ is very close to the minimum of 0.25, and hence Cantelli’s bound compares more favorably than is generally expected to be. In the online supplement we provide a similar comparison with the variance-based bound of De Schepper & Heijnen (1995), and also perform numerical experiments with the other bounds derived in the previous subsection.

3. Distribution-free analysis of OR models

We now turn to three OR applications: the newsvendor problem, stop-loss reinsurance and radiotherapy optimization. These three models can be subjected to distribution-free analyses that make use of the novel tight bounds. The common theme is that with ambiguity described in terms of mean, MAD and restricted support, distribution-free analysis often leads to valuable structural insights.

3.1. Newsvendor problem

The newsvendor problem serves to find the order quantity that maximizes the expected profit for a single period given a stochastic demand. Denote by $q$ the order quantity (number of units) and by $D$ the stochastic demand during a single selling period. Per unit, $p$ denotes the selling price and $c$ the purchase cost. Let $p > c$, and assume without loss of generality that unsold units have zero salvage value. The expected profit is $E_\theta[\pi(q, D)]$ with $D \sim F$ and $\pi(q, D) = p \min(q, D) - cq$.

The decision maker then chooses the optimal order quantity $q^*$ that solves $\max_q E_\theta[\pi(q, D)]$ with $D \sim F$ and $\pi(q, D) = p \min(q, D) - cq$.

The intervals $[q^*, q^d]$ for mean-MAD ambiguity with $\mu = 5$, $d = 1.5$ and various profit margins $\eta$.

Table 1

| $\eta$ | $b = 10$ | $b = 15$ | $b = 20$ | $b = \infty$ |
|--------|----------|----------|----------|---------------|
| 0.01   | [0.00, 4.17] | [0.00, 4.23] | [0.00, 4.26] | [0.00, 4.29] |
| 0.1    | [0.00, 4.67] | [0.00, 4.70] | [0.00, 4.71] | [0.00, 4.72] |
| 0.2    | [1.25, 5.94] | [1.25, 5.94] | [1.25, 5.94] | [1.25, 5.94] |
| 0.3    | [3.13, 6.25] | [3.13, 6.25] | [3.13, 6.25] | [3.13, 6.25] |
| 0.5    | [3.50, 6.50] | [3.50, 6.50] | [3.50, 6.50] | [3.50, 6.50] |
| 0.7    | [3.93, 7.50] | [3.93, 7.50] | [3.93, 7.50] | [3.93, 7.50] |
| 0.9    | [5.33, 10.00] | [4.17, 12.50] | [4.17, 12.50] | [4.17, 12.50] |
| 0.95   | [5.63, 10.00] | [5.31, 15.00] | [5.00, 20.00] | [4.21, 20.00] |
| 0.99   | [5.83, 10.00] | [5.77, 15.00] | [5.71, 20.00] | [4.24, 80.00] |

Fig. 7. An illustration of $t$, $t_2$, $\hat{\sigma}$ and $\sigma$ for the parameter values $a = -1$, $\mu = 0$, $b = 1$ and $d = 0.25$.

The newsvendor problem into a maximin decision maker that solves

$$\max_q \inf_{\theta \in \mathcal{P}_{(\mu, \sigma)}} E_\theta[\pi(q, D)]$$

with $\mathcal{P}_{(\mu, \sigma)}$ the ambiguity set that contains all distributions with a given mean $\mu$ and variance $\sigma^2$, and solution

$$q^* = \left\{ \begin{array}{ll} 0, & \text{if } \eta < \frac{\sigma^2}{\sigma^2 + \sigma^2} \\ \mu + \frac{\sigma^2 - 1}{\sqrt{\eta(1-\eta)}}, & \text{if } \frac{\sigma^2 - 1}{\sigma^2 + \sigma^2} \leq \eta \leq 1 - \frac{\sigma^2}{\sigma^2 + \sigma^2} \\ \eta & \text{if } \eta \geq 1 - \frac{\sigma^2}{\sigma^2 + \sigma^2} \end{array} \right.$$ (21)

We shall instead consider all demand distributions with given mean $\mu$, MAD $d$ and support $[0, b]$, and consider

$$\max_q \inf_{\theta \in \mathcal{P}_{(\mu, 0, b)}} E_\theta[\pi(q, D)]$$

This is the counterpart of problem (20). Scarf (1958) solved (20) directly, computing the lower bound $\inf_{\theta \in \mathcal{P}_{(\mu, 0, b)}} E_\theta[\pi(q, D)]$ via a linear program. Instead, we do not solve (22) directly, but apply the tail probability bound from Theorem 1 to the first-order condition for $q^*$ in (19). Clearly, the tight lower and upper bounds for $q^*$ follow from substituting $\inf_{\theta \in \mathcal{P}_{(\mu, 0, b)}} P(D > q)$ and $\sup_{\theta \in \mathcal{P}_{(\mu, 0, b)}} P(D > q)$ into (19), respectively.

Proposition 1 (Order quantity bounds under mean-MAD-range ambiguity). Suppose the newsvendor knows the mean $\mu$, the absolute deviation $d$ and the support’s upper bound $b$ of the demand distribution $\mathbb{P}(D \leq q)$. The optimal order quantity $q^*$ that solves $\max_q E_\theta[\pi(q, D)]$ is then contained in the interval $[q^*, q^d]$ with

$$[q^*, q^d] := \left\{ \begin{array}{ll} \left[ \frac{2\mu(b-q)-bd}{d(b-\mu)} + d(1-\mu), \mu + \frac{d}{2}\right], & \text{if } \eta < \frac{d}{2d} \\ \left[ \frac{d}{2d}, \mu + \frac{d}{2}\right], & \text{if } \frac{d}{2d} \leq \eta \leq 1 - \frac{d}{2d} \\ \left[ \mu - \frac{d}{2d}, b \right], & \text{if } \eta \geq 1 - \frac{d}{2d} \end{array} \right.$$ (23)

where $\eta = (p-c)/p$ is the critical quantile of the distribution of $D$.

The proposition provides various handles for a robust policy that responds to the uncertainty captured in $\mathcal{P}_{(\mu, 0, b)}$. The lower bound $q^*$ follows from the worst-case demand distribution. Observe that $q^*$ is larger than $\mu$ when the profit margin $\eta$ exceeds $1 - d/2(b-\mu)$, and smaller than $\mu$ otherwise. This insight can be contrasted with $q^d$ in (21) that also considers the worst-case scenario, but then in view of $\mathcal{P}_{(\mu, 0, b)}$ ambiguity. Scarf’s $q^d$ is larger than $\mu$ if $\eta > 1/2$ and smaller than $\mu$ otherwise. Hence, $q^d$ quantifies the dependency on $b$, where $q^d$ does not. In particular, when the profit margin $\eta$ is fixed, the pessimistic newsvendor that uses $q^d$ will only order above the mean when $b$ does not exceed $\mu + d/(2(1-\eta))$.

Table 1 shows that the support $[0, b]$ also influences the intervals $[q^*, q^d]$, in particular for low and high profit margins. We also recognize the three different regimes in Proposition 1 that correspond to low margins, average margins and high margins.

938
We mention two further works related to Proposition 1. Ben-Tal & Hochman (1976) use general techniques for stochastic programs with limited information such as (22). For such stochastic programs the available information is often not sufficient to find the optimal solution. Ben-Tal & Hochman (1976) develop a method to construct the minimal set that should contain the optimum. They also demonstrate this technique for the newsvendor model with given mean and MAD, but unbounded support, and obtain intervals that indeed arise from Proposition 1 for the limit $b \to \infty$:

$$[q^L, q^U] = \begin{cases} 
0, \frac{\mu - db}{\eta - \beta d}, & \text{if } \eta < \frac{d}{\beta d}, \\
\mu - \frac{d}{\beta d}, & \text{if } \frac{d}{\beta d} \leq \eta < 1 - \beta, \\
\mu, & \text{if } \eta = 1 - \beta, \\
\mu + \frac{d(1 + d)}{\beta(1 - \beta)}, & \text{if } \eta > 1 - \frac{d}{\beta(1 - \beta)}, \\
\frac{b - \beta d - \mu}{1 - \beta - \mu}, & \text{if } \eta < \frac{d}{\beta d}.
\end{cases}$$  

(24)

Natarajan et al. (2018) introduce semi-variance as a piece of information about the skewness of the distribution. Together with the mean and variance, this results in a more restrictive ambiguity set (compared to Scarf), and therefore a less conservative (or sharper) estimation of $q^*$. In our case, we restrict the ambiguity set further with $\mathbb{P}(X \geq \mu) = \beta$ information that, like semi-variance, measures skewness. The following problem is the mean-MAD counterpart of the mean-variance-semivariance model discussed in the work of Natarajan et al. (2018):

$$\max q \inf_{(\mu, d, \beta) \in \mathcal{P}} \mathbb{E} \left[ \eta \mathcal{Q}(q, D) \right],$$

where $\beta$ adopts the role of the semivariance as a measure of skewness. We use Theorems 3 and 4 to bound the total demand distribution of the demand, and obtain sharper bounds for $q^*$. Proposition 2 (Order quantity bounds under mean-MAD-β ambiguity). Suppose the newsvendor knows that $\mathbb{P} \in \mathcal{P}_{(\mu, d, \beta)}$ and $\mathbb{P}(D \geq \mu) = \beta$. The optimal order quantity $q^*$ that solves $\max_q \mathbb{E} \left[ \eta \mathcal{Q}(q, D) \right]$ is then contained in the interval $[q^L, q^U]$ with

$$[q^L, q^U] = \begin{cases} 
0, \frac{(1 - \beta)d - db + \mu}{\eta - \beta d}, & \text{if } \eta < \frac{d}{\beta d}, \\
\mu - \frac{d}{\beta d}, & \text{if } \frac{d}{\beta d} \leq \eta < 1 - \beta, \\
\mu, & \text{if } \eta = 1 - \beta, \\
\mu + \frac{d(1 + d)}{\beta(1 - \beta)}, & \text{if } \eta > 1 - \frac{d}{\beta(1 - \beta)}, \\
\frac{b - \beta d - \mu}{1 - \beta - \mu}, & \text{if } \eta < \frac{d}{\beta d}.
\end{cases}$$  

(26)

where $\eta = (p - c)/p$ is the critical quantile of the distribution of $D$.

This result provides a robust policy that protects against uncertainty contained in $\mathcal{P}_{(\mu, d, \beta, \beta)}$. Obviously, ordering the mean is optimal if $\eta = 1 - \beta$. The lower bound $q^L$ relates to the worst-case demand distribution. Similar to the case with mean-MAD-range information, $q^L$ is larger than $\mu$ when the profit margin $\eta$ exceeds $1 - d/2(b - \mu)$. Hence, skewness information does not determine whether the pessimistic newsvendor orders more than the mean, since the upper bound $b$ again plays a decisive role. This can be contrasted with the results of Natarajan et al. (2018), who show that for $\mathcal{P}_{(\mu, \sigma, \beta)}$ ambiguity, the order quantity is greater than $\mu$ if $\eta > \frac{1}{2}(1 + s)$, where $s$ is the normalized semivariance.

Table 2 shows that the bounded support $[0, b]$ again influences the intervals for low and high profit margins. The new intervals are sharper than the ones found in Table 1. This is, of course, an obvious result of incorporating more distributional information.

Apart from modifying or narrowing the ambiguity set, conservatism can be alleviated by choosing alternate objective functions, for instance by replacing the profit function by a regret function (opportunity cost of not making the optimal decision) (Perakis & Roels, 2008; Yue, Chen, & Wang, 2006), or by extending the profit function with a utility function $u(.)$ for max-min analysis of $\mathbb{E}[u(q, D)]$ (Han, Du, & Zhuang, 2014). See Natarajan et al. (2018) for an extensive review of many other studies on distribution-free newsvendor models. The tight bounds developed in this paper can be used for distribution-free analysis of more advanced models, including those modeling regret and utility mentioned above, the risk-averse newsvendor with stochastic price-dependent demand (Chen, Xu, & Zhang, 2009) and multi-product settings (Choi, Ruszczyński, & Zhao, 2011).

3.2. Stop-loss reinsurance

Reinsurance is a classical topic in the actuarial sciences and insurance mathematics and implies that an insurance company transfers part of its risk to a reinsurance company (Asmussen & Albrecher, 2010; Kaas, Goovaerts, Dhaene, & Denuit, 2008). Say an insurance company faces a total claim $S$ that is the sum of $n$ individual claims $X_i, i = 1, \ldots, n$. The insurance company pays the claim up to a level $z$, and the reinsurance company covers the remainder. This gives rise to the so-called retention function $\psi(z) = \min[S, z]$ that represents the payment of the insurer. We provide an upper bound for the standard stop-loss reinsurance function in Proposition 3.

Proposition 3. The worst-case expected claim payment of the direct insurer as a function of the retention limit $z$ is given by

$$\sup_{\mathbb{P} \in \mathcal{P}_{(\mu, d, \beta)}} \mathbb{E}_\mathbb{P} \left[ \psi(z, \mathcal{Q}(q, D)) \right] = \begin{cases} 
z, & \text{if } z \in [0, \tau_1], \\
\mu - \frac{d(b - \mu)}{2(b - \mu)} - d, & \text{if } z \in [\tau_1, \mu], \\
z(1 - \frac{d}{2(b - \mu)}), & \text{if } z \in [\mu, \mu_2], \\
\mu, & \text{if } z \in [\tau_2, b].
\end{cases}$$  

(27)

where

$$\tau_1 = \mu - \frac{d(b - \mu)}{2(b - \mu)} - d, \quad \tau_2 = \mu + \frac{d\mu}{2\mu - d}.$$  

Proof. First, note that $\psi(z) = \min[S, z] = S - \max[S - z, 0]$, and hence

$$\sup_{\mathbb{P} \in \mathcal{P}_{(\mu, d, \beta)}} \mathbb{E}_\mathbb{P} \left[ \psi(z, \mathcal{Q}(q, D)) \right] = \mu - \sup_{\mathbb{P} \in \mathcal{P}_{(\mu, d, \beta)}} \mathbb{E}_\mathbb{P} \left[ \max[S - z, 0] \right].$$  

(28)

The second term is convex in $S$, and thus equivalent to, see equation (12) of Postek et al. (2018)

$$\min_{z \in [\mu, \mu_2]} \sum_{(\mathbb{P} \in \mathcal{P}_{(\mu, d, \beta)})} \mathbb{E}_\mathbb{P} \left[ \sup_{z \in [0, b]} \left\{ \theta \max[\mu + \frac{d}{\beta d} - z, 0] + (1 - \theta) \max[\mu - \frac{d}{2(1 - \theta)} - z, 0] \right\} \right].$$  

(29)

a convex optimization problem with a piecewise linear objective function. The optimal value depends on the retention limit $z$. Solving problem (29) for $z \in [0, b]$ and subtracting the optimal value from $\mu$ results in the four cases mentioned in (27).

The payment function of the reinsurance company constitutes a more challenging problem when the insurance coverage is limited. In this case, a relevant performance characteristic is to what extent
the insurance company benefits from the reinsurance contract. This benefit is measured with the function

$$\phi(z, m, S) = \begin{cases} m, & \text{if } z \geq z + m, \\ S - z, & \text{if } z \leq S \leq z + m, \\ 0, & \text{if } z \leq S. \end{cases}$$

(30)

When the total claim \( S \) stays below the retention limit \( z \), the insurance company covers the entire claim, but when \( S \) exceeds \( z \) the reinsurer pays the excess claim up to a maximum \( m \). Thus, the reinsurance company does not compensate large claims that exceed the exit point \( m + z \). Above this level the risk is retained by the insurance company. We obtain a novel bound by primal-dual arguments.

**Proposition 4.** The expected insurer’s benefit is bounded by

$$\sup_{P \in \mathcal{P}_{(\mu,b,d)}} \mathbb{E}[\phi(z, m, S)] = \begin{cases} \min[m, m2(\mu - \frac{db(c-m)(c-r))}{2(b-c)}), & \text{if } z \leq z + m \leq \mu, \\ \min[m(1 - \frac{d}{b})n(z), (z - \mu) + \mu], & \text{if } z \leq S \leq z + m \leq b, \\ \min[m(1 - \frac{d}{b}) + \mu, (z + m - \mu)], & \text{if } z \leq z + m \leq b, \\ \end{cases}$$

(31)

where the function \( \phi(z, m, S) \) degenerates to \( \max(S - z, 0) \) if \( z + m > b \).

In this case,

$$\sup_{P \in \mathcal{P}_{(\mu,b,d)}} \mathbb{E}[\phi(z, m, S)] = \begin{cases} z(\mu - \frac{d}{b} - 1) + \mu, & \text{if } z \leq \mu, \\ \frac{db(c-m)(c-r))}{2(b-c)} + \mu, & \text{if } z \geq \mu. \end{cases}$$

(32)

For the sake of conciseness, we only sketch the proof. The full details are highly similar to the derivations used in the proof of Theorem 1.

We will show via primal-dual reasoning that the stated stop-loss formulas are tight upper bounds. We consider the measurable function \( \phi(z, m, S) \). For a random variable \( S \) with distribution \( P \in \mathcal{P}_{(\mu,b,d)} \), we solve

$$\sup_{P \in \mathcal{P}_{(\mu,b,d)}} \mathbb{E}[\phi(z, m, S)] \text{ s.t. } \int_S \phi(z, m, S) dP(s) = 1, \int_S s dP(s) = \mu, \int_S |s - \mu| dP(s) = d.$$ 

(33)

Consider the dual of (33),

$$\inf_{\lambda_0, \lambda_1, \lambda_2} \lambda_0 + \lambda_1 \mu + \lambda_2 d \text{ s.t. } \phi(z, m, s) \leq \lambda_0 + \lambda_1 s + \lambda_2 (s - \mu), \forall s \in [0, b].$$

(34)

Define \( F(s) := \lambda_0 + \lambda_1 s + \lambda_2 (s - \mu) \). Then the inequality in (34) can be written as \( \phi(z, m, s) \leq F(s) \), i.e., \( F(s) \) majorizes the ‘staircase’ function \( \phi(z, m, s) \). Note that \( F(s) \) has a ‘kink’ at \( s = \mu \). There are six candidate scenarios, which are displayed in Fig. 8. When \( m + z \leq \mu \), \( F(s) = 1 \) and touches \( \phi(z, m, s) \) in \([0, m + z, b]\) (scenario 1a), or \( F(s) \) touches \( \phi(z, m, s) \) in \([m + z, b]\) (scenario 1b). When \( z \leq \mu \leq m + z \), \( F(s) \) touches in \([0, \mu, m + z, b]\) (scenario 2a) or in \([0, \mu, m + z, b]\) (scenario 2b). Finally, if \( \mu \leq z \leq m + z \), \( F(s) \) coincides with \( \phi(z, m, s) \) in \([0, \mu, m + z, b]\) (scenario 3a) or in \([0, m + z, b]\) (scenario 3b).

Scenario 1a implies that \( F(0) = F(m + z) = F(\mu) = F(b) = m \), and hence \( \lambda_0 = m, \lambda_1 = \lambda_2 = 0 \) with objective value \( m \). It is clear that the optimal primal objective value is also equal to \( m \) as the primal solution can only assign probability to values greater than or equal to \( m + z \) (which is a consequence of complementary slackness).

Scenario 1b implies \( F(0) = 0, F(m + z) = F(b) = m \). It can be shown that allocating probability mass to the points \([0, m + z, b]\) in the primal problem (33) yields the same objective value, and hence the corresponding solutions are optimal.

Similarly, scenarios 2a, 2b, 3a and 3b imply values for at least three of \( F(0), F(\mu), F(m + z) \) and \( F(b) \), from which a dual and primal solution with equal objective value can be derived. The proof of the first part of the theorem is then completed by taking the minimum for each scenario. The second part is an immediate consequence of upper bound (8) in Postek et al. (2018), which is a result that was already shown by Ben-Tal & Hochman (1972).

An illustration of the bounds for the stop-loss payments is provided in Fig. 9, where we display payments as functions of \( z \) with \( \mu = 5, d = 1.77, \) and \( m = 3, m = 5 \) and \( m \to \infty \). We assume that the ‘true’ total claim \( S \) follows a Poisson(5) distribution. Note the resemblance between the shape of the stop-loss bound in Fig. 9 and the mean-MAD tail probability bound in Theorem 1. The former bound, however, has an additional linear part with a negative slope for \( \mu - m \leq z \leq \mu \). This linear segment is only present when \( m \) exceeds \( d/2 \); moreover, the bound approaches a linear function for \( z \leq \mu \) when \( m \) is chosen sufficiently large. Additionally, letting \( m \to \infty \), our example results in a bound equal to the constant \( d/2 \) if \( z \geq \mu \), and thus the bound for the stop-loss payment of the reinsurer degenerates to a piecewise linear function consisting of two parts (a linear part with negative slope \( d/2 \mu - 1 \) and a constant part equal to \( d/2 \)).

These results complement the vast literature on tight bounds for expected claim payments. Cox (1991) considers bounded sup-
port and known first and second moment and obtains tight bounds using general results for moment problems. Other related works explore ways to sharpen the bounds using additional information. When modifying the ambiguity set by incorporating skewness information, imposing unimodality and symmetry conditions or using higher order moments, the gap between the upper and lower bounds narrows significantly; see Heijnen (1990), De Vylder & Goovaerts (1982) and Jansen, Haezendonck, & Goovaerts (1986). Note that mean–MAD information can easily be extended with additional parameters, such as the probability $\beta = \Pr(S \geq \mu)$ or the median.

3.3. Radiotherapy optimization

We consider a continuous optimization problem that arises in radiotherapy. Here, the biological effective radiation dose delivered to a tumor is to be maximized subject to a constraint on the biological effective dose delivered to the surrounding healthy tissue. Mathematically, the biological effective dose (BED) for a dose $x \in \mathbb{R}^n$ delivered over $n$ fractions is given by

$$B(x) = \sum_{i=1}^{n} x_i + \frac{1}{\rho}\left(x_i^2 + x_i^3\right),$$

where $\rho$ is the radiosensitivity parameter of the irradiated tissue. More specifically, it can be interpreted as the tissue’s sensitivity to fractionation, where a low value indicates a high sensitivity to fractionation, i.e., the distribution of treatment over multiple fractions.

While there is an extensive body of research on the value of $\rho$ for different tumor sites, it remains subject to significant uncertainty (Joiner & Van der Kogel, 2016). Moreover, since this value can differ from patient to patient, there is a very limited amount of data available and there is little evidence to suggest it follows some well known distribution. Throughout the rest of the example, we denote the sensitivity to fractionation by $\rho_1$ and $\rho_2$ for the tumor and the surrounding healthy tissue, respectively.

For illustrative purposes, we consider a setting in which it has been decided to deliver the treatment over two fractions, i.e., the optimization variables are limited to the dose in the first and second fraction. Moreover, we focus on the uncertainty of $\rho_2$, and thus model the restriction of sparing the healthy tissue through an ambiguous chance constraint. Mathematically, we wish to solve the following optimization problem (Ten Eikelder, Ajdari, Bortfeld, & Den Hertog, 2019):

$$\max_{x_1, x_2} x_1 + x_2 + \frac{1}{\rho_1}\left(x_1^2 + x_2^2\right)$$

subject to:

$$\Pr\left(\sigma(x_1 + x_2) + \frac{1}{\rho_2}\sigma^2(x_1^2 + x_2^2) \leq t(\rho_2)\right) \geq 1 - \epsilon, \quad \forall \epsilon \in \mathcal{F}_{(s, h, \delta)},$$

$$x_1, x_2 \geq x_{\min},$$

where $\sigma$ is the generalized dose-sparing factor that denotes the fraction of the mean tumor dose that the healthy tissue receives.

Fig. 9. The bounds and ‘true’ values of the expected claim payment of the insurance company $E_x[\psi(x, S)]$ and the insurer’s benefit $E_x[\psi(x, S, m, S)]$ as functions of the retention limit $z$ with $\mu = 5$, $d = 1.77$, and $m = 3$, $m = 5$ or $m = \infty$. The red lines depict the upper bounds and the blue lines give the true expected claim payments when $S \sim \text{Poisson}(5)$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
on average, $x_{\text{min}}$ is the minimum dose that must be delivered in each fraction, and $t(\rho_2)$ denotes the tolerance level of the healthy tissue and is given by

$$ t(\rho_2) = \phi D \left( 1 + \frac{\phi D}{\rho_2} \right). $$

In other words, the healthy tissue is known to tolerate a total dose of $D$ gray if it is delivered in $T$ fractions under dose shape factor $\phi$. This dose shape factor is a parameter that characterizes the spatial heterogeneity of a dose distribution (Perkó, Bortfeld, Hong, Wolfgang, & Unkelbach, 2018).

The ambiguity of $\rho$ is modeled through the mean-MAD ambiguity set, where the lower bound of the support is given by $a$ instead of 0. In general, convex ambiguous chance constraints in which the uncertain parameter appears on the right-hand side can be reformulated as a tractable convex constraint.

**Proposition 5.** Let $g : \mathbb{R}^n \rightarrow \mathbb{R}$, $h \in \mathbb{R}$ and let $Z$ be a random variable whose distribution lies in the ambiguity set

$$ \mathcal{P} = \{ P : P[Z \in [-1, b]] = 1, \mathbb{E}[Z] = 0, \mathbb{E}[|Z|] = d \}, $$

for some $d \in [0, \frac{2b}{1+b}]$. For any $\epsilon \in (0, \frac{1}{1+b})$ and $x \in \mathbb{R}^n$ it holds that

$$ \inf_{P \in \mathcal{P}} P[g(x) + Z \leq 0] \geq 1 - \epsilon, $$

if and only if

$$ g(x) + \min \left\{ b, \frac{d}{2\epsilon} \right\} \leq 0. $$

**Proof.** We first rewrite (36) to

$$ \sup_{P \in \mathcal{P}} P[Z > -g(x)] \leq \epsilon. $$

From Theorem 1 and the fact that $\epsilon < 1/(1+b)$ we know that it must hold that $-g(x) > \mathbb{E}[Z] = 0$. Given that requirement, we know by Theorem 1 that

$$ \sup_{P \in \mathcal{P}} P[Z > -g(x)] = \begin{cases} \min \left\{ -\frac{d}{2\epsilon x_i}, 1 - \frac{d}{2} \right\} & \text{if} \ -g(x) < 1 \\ 0 & \text{if} \ -g(x) \geq 1. \end{cases} $$

From $d \in [0, \frac{2b}{1+b}]$ and $\epsilon \in (0, \frac{1}{1+b})$, it follows that $1 - d/2 > \epsilon$, and thus any feasible solution $x$ must satisfy $-g(x) \geq 1$ and/or $-d/2g(x) \leq \epsilon$. The latter can be equivalently stated as

$$ -g(x) \geq \frac{d}{2\epsilon}. $$

which can easily be combined with the former as

$$ -g(x) \geq \min \left\{ 1, \frac{d}{2\epsilon} \right\} \iff g(x) + \min \left\{ 1, \frac{d}{2\epsilon} \right\} \leq 0. $$

Because $d/2\epsilon > 0$, we find that the requirement $-g(x) > 0$ is redundant, and thus (36) is equivalent to (37). □

The ambiguous chance constraint (35b) is not naturally stated in the form (36). It can be rewritten, however, as

$$ \mathbb{P}\left( \rho \cdot (x_1 + x_2 - \phi D) > \frac{\phi^2 D^2}{\mu} - \sigma^2 (x_1^2 + x_2^2) \leq \epsilon \right). $$

where we note multiplication by $\rho$ is allowed as its support is nonnegative. Leveraging the tail probability bound, we find for $\epsilon \in (0, \frac{\mu - a}{b - a})$ that (38) is equivalent to

$$ \mu \sigma (x_1 + x_2) + \sigma^2 (x_1^2 + x_2^2) - \frac{\sigma^2 (x_1^2 + x_2^2)}{\mu} \leq \mu \phi D + \frac{\phi^2 D^2}{\mu}. $$

The resulting optimization problem can be solved efficiently, as (39) can be equivalently stated as two conic quadratic inequalities through the introduction of an auxiliary variable.

We solve (35) for a specific, realistic set of parameters taken from Ten Eikelder et al. (2019), which are reported in Table 3.

Figure 10 shows the feasible region and optimal solution of (35) for different values of $\epsilon$ as well as the feasible region when we assume having the exact knowledge that $\rho = \mu$ and the feasible region when $\epsilon = 0.1$ and $\rho \sim \text{Beta}(6.6, 13.2)$, which is a member of the ambiguity set. Remarkable in this example is the similarity between the feasible region of the problem without any uncertainty, the specified Beta distribution and that of the ambiguous problem for $\epsilon = 0.1$ and $\epsilon = 0.05$. From the feasible region for

| Parameter | Value |
|-----------|-------|
| $\epsilon$ | 0.01  |
| $\sigma$  | 0.9   |
| $\phi$    | 2     |
| $D$       | 27    |
| $T$       | 5     |
| $x_{\text{min}}$ | 1.5 |
| $a$       | 3     |
| $b$       | 6     |
| $\mu$     | 4     |
| $d$       | 0.25  |

Table 3. Parameter values used for solving (35).
$\epsilon = 0.01$, however, it is clear that requiring that a low risk of violation results in a solution that is much worse in terms of tumor BED. Remarkable, also, is that the feasible region for the specified Beta distribution hardly changes with $\epsilon$ compared to the behavior under ambiguity. In fact, even for $\epsilon = 0.01$, the feasible region when we assume $\rho \sim \text{Beta}(6.6, 13.2)$ contains the ambiguous feasible region for $\epsilon = 0.01$. Note that the figure does illustrate how the shape of the feasible region changes with $\epsilon$: the feasibility of unbalanced solutions, i.e., solutions that administer a different dose in the two fractions, is impacted much more severely than that of balanced solutions.

We mention two related works on ambiguous chance constraints. Hansansanto, Roitch, Kuhn, & Wiesemann (2017) present a tractable framework for joint ambiguous chance constraints under a few simplifying conditions. In particular, they assume a concic, hence bounded, support, which is a key difference to our approach. Their approach is very powerful in settings for which an unbounded support makes sense, however, as they are able to elegantly deal with joint ambiguous chance constraints. Xie & Ahmed (2018), on the other hand, consider ambiguous chance constraints given a bounded support and moment information. Their assumptions on the ambiguity set do, however, exclude exact distributional information on nonlinear functions of the uncertain parameters, which we do assume in exact knowledge of the mean absolute deviation.

4. Conclusion and outlook

Tail probabilities are ubiquitous in probabilistic studies in many areas of science and application domains. Just as the original Chebyshev’s inequality for mean-variance ambiguity, we expect our novel tail bounds for mean-MAD ambiguity to find many applications.

In our search for tight bounds under limited information, we had to solve for the worst-case distribution and worst-case value of the expectation of the indicator function $1\{X \geq z\}$. In this paper the limited information was captured through ambiguity sets $\mathcal{P}_{\{a,b,d\}}$, $\mathcal{P}_{\{a,b,d,a\}}$, and it turned out that the combination of the non-convex indicator function with these ambiguity set gave rise to semi-infinite linear programs with easy, closed-form solutions.

In future work, we expect to find more such solvable classes, i.e. specific combinations of objective function (other than the indicator function) and ambiguity sets that together give rise to solvable linear programs and hence easy extremal distributions. In this way, one could try to sharpen the tail bounds by including more information (e.g. higher moments or percentiles), or to consider objective functions other than the tail probability. Our proof method based on solving the dual problem with piecewise-linear majorants is not tailor-made for the indicator function, and could potentially work for a much larger class of (measurable) objective functions, as shown in Section 3.2. Another direction we shall pursue is the application of the bounds to more complex, and possibly high-dimensional robust optimization problems. To do so, we shall leverage the connection with the quickly evolving field of DRO, as illustrated by examples in Section 3.3. Indeed, minmax and maxmin decision problems arise naturally, and the bounds and proof techniques can help in advancing that field.

Acknowledgements

We thank Vera van den Dool (ORTEC) for her ideas and efforts in the preliminary stages of this research and for sparking our interest in the topic. The research of the first author was funded by the Netherlands Organisation for Scientific Research (NWO) Research Talent [Grant 406.17.511].

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2021.12.010.

References

Asmussen, S., & Albrecher, H. (2010). Ruin probabilities (2nd ed.). World scientific.
Ben-Tal, A., & Hochman, E. (1972). More bounds on the expectation of a convex function of a random variable. Journal of Applied Probability, 9(4), 803–812.
Ben-Tal, A., & Hochman, E. (1976). Stochastic programs with incomplete information. Operations Research, 24(2), 336–347.
Bertsimas, D., & Popescu, I. (2005). Optimal inequalities in probability theory: A convex optimization approach. SIAM Journal on Optimization, 15(3), 780–804.
Bienaymé, I.-J. (1853). Considérations à l’appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés, imprimerie de Malher–Bucheler.
Casella, G., & Berger, R. L. (2002). Statistical inference (2nd ed.). Duxbury Thomson Learning.
Chebyshev, P. L. (1867). Des valeurs moyennes. Journal de Mathématiques Pures et Appliqués, 12, 177–184.
Chen, Y., Xu, M., & Zhang, Z. G. (2009). A risk-averse newsvendor model under the CVaR criterion. Operations Research, 57(4), 1040–1044.
Choi, S., Ruszczyński, A., & Zhao, Y. (2011). A multiproduct risk-averse newsvendor with law-invariant coherent measures of risk. Operations Research, 59(2), 346–364.
Cox, S. H. (1991). Bounds on expected values of insurance payments and option prices. Transactions of the Society of Actuaries, 45, 231–260.
Das, B., Dhara, A., & Natarajan, K. (2021). On the heavy-tail behavior of the distributionally robust newsvendor. Operations Research, 69(4), 1077–1099.
De Schepper, A., & Heijn, B. (1995). General restrictions on tail probabilities. Journal of Computational and Applied Mathematics, 64(1–2), 177–188.
De Vylder, F., & Goovaerts, M. J. (1982). Analytical upper bounds on stop-loss premiums. Insurance: Mathematics and Economics, 1(3), 163–175.
Gauss, C. F. (1821). Theoria combinationis observationum erroribus minimis obnoxiae, pars prior. SIAM: Philadelphia.
Gholipour, S., & Wiesemann, W. (2020). The distributionally robust chance-constrained vehicle routing problem. Operations Research, 68(3), 716–732.
Grechuk, B., Molyboha, A., & Zabarankin, M. (2010). Chebyshev inequalities with law invariant deviation measures. Probability in the Engineering and Informational Sciences, 24(1), 145–170.
Han, Q., Du, D., & Zuluaga, L. F. (2014). A risk-and-ambiguity-averse extension of the max-min newsvendor model. Operations Research, 62(3), 535–542.
Han, H., Shen, M., Topcu, U., Ovchadi, H., & Murray, R. M. (2015). Convex optimal uncertainty quantification. SIAM Journal on Optimization, 25(3), 1368–1387.
Hanusansanto, G. A., Roitch, V., Kuhn, D., & Wiesemann, W. (2017). Ambiguous joint chance constraints under mean and dispersion information. Operations Research, 65(3), 751–767.
Heijn, B. (1990). Best upper and lower bounds on modified stop loss premiums in case of known range, mode, mean and variance of the original risk. Insurance: Mathematics and Economics, 9(2–3), 207–220.
Isii, K. (1962). On sharpness of Chebyshev-type inequalities. Annals of the Institute of Statistical Mathematics, 14(1), 185–197.
Jansen, K., Haezendonck, J., & Goovaerts, M. (1986). Upper bounds on stop-loss premiums in case of known moments up to the fourth order. Insurance: Mathematics and Economics, 5(4), 315–334.
Joiner, M. C., & Van der Kogel, A. (2016). Basic clinical radiobiology: vol. 1. CRC press.
Kaar, R., Goovaerts, M., Dhaene, J., & Denuit, M. (2008). Modern actuarial risk theory: Using R (2nd ed.). Springer-Verlag.
Natarajan, K., Sim, M., & Uichanco, J. (2018). Asymmetry and ambiguity in newsvendor models. Management Science, 64, 3146–3167.
Natarajan, K., & Zhou, L. (2007). A mean–variance bound for a three-piece linear function. Probability in the Engineering and Informational Sciences, 21(4), 611–621.
Périsic, C., & Roels, C. (2008). Regret in the newsvendor model with partial information. Operations Research, 56(1), 188–203.
Perkiss, Z., Borfitt, T., Hong, T., Wolfgang, J., & Unkelbach, J. (2018). Derivation of mean dose tolerances for new fractionalisation schemes and treatment modalities. Physics in Medicine & Biology, 63(2), 025012.
Popescu, I. (2005). A semidefinite programming approach to optimal-moment bounds for convex classes of distributions. Mathematics of Operations Research, 30(2), 550–567.
Postek, K., Ben-Tal, A., Den Hertog, D., & Melenberg, B. (2018). Robust optimization with ambiguous stochastic constraints under mean and dispersion information. Operations Research, 66(3), 814–833.
Roginska, W. W. (1958). Moments of non-negative mass. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 245(1240), 1–27.
Rujeeraipiboon, N., Kuhn, D., & Wiesemann, W. (2018). Chebyshev inequalities for products of random variables. Mathematics of Operations Research, 43(3), 887–918.
Scarf, H. E. (1958). A min-max solution of an inventory problem. In K. J. Arrow, S. Karlin, & H. E. Scarf (Eds.), Studies in the mathematical theory of inventory and production. Stanford University Press.
Sharpe, A., Dentcheva, D., & Ruszczyński, A. (2009). Lectures on stochastic programming. Modeling and theory. SIAM.
Ten Eikelder, S., Adjari, A., Borfitt, T., & Den Hertog, D. (2019). Adjustable robust treatment-length optimization in radiation therapy arXiv:1906.12116.
Van Parys, B. P., Goulart, P. J., & Kuhn, D. (2016). Generalized Gauss inequalities via semidefinite programming. *Mathematical Programming, 156*(1–2), 271–302.
Vandenberghe, L., Boyd, S., & Comanor, K. (2007). Generalized Chebyshev bounds via semidefinite programming. *SIAM Review, 49*(1), 52–64.
Xie, W., & Ahmed, S. (2018). On deterministic reformulations of distributionally robust joint chance constrained optimization problems. *SIAM Journal on Optimization, 28*(2), 1151–1182.
Xin, L., & Goldberg, D. A. (2021). Time (in)consistency of multistage distributionally robust inventory models with moment constraints. *European Journal of Operational Research, 289*(3), 1127–1141.
Yue, J., Chen, B., & Wang, M.-C. (2006). Expected value of distribution information for the newsvendor problem. *Operations Research, 54*(6), 1128–1136.