A Chebyshev Inequality Based on Bounded Support and Mean Absolute Deviation

Ernst Roos, Ruud Brekelmans, Dick den Hertog
Tilburg University, Department of Econometrics and Operations Research
July 12, 2019

Abstract
The Chebyshev inequality is one of the most well-known results in classical probability theory. This inequality provides an upper bound on the tail probability of a random variable based on the first two moments of its distribution. While this upper bound is tight, it 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. In this paper, we provide an alternative tight lower and upper bound on the tail probability given a bounded support, mean and mean absolute deviation of the random variable. This result allows us to find convex reformulations of single and joint ambiguous chance constraints with right-hand side uncertainty, and left-hand side uncertainty with specific structural properties as well as safe approximations to a wider class of ambiguous chance constraints. Applications of such ambiguous chance constraints include radiation therapy and inventory control problems, both of which we illustrate numerically.

Keywords: Chebyshev inequality, probability bounds, distributionally robust optimization, chance constraints

1 Introduction

The Chebyshev inequality is a well-known result in classical probability theory that provides an upper bound on the tail probability of a random variable given limited distributional information: it only uses the first two moments of a distribution (Bienaymé, 1853; Chebyshev, 1867). Its popularity stems from this distribution-free nature, that is, one does not need a specified distribution or family of distributions for the Chebyshev inequality to hold.

*Corresponding author: e.j.roos@tilburguniversity.edu
Over the years, multivariate generalizations of the Chebyshev inequality have also been studied. Bertsimas and Popescu (2005) and Vandenberghe et al. (2007) both study such a generalization through formulating a convex optimization problem, given that the prescribed confidence region can be described by polynomial or linear and quadratic inequalities, respectively. Grechuk et al. (2010) on the other hand, provide closed-form variants of the Chebyshev inequality for different dispersion measures than the variance. Generalized versions of the Chebyshev inequality for products of random variables that focus on a one-sided inequality have also received some attention recently (Rujeerapaiboon et al., 2018).

While the Chebyshev inequality is tight, it 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 (Van Parys et al., 2016). A variant of the Chebyshev inequality that was already considered by Gauss (1821) restricts the distributions it considers to be unimodal. This yields an improvement by a factor $\frac{4}{3}$ 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).

All the abovementioned inequalities, however, still assume an unbounded support. In many practical applications, however, some information on the minimum and/or maximum of uncertain parameters is known. This is particularly true for applications that consider uncertain parameters that are known to be nonnegative, such as inventory management. We provide an upper and lower bound on a random variable’s tail probability based on a known, bounded support, mean and mean absolute deviation from the mean. These bounds are attained by a multitude of distributions: both discrete distributions with 3 elements in their support as well as mixed distributions with both a continuous and discrete part.

Recent advances in Distributionally Robust Optimization, which amongst others, also considers ambiguous chance constraints, indicate that knowledge on the support, mean and mean absolute deviation can lead to closed-form expressions for stochastic quantities such as the minimum and maximum expectation of a convex function (Postek et al., 2018). In particular, Postek et al. (2018) use results by Ben-Tal and Hochman (1972) who find tight upper and lower bounds on the expectation of convex function of a random variable. The results in this paper could be regarded as a variant of the results first described by Ben-Tal and Hochman (1972), who provide a tight lower and upper bound for the expectation of a random variable, based on the same distributional information we use to provide a tight lower and upper bound on the tail probability.

Ghosal and Wiesemann (2018) also present results on pessimistic probabilities by using a variance-at-risk reformulation and use them to solve a distributionally robust vehicle
routing problem. The ambiguity set they consider is highly similar to ours with a single important difference: they consider all distributions with a dispersion measure smaller than or equal to a specified value, which has both theoretical advantages and disadvantages. We have included a detailed comparison just after our results are presented.

Because we provide tight bounds on the tail probability, we are able to find convex reformulations of ambiguous chance constraints. In particular, we discuss the convex reformulation of ambiguous chance constraints in which the uncertainty is present as a single random variable. Besides that, our results can be applied to find safe approximations to ambiguous joint chance constraints with right-hand side uncertainty and ambiguous chance constraint based on multiple random variables. Hanususanto et al. (2017) provide a tractable framework for joint ambiguous chance constraints, given a few simplifying conditions. In particular, they assume the support to be conic and hence unbounded, which is a key difference from our approach. Applications that include ambiguous joint chance constraints include, but are not limited to radiation therapy optimization, inventory management, floodwater protection and production planning.

We also extend our results to an upper bound on the tail probability of the sum of random variables. This allows us to provide safe approximations to larger class of ambiguous chance constraints: those that involve more than a single uncertain parameter per underlying constraint.

We highlight the following main contributions of this work:

- We provide a tight upper and lower bound for the probability that a random variable exceeds a specified threshold under a known support, mean and mean absolute deviation, that is, a variant of the Chebyshev inequality given bounded support, mean and mean absolute deviation.

- We apply our results to ambiguous chance constraints and derive convex reformulations of such constraints under mild assumptions, as well as safe approximations for more general classes.

- We provide an upper bound for the tail probability of the sum of random variables, given a known support, mean and mean absolute deviation, and present a safe approximation for ambiguous chance constraints based on this upper bound.

- We numerically illustrate our results through applications in inventory management and radiation therapy.

The paper is organized as follows. In Section 2 we present and prove the above-mentioned theoretical tight upper and lower bounds on the tail probability of a random variable. Section 3 uses these results to find convex representations of ambiguous chance constraints.
constraints. Section 4 generalizes the upper bound to sums of random variables and provides safe approximations to ambiguous chance constraints based on it. Section 5 illustrates two applications of the developed theory: one in inventory management and one in radiation therapy.

2 Tight Chebyshev Inequalities

2.1 Upper Bound

In this section, we derive a tight upper bound for the probability that a random variable exceeds a specified threshold, given a specified bounded support, mean and mean absolute deviation. Mathematically, this means we consider the following ambiguity set of probability measures:

\[ P = \{ P : B \to [0, 1] \mid P[X \in [a, b]] = 1, \ E_P[X] = \mu, \ E_P[|X - \mu|] = d \}, \]  

(1)

where \(B\) is the Borel \(\sigma\)-algebra and \(a, b, \mu, d \in \mathbb{R}\) are given parameters that describe all known properties of the distribution. The following theorem then gives a tight upper bound on a random variable \(X\) with distribution \(P \in P\) exceeding a threshold \(\tau \in \mathbb{R}\).

**Theorem 1.** The maximum probability of a random variable \(X\) exceeding a threshold \(\tau\) when its distribution resides in \(P\) is given by:

\[
\sup_{P \in P} P[X > \tau] = \sup_{P \in P} P[X \geq \tau] = \begin{cases} 
1 & \text{if } \tau \leq a \\
\min \left\{ 1, \frac{\mu - a}{\tau - a} - \frac{d(b - \tau)}{2(\tau - a)(b - \mu)} \right\} & \text{if } \tau \in (a, \mu] \\
\min \left\{ \frac{d}{2(\tau - \mu)}, 1 - \frac{d}{2(\mu - a)} \right\} & \text{if } \tau \in (\mu, b) 
\end{cases}
\]  

(2)

\[ \sup_{P \in P} P[X > b] = 0 \]

\[ \sup_{P \in P} P[X \geq b] = \frac{d}{2(b - \mu)}. \]

**Proof.** We first prove the result for \(P[X \geq \tau]\). It it obvious that for any \(\tau \leq a\), it holds that

\[ \sup_{P \in P} P[X \geq \tau] = 1, \]

as \(a\) is the lower bound of the support of \(X\). We now discuss the case where \(\tau \in (a, \mu]\). We start by defining the set of distributions that maximize \(P(X \geq \tau)\), that is, we define

\[ \mathcal{P}^* = \left\{ P^* \in \mathcal{P} \mid P^*[X \geq \tau] = \max_{P \in P} P[X \geq \tau] \right\}. \]

We note that we use the maximum here instead of the supremum to indicate that \(\mathcal{P}^*\) is nonempty, i.e., there are distributions that attain this maximum and we note that
this does not hold when we consider $\mathbb{P}[X > \tau]$. We show that there exists a probability measure $\mathbb{P}^* \in \mathcal{P}^*$ of the form

$$
\mathbb{P}^* = p_1 \cdot \delta_a + p_2 \cdot \delta_\tau + p_3 \cdot \delta_\mu + p_4 \cdot \delta_b,
$$

for some $p_1, p_2, p_3, p_4 \in [0, 1]$ and subsequently derive closed-form expressions for these unknowns. Here, for any $x \in \mathbb{R}$, we denote the Dirac distribution that places all probability mass on $x$ by $\delta_x$.

More specifically, we show that given a probability measure in $\mathcal{P}^*$ with a positive probability on $(\tau, \mu)$ and/or $(\mu, b)$, there exists a probability measure in $\mathcal{P}^*$ with zero probability on that interval. Moreover, we first show that a probability measure in $\mathcal{P}^*$ with a positive probability on $(a, \tau)$ cannot exist.

Let $\hat{\mathbb{P}} \in \mathcal{P}^*$ be such that $q := \hat{\mathbb{P}}[X \in (a, \tau)] > 0$. We denote $\hat{\mu} := \frac{1}{q} \int_{(a,\tau)} x \, d\hat{\mathbb{P}}(x)$ and define a new probability measure $\bar{\mathbb{P}}$ by

$$
\bar{\mathbb{P}}[X \in C] = 1_C(a) \cdot \left( q \cdot \frac{\tau - \hat{\mu}}{\tau - a} + \hat{\mathbb{P}}[X = a] \right) + 1_C(\tau) \cdot q \cdot \frac{\hat{\mu} - a}{\tau - a} + \hat{\mathbb{P}}[X \in C \cap (\tau, b)].
$$

It is important to note that

$$
q \cdot \frac{\tau - \hat{\mu}}{\tau - a} + q \cdot \frac{\hat{\mu} - a}{\tau - a} = q,
$$

and thus

$$
\int_{[a,b]} d\bar{\mathbb{P}}(x) = 1.
$$

Moreover, it holds that

$$
\int_{[a,b]} x \, d\bar{\mathbb{P}}(x) = q \cdot \frac{\tau - \hat{\mu}}{\tau - a} \cdot a + q \cdot \frac{\hat{\mu} - a}{\tau - a} \cdot \tau + \int_{\{a\} \cup [\tau,b]} x \, d\hat{\mathbb{P}}(x)
$$

$$
= q \cdot \frac{(\tau - a) \hat{\mu}}{\tau - a} + \int_{\{a\} \cup [\tau,b]} x \, d\hat{\mathbb{P}}(x)
$$

$$
= q \hat{\mu} + \int_{\{a\} \cup [\tau,b]} x \, d\hat{\mathbb{P}}(x)
$$

$$
= q \int_{(a,\tau)} x \, d\hat{\mathbb{P}}(x) + \int_{\{a\} \cup [\tau,b]} x \, d\hat{\mathbb{P}}(x)
$$

$$
= \int_{[a,b]} x \, d\bar{\mathbb{P}}(x) = \mu,
$$

5
and
\[
\begin{align*}
\int_{[a,b]} |x - \mu| \, d\hat{P}(x) &= q \cdot \frac{\tau - \hat{\mu}}{\tau - a} \cdot (\mu - a) + q \cdot \frac{\hat{\mu} - a}{\tau - a} \cdot (\mu - \tau) \\
&\quad + \int_{(a,\tau]} |x - \mu| \, d\tilde{P}(x) \\
&= q \cdot \frac{(\tau - a)(\mu - \hat{\mu})}{\tau - a} + \int_{\{a\} \cup [a,\tau]} |x - \mu| \, d\tilde{P}(x) \\
&= q (\mu - \hat{\mu}) + \int_{\{a\} \cup [a,\tau]} |x - \mu| \, d\tilde{P}(x) \\
&= \mu \int_{(a,\tau)} d\tilde{P}(x) - \int_{(a,\tau]} x \, d\tilde{P}(x) + \int_{\{a\} \cup [a,\tau]} |x - \mu| \, d\tilde{P}(x) \\
&= \int_{[a,b]} |x - \mu| \, d\tilde{P}(x) = d,
\end{align*}
\]
and thus \(\tilde{P} \in \mathcal{P}\). We also note that because \(q > 0\) implies that \(\hat{\mu} > a\), it holds that
\[
\int_{[a,b]} d\tilde{P}(x) = \int_{[a,b]} d\tilde{P}(x) + q \cdot \frac{\hat{\mu} - a}{\tau - a} > \int_{[a,b]} d\tilde{P}(x),
\]
which is a contradiction with \(\tilde{P} \in \mathcal{P}^*\). Therefore, there cannot exist a probability measure \(\hat{P} \in \mathcal{P}^*\) such that \(\int_{(a,\tau)} d\hat{P}(x) > 0\).

Now, consider a probability measure \(\hat{P} \in \mathcal{P}^*\) such that \(q := \int_{(\tau,\mu)} d\hat{P}(x) > 0\). We once again define \(\hat{\mu} := \int_{[\tau,\mu)} x \, d\hat{P}(x)\) and construct a new probability measure \(\bar{P}\) by
\[
\bar{P}[X \in C] = 1_{C}(\tau) \cdot q \cdot \frac{\mu - \hat{\mu}}{\mu - \tau} + 1_{C}(\mu) \cdot q \cdot \frac{\hat{\mu} - \tau}{\mu - \tau} + \hat{P}[X \in C \setminus (\tau,\mu)].
\]
Analogous to the previous case it follows that \(\bar{P} \in \mathcal{P}\) and that
\[
\bar{P}[X \geq \tau] = \hat{P}[X \geq \tau],
\]
and thus \(\bar{P} \in \mathcal{P}^*\). Similarly, we can show that from a probability measure \(\hat{P} \in \mathcal{P}^*\) such that \(q := \int_{(\mu,b)} d\hat{P}(x) > 0\), we can construct a probability measure \(\bar{P} \in \mathcal{P}^*\) such that \(\int_{(\mu,b)} d\hat{P}(x) = 0\). Together, this all implies that if \(\mathcal{P}^*\) is nonempty, a probability measure \(\mathcal{P}^* \in \mathcal{P}^*\) of the form
\[
\mathcal{P}^* = p_1 \cdot \delta_a + p_2 \cdot \delta_\tau + p_3 \cdot \delta_\mu + p_4 \cdot \delta_b,
\]
must exist. In other words, if we can solve the following simple linear optimization problem, we find a probability measure that maximizes \(P(X \geq \tau)\): 

\[
\begin{align*}
\max_{p_1,p_2,p_3,p_4} & \quad p_2 + p_3 + p_4 \\
\text{s.t.} & \quad p_1 + p_2 + p_3 + p_4 = 1 \\
& \quad p_1 a + p_2 \tau + p_3 \mu + p_4 b = \mu \\
& \quad p_1 (\mu - a) + p_2 (\mu - \tau) + p_4 (b - \mu) = d \\
& \quad p_1, p_2, p_3, p_4 \geq 0.
\end{align*}
\]
By careful inspection of (3a)-(3c), one finds that any solution to (3) must have
\[ p_4 = \frac{d}{2(b - \mu)}. \]

Therefore, it is equivalent to simply focus on maximizing \( p_2 + p_3 \). More inspection yields that
\[ p_2 = \frac{\mu - a}{\tau - a} - \frac{d(b - a)}{2(b - \mu)(\tau - a)} - \frac{\mu - a}{\tau - a} p_3, \]
which implies that we can focus on choosing \( p_3 \) such that
\[ \left[ 1 - \frac{\mu - a}{\tau - a} \right] p_3 = \frac{\tau - \mu}{\tau - a} p_3, \]
is maximized. Since \( \frac{\tau - \mu}{\tau - a} < 0 \), it must be optimal to choose \( p_3 = 0 \) and thus
\[ p_2 = \frac{\mu - a}{\tau - a} - \frac{d(b - a)}{2(b - \mu)(\tau - a)}. \]

From this, it readily follows from (3a) that
\[ p_1 = \frac{d(b - \tau)}{2(\tau - a)(b - \mu)} - \frac{\mu - \tau}{\tau - a}. \]

(4)

Note that we have ignored (3d) so far and that for any \( d < \frac{2(\mu - \tau)(b - \mu)}{b - \tau} \), (4) would imply \( p_1 < 0 \), which contradicts (3d). This does, however, imply that there exists a solution with \( p_1 = 0 \), such that the optimal value of the linear optimization problem is clearly 1, that is, there exists a probability measure such that \( \mathbb{P}[X \geq \tau] = 1. \)

For \( d \geq \frac{2(\mu - \tau)(b - \mu)}{b - \tau} \) we thus find an explicit probability measure \( \mathbb{P}^* \in \mathcal{P}^* \) defined by
\[
\mathbb{P}^* = \left[ \frac{d(b - \tau)}{2(\tau - a)(b - \mu)} - \frac{\mu - \tau}{\tau - a} \right] \cdot \delta_a + \left[ \frac{\mu - a}{\tau - a} - \frac{d(b - a)}{2(b - \mu)(\tau - a)} \right] \cdot \delta_\tau + \frac{d}{2(b - \mu)} \cdot \delta_b,
\]
and for \( d < \frac{2(\mu - \tau)(b - \mu)}{b - \tau} \) we find that
\[
\mathbb{P}^* = \left[ 1 - \frac{d}{2(b - \mu)} \right] \delta_\tau + \frac{d}{2(b - \mu)} \delta_b \in \mathcal{P}^*,
\]
which shows that
\[ \sup_{\mathbb{P} \in \mathcal{P}((\mu, \tau, d))} \mathbb{P}[X \geq \tau] = \min \left\{ 1, \frac{\mu - a}{\tau - a} - \frac{d(b - \tau)}{2(b - \mu)(\tau - a)} \right\}. \]

We now prove the result for \( \tau \in (\mu, b] \). We will once again use the set of optimal probability measures \( \mathcal{P}^* \), and show that there exists a probability measure in this set of the form
\[ \mathbb{P}^* = p_1 \cdot \delta_a + p_2 \cdot \delta_\mu + p_3 \cdot \delta_\tau + p_4 \cdot \delta_b, \]
for some $p_1, p_2, p_3, p_4 \in [0, 1]$.

Similar to the analysis above, one can show that given a probability measure in $\mathcal{P}^*$ with a positive probability on $(a, \mu)$ and/or $(\tau, b)$, there exists a probability measure in $\mathcal{P}^*$ with zero probability on that interval. Furthermore, one can show that a probability measure in $\mathcal{P}^*$ with a positive probability on $(\mu, \tau)$ cannot exist. For sake of brevity we omit the detailed proof of this as it is highly similar to the analysis above.

One thus arrives at the following linear optimization problem:

$$
\max_{p_1, p_2, p_3, p_4} \quad p_3 + p_4 \\
\text{s.t.} \quad p_1 + p_2 + p_3 + p_4 = 1 \\
\quad p_1a + p_2\mu + p_3\tau + p_4b = \mu \\
\quad p_1(\mu - a) + p_3(\tau - \mu) + p_4(b - \mu) = d \\
\quad p_1, p_2, p_3, p_4 \geq 0.
$$

(5a) (5b) (5c) (5d)

Inspection of the equality constraints in this problem yields

$$
p_1 = \frac{d}{2(\mu - a)},
$$

for any feasible solution. Moreover, we find that

$$
p_3 + \frac{b - \mu}{\tau - \mu}p_4 = \frac{d}{2(\tau - \mu)},
$$

and thus

$$
p_3 + p_4 = \frac{d}{2(\tau - \mu)} + p_4 \left[ 1 - \frac{b - \mu}{\tau - \mu} \right],
$$

which is maximized by setting $p_4 = 0$, because $1 - \frac{b - \mu}{\tau - \mu} < 0$. This does additionally imply that

$$
p_2 = 1 - \frac{d}{2(\mu - a)} - \frac{d}{2(\tau - \mu)},
$$

which is not necessarily nonnegative. In fact, whenever $d > \frac{2(\tau - \mu)(\mu - a)}{\tau - a}$, this would mean $p_2 < 0$, which is clearly infeasible. In this case, it is optimal to choose $p_4$ as small as possible, that is, choose it such that $p_2 = 0$:

$$
p_4 = \frac{d(\tau - a)}{2(b - \tau)(\mu - a)} - \frac{\tau - \mu}{b - \tau}.
$$

Analyzing the optimal values of the linear optimization problem we thus find that the optimal value of (5) is given by

$$
\min \left\{ \frac{d}{2(\tau - \mu)}, 1 - \frac{d}{2(\mu - a)} \right\}.
$$
A Chebyshev inequality based on mean absolute deviation
Roos et al.

which shows that
\[
\sup_{P \in P} \mathbb{P}[X \geq \tau] = \min \left\{ \frac{d}{2(\tau - \mu)}, 1 - \frac{d}{2(\mu - a)} \right\},
\]
when \( \tau \in (\mu, b] \).

To show that \( \sup_{P \in P} \mathbb{P}[X > \tau] = \sup_{P \in P} \mathbb{P}[X \geq \tau] \) for \( \tau \in [a, b] \), we note that
\[
\sup_{P \in P} \mathbb{P}[X \geq \tau + \epsilon] = \min \left\{ 1, \frac{\mu - a}{\tau + \epsilon - a} - \frac{d(b - (\tau + \epsilon))}{2(b - \mu)(\tau + \epsilon - a)} \right\},
\]
for any \( \tau \in [a, \mu] \) and \( \epsilon \in (0, \mu - \tau] \) and
\[
\sup_{P \in P} \mathbb{P}[X \geq \tau + \epsilon] = \min \left\{ \frac{d}{2(\tau + \epsilon - \mu)}, 1 - \frac{d}{2(\mu - a)} \right\},
\]
for any \( \tau \in [\mu, b] \) and \( \epsilon \in (0, b - \tau) \). We thus find that
\[
\sup_{P \in P} \mathbb{P}[X > \tau] = \lim_{\epsilon \downarrow 0} \sup_{P \in P} \mathbb{P}[X \geq \tau + \epsilon]
\]
\[
= \begin{cases} 
\min \left\{ 1, \frac{\mu - a}{\tau - a} - \frac{d(b - \tau)}{2(b - \mu)(\tau - a)} \right\} & \text{if } \tau \in [a, \mu] \\
\min \left\{ \frac{d}{2(\tau - \mu)}, 1 - \frac{d}{2(\mu - a)} \right\} & \text{if } \tau \in (\mu, b),
\end{cases}
\]
since for \( \tau = \mu \)
\[
\lim_{\epsilon \downarrow 0} \min \left\{ \frac{d}{2(\tau + \epsilon - \mu)}, 1 - \frac{d}{2(\mu - a)} \right\} = 1 - \frac{d}{2(\mu - a)}
\]
\[
= \frac{\mu - a}{\tau - a} - \frac{d(b - \tau)}{2(b - \mu)(\tau - a)}
\]
\[
= \min \left\{ 1, \frac{\mu - a}{\tau - a} - \frac{d(b - \tau)}{2(b - \mu)(\tau - a)} \right\}.
\]

For \( \tau = b \), the suprema clearly cannot be equal, as \( \sup_{P \in P} \mathbb{P}[X > b] = 0 \) by definition of the support, while
\[
\sup_{P \in P} \mathbb{P}[X \geq b] = \frac{d}{2(b - \mu)}.
\]

Having discussed \( \tau \in [a, \mu) \), \( \tau \in (\mu, b) \) and \( \tau = b \), the proof is complete. \( \square \)

The result presented in Theorem 1 has a couple of noteworthy characteristics:

1. If the support is symmetric around \( \mu \), that is, \( b - \mu = \mu - a \), then the worst-case probability is at least \( \frac{1}{2} \) for \( \tau \in [a, \mu] \).

2. Both minima in (2) have a threshold for \( d \) that determines the value: \( \frac{2(b - \mu)(\mu - \tau)}{b - \tau} \) and \( \frac{2(\tau - \mu)(\mu - a)}{\tau - a} \) for \( \tau \in [a, \mu] \) and \( \tau \in (\mu, b) \), respectively.

3. The upper bound is continuous in \( \tau = \mu \).
4. The upper bound for $\tau \in (\mu, b)$ is increasing for $d < \frac{2(\tau - \mu)(\mu - a)}{\tau - a}$ and decreasing for larger values of $d$.

This last observation in particular is interesting as one might expect that such a maximum probability is increasing in the mean absolute deviation (or any other dispersion measure). Moreover, if the mean absolute deviation is unknown, this means that the worst-case probability based on only the support and mean is given by the result of Theorem 1 for $d = \frac{2(\tau - \mu)(\mu - a)}{\tau - a}$.

**Corollary 1.** The maximum probability of a random variable $X$ exceeding a threshold $\tau$ when its distribution resides in

$$
P_\mu = \{ P : B \to [0, 1] | P [X \in [a, b]] = 1, \ E_P [X] = \mu \} ,$$

is given by

$$
\sup_{P \in P_\mu} [X > \tau] = \sup_{P \in P_\mu} [X \geq \tau] = \begin{cases} 
1 & \text{if } \tau \leq \mu \\
\frac{\mu - a}{\tau - a} & \text{if } \tau \in (\mu, b) 
\end{cases}
$$

$$
\sup_{P \in P_\mu} [X > b] = 0 \quad \sup_{P \in P_\mu} [X \geq b] = \frac{\mu - a}{b - a}.
$$

**Proof.** A distribution in $P_\mu$ can have a mean absolute deviation at most equal to $\delta = \frac{2(b - \mu)(\mu - a)}{b - a}$ (Postek et al., 2018), and therefore we know that

$$
\sup_{P \in P_\mu} [X \geq \tau] = \begin{cases} 
\max_{d \in [0, \delta]} \left\{ \min \left\{ 1, \frac{\mu - a}{\tau - a} - \frac{d(b - \tau)}{2(\tau - a)(b - \mu)} \right\} \right\} & \text{if } \tau \leq \mu \\
\max_{d \in [0, \delta]} \left\{ \min \left\{ \frac{d}{2(\tau - a)}, 1 - \frac{d}{2(\mu - a)} \right\} \right\} & \text{if } \tau \in (\mu, b) 
\end{cases}
$$

Here, the result for $\tau \in (\mu, b]$ follows from the observation that $\min \left\{ \frac{d}{2(\tau - a)}, 1 - \frac{d}{2(\mu - a)} \right\}$ is increasing for $d < \frac{2(\tau - \mu)(\mu - a)}{\tau - a}$ and decreasing for larger values of $d$, and its maximum is thus at that threshold.

We similarly find

$$
\sup_{P \in P_\mu} [X > \tau] = \begin{cases} 
1 & \text{if } \tau \leq \mu \\
\frac{\mu - a}{\tau - a} & \text{if } \tau \in (\mu, b) \\
0 & \text{if } \tau \geq b.
\end{cases}
$$

Just as the result of Theorem 1 can be interpreted as the probabilistic version of the results of Ben-Tal and Hochman (1972), Corollary 1 is the probabilistic version of the Edmundson-Madansky bound (Madansky, 1959).
2.2 Lower Bound

For a tight lower bound on $P[X > \tau]$, we can use the results obtained above on a slightly altered version of the input. The exact specification of this idea is formalized in the following theorem.

**Theorem 2.** The minimum probability of a random variable $X$ exceeding a threshold $\tau$ is given by:

$$
\inf_{P \in \mathcal{P}} P[X \geq \tau] = \inf_{P \in \mathcal{P}} P[X > \tau] = \begin{cases} 
\max \left\{ 1 - \frac{d}{2(\mu - \tau)}, \frac{d}{2(b - \mu)} \right\} & \text{if } \tau \in (a, \mu) \\
\max \left\{ 0, \frac{\mu - \tau}{b - \tau} + \frac{d(\tau - a)}{2(b - \tau)(\mu - a)} \right\} & \text{if } \tau \in [\mu, b) \\
0 & \text{if } \tau \geq b.
\end{cases} \quad (6)
$$

Proof. We reformulate the infimum as follows:

$$
\inf_{P \in \mathcal{P}} \mathbb{P}(X > \tau) = 1 - \sup_{P \in \mathcal{P}} \mathbb{P}(X \leq \tau) = 1 - \sup_{P \in \mathcal{P}} \mathbb{P}(X \geq \tilde{\tau}),
$$

where

$$
\mathcal{P} = \left\{ P : \mathcal{B} \to [0, 1] \mid \mathbb{P}(X \in [\tilde{a}, \tilde{b}]) = 1, \ \mathbb{E}_P[X] = \mu, \ \mathbb{E}_P[|X - \mu|] = d \right\},
$$

and $\tilde{a} = 2\mu - b$, $\tilde{b} = 2\mu - a$ and $\tilde{\tau} = 2\mu - \tau$. Plugging in the results from Theorem 1 for $\tau \in (a, b)$ then yields (6). Similarly, the result for $\inf_{P \in \mathcal{P}} \mathbb{P}[X \geq \tau]$ can be obtained.

To illustrate the results in Theorems 1 and 2, Figure 1 shows limiting distributions that attain the bounds we derive for a variety of parameter values. The distribution corresponding to the maximum (or minimum) probabilities are shown in red (or green) and its values can be found on the left (or right) axis.

Figures 1a and 1b depict an example where $\tau \in (\mu, b)$, with different values for $d$: $\frac{1}{4}$ and 1, respectively. For $d = \frac{1}{4}$, the distributions are in fact very similar, with the difference that some probability mass is just below or above $\tau$. When $d = 1$, on the other hand, the distributions are very different. We observe the same difference when $\tau \in (a, \mu)$. The idea of the minimum being similar to the maximum when all parameters are reflected in $\mu$ is also visible when comparing Figures 1b and 1d. The worst-case distribution in 1b is similar to the best-case distribution in 1d and vice versa.
A Chebyshev inequality based on mean absolute deviation

Figure 1: Examples of distributions that attain the lower and upper bound probability of exceeding the threshold $\tau$. 
\section{2.3 Comparison to Existing Bounds}

Closely related to our results is the discussion by Ghosal and Wiesemann (2018) in Section 4.1. In particular, they consider, amongst others, an ambiguity set given by

$$\mathcal{P} = \{ \mathbb{P} : \mathcal{B} \to [0,1] \mid \mathbb{P}[X \in [a,b]] = 1, \ \mathbb{E}_\mathbb{P}[X] = \mu, \ \mathbb{E}_\mathbb{P}[|X - \mu|] \leq d \}.$$ 

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

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[X \geq \tau] = 1.$$ 

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

Clearly, because of the above observations, the theoretical maximum of $\mathbb{P}[X > \tau]$ has a much simpler closed-form solution than (2) for the ambiguity set $\mathcal{P}$. A big downside is that many of the extra distributions contained in $\mathcal{P}$ but not in $\mathcal{P}$ might be 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 $\tau$ in particular, using $\mathcal{P}$ can lead to an overestimation of the maximum value of $\mathbb{P}[X > \tau]$. The observation that the maximum value of $\mathbb{P}[X > \tau]$ 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 $\mathbb{P}[X > \tau]$.

Comparing the result of Theorem 1 directly to the classic Chebyshev inequality is also possible, but somewhat unfair as the classic Chebyshev inequality is two-sided:

$$\mathbb{P}[|X - \mu| \geq \tau \sigma] \leq \frac{1}{\tau^2}.$$ 

We instead compare our result to a tight one-sided variant of the Chebyshev inequality, also referred to as Cantelli’s inequality (Chebyshev, 1867):

$$\mathbb{P}[X \geq \tau] \leq \frac{\sigma^2}{\sigma^2 + (\tau - \mu)^2}, \quad (7)$$

where $X$ has mean $\mu$ and variance $\sigma^2$, and $\tau > \mu$. Since we assume the mean absolute deviation to be known, but not the variance, some relation between these two quantities is needed to be able to make a comparison. In particular, we will use that

$$\frac{d^2}{4\beta(1-\beta)} \leq \sigma^2 \leq \frac{d(b-a)}{2}, \quad (8)$$
A Chebyshev inequality based on mean absolute deviation
Roos et al.

Figure 2: A comparison of the bound in Theorem 1 with Cantelli’s bound for three different values of $\sigma$ with the parameter values $a = -1$, $\mu = 0$, $b = 1$ and $d = \frac{1}{4}$.

where $\beta = \mathbb{P}[X > \mu]$ (Ben-Tal and Hochman, 1985). We note that this also implies $d \leq \sigma$. 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 (8). Figure 2 illustrates this comparison for a simple numerical example with the following parameters: $a = -1$, $\mu = 0$, $b = 1$, $d = \frac{1}{4}$. We consider three values for $\sigma$: $\sigma = d = \frac{1}{4}$, $\sigma = \frac{1}{3}$ and $\sigma = \sqrt{\frac{d(b-a)}{2}} = \frac{1}{2}$.

Figure 2 gives rise to a number of interesting observations. First of all, we note that the flat area in the blue line corresponds to the values of $\tau$ such that

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

which corresponds to all $\tau \leq \tau^* := \mu + \frac{d(\mu - a)}{2(\mu - a) - d}$. Moreover, we note that for $\sigma = d$, which corresponds to the red line in Figure 2, Cantelli’s bound is lower than (2) for all $\tau^* \leq \tau \leq b$. This is true for all parameters as:

$$\frac{d^2}{d^2 + (\tau - \mu)^2} = \frac{d^2}{d^2 + (\tau - \mu)^2 - 2d(\tau - \mu) + 2d(\tau - \mu)}$$

$$= \frac{d^2}{(d - (\tau - \mu))^2 + 2d(\tau - \mu)}$$

$$\leq \frac{d^2}{2d(\tau - \mu)}$$

$$= \frac{d}{2(\tau - \mu)}.$$  

In particular, for $\sigma = d$ Cantelli’s bound and (2) always coincide at $\tau = \mu + d$, since:

$$\frac{d^2}{d^2 + d^2} = \frac{1}{2} = \frac{d}{2d}.$$
A Chebyshev inequality based on mean absolute deviation
Roos et al.

Figure 3: An illustration of $\hat{\tau}$, $\tau^*$, $\underline{\tau}$ and $\overline{\tau}$ for the parameter values $a = -1$, $\mu = 0$, $b = 1$ and $d = 0.25$.

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

For intermediate values of $\sigma$, we observe behavior similar to the orange line in Figure 2. More specifically, we find that (2) is lower than Cantelli’s bound for all $\tau$ in the two intervals $[0, \hat{\tau}]$ and $[\underline{\tau}, \tau^*]$, with the three boundaries given by

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

$$\underline{\tau} = \frac{\sigma^2}{d} - \sigma\sqrt{\frac{\sigma^2}{d^2} - 1}$$

$$\tau^* = \min\left\{b, \frac{\sigma^2}{d} + \sigma\sqrt{\frac{\sigma^2}{d^2} - 1}\right\}.$$  

Note that for some $\sigma$, such as $\sigma = \sqrt{\frac{d(b-a)}{2}}$ in Figure 2, it holds that $\hat{\tau} \geq \underline{\tau}$, that is, (2) is lower than Cantelli’s bound for all $\tau \in [0, \overline{\tau}]$. To visually clarify all boundaries discussed above, Figure 3 only shows Cantelli’s bound for $\sigma = 0.27$ and marks $\tau^*$, $\hat{\tau}$, $\underline{\tau}$ and $\overline{\tau}$.

3 Reformulation of Chance Constraints

In this section, we use the tight bounds derived in Section 2 to reformulate ambiguous chance constraints to convex constraints. In particular, we present a reformulation of
chance constraints with right-hand side uncertainty in Theorem 3 and 5, and a safe approximation to joint chance constraints with right-hand side uncertainty, where the uncertain parameters are independent in Theorem 4. We also treat right-hand side uncertainty with identical uncertain parameters in Theorem 6. Last but not least, we consider a single ambiguous chance constraint with left-hand side uncertainty dependent on a single uncertain parameter in Theorem 7.

We first present a convex reformulation of an ambiguous chance constraint based on a convex inequality with right-hand side uncertainty.

**Theorem 3.** Let $g : \mathbb{R}^n \to \mathbb{R}$, $b \in \mathbb{R}$ and let $\zeta$ be an 1-dimensional random variable whose distribution lies in the ambiguity set

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

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

$$
\inf_{P \in \mathcal{P}} P[g(x) \leq b - \zeta] \geq 1 - \epsilon, \quad (9)
$$

if and only if

$$
g(x) + \min\left\{1, \frac{d}{2\epsilon}\right\} \leq b. \quad (10)
$$

**Proof.** We first rewrite (9) to

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

From Theorem 1 and the fact that $\epsilon < \frac{1}{2}$ we know that it must hold that $b - g(x) > E[\zeta] = 0$. Given that requirement, we know by Theorem 1 that

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

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

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

which can easily be combined with the former as

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

Because $\frac{d}{2\epsilon} > 0$, we find that the requirement $b - g(x) > 0$ is redundant, and thus (9) is equivalent to (10). □
A similar reasoning to that in Theorem 3 can be applied to joint chance constraints with independent right-hand side uncertainty. Here, because of the use of the support information of the random variable, we do not provide an exact reformulation, but a safe approximation instead.

**Theorem 4.** Let $g_i : \mathbb{R}^n \to \mathbb{R}$, $b_i \in \mathbb{R}$ for $i = 1, \ldots, m$ and let $\zeta$ be an $m$-dimensional random variable whose distribution lies in the ambiguity set

$$
\mathcal{P} = \{ \mathbb{P} : \mathbb{P}[\zeta_i \in [-1, 1]] = 1, \ \mathbb{E}[\zeta_i] = 0, \ \mathbb{E}[|\zeta_i|] = d_i \ \forall i, \ \zeta_i \perp \zeta_j \ \forall i \neq j \},
$$

for some $d \in \mathbb{R}^m$ such that $d_i \in [0, 1]$ for all $i$. Let $\epsilon \in (0, \frac{1}{2})$ and $\mathcal{I}$ be the set of indices $i$ such that $\frac{d_i}{2\epsilon} \leq 1$. For any $x \in \mathbb{R}^n$ it holds that

$$
\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[g_i(x) \leq b_i - \zeta_i] \geq 1 - \epsilon,
$$

if

$$
\sum_{i \in \mathcal{I}} \log \left[ 1 - \frac{d_i}{2 (b_i - g_i(x))} \right] \geq \log [1 - \epsilon] \tag{12a}
$$

$$
g_i(x) + 1 \leq b_i \quad i \notin \mathcal{I} \tag{12b}
$$

$$
g_i(x) < b_i \quad i \in \mathcal{I}, \tag{12c}
$$

which is a convex set of constraints if all $g_i$ are convex functions.

**Proof.** Using the pairwise independence of $\zeta$ we find

$$
\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[g_i(x) \leq b_i - \zeta_i] = \inf_{\mathbb{P} \in \mathcal{P}} \prod_{i=1}^{m} \mathbb{P}[g_i(x) \leq b_i - \zeta_i]
$$

$$
= \prod_{i=1}^{m} \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[g_i(x) \leq b_i - \zeta_i]
$$

$$
= \prod_{i=1}^{m} \left[ 1 - \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[g_i(x) > b_i - \zeta_i] \right]
$$

$$
= \prod_{i=1}^{m} \left[ 1 - \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\zeta_i > b_i - g_i(x)] \right]. \tag{13}
$$

From this, it readily follows that it must at least hold that $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\zeta_i > b_i - g_i(x)] \leq \epsilon$ for all $i$. From Theorem 1 we consequently know that it must hold that $b_i - g_i(x) > \mathbb{E}[\zeta] = 0$ for all $i$, as we know that $\epsilon < \frac{1}{2}$. Given that $b_i - g_i(x) > 0$, we know that

$$
\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\zeta_i > b_i - g_i(x)] = \begin{cases} 
\min \left\{ \frac{d_i}{2(b_i - g_i(x))}, \ 1 - \frac{d_i}{2} \right\} & \text{if } b_i - g_i(x) < 1 \\
0 & \text{if } b_i - g_i(x) \geq 1. \end{cases} \tag{14}
$$

Since $1 - \frac{d_i}{2} \geq \frac{1}{2}$ it follows from $\epsilon < \frac{1}{2}$ that it must hold for any feasible solution $x$ that

$$
\min \left\{ \frac{d_i}{2(b_i - g_i(x))}, \ 1 - \frac{d_i}{2} \right\} = \frac{d_i}{2(b_i - g_i(x))} \quad i = 1, \ldots, m.
$$
Moreover, we note that $\frac{d_i}{2(b_i - g_i(x))} \leq \epsilon$ is equivalent to $b_i - g_i(x) \geq \frac{d_i}{2\epsilon}$ and thus if $\frac{d_i}{2\epsilon} \geq 1$, imposing this is overly restrictive, as we know from (14) that the worst-case probability of violation is 0, not $\epsilon$ in that situation. For all such $i$, we thus simply require that $\epsilon$.

Given this analysis, we find that (13) is equal to

$$\prod_{i \in \mathcal{I}} \left[ 1 - \sup_{P \in \mathcal{P}} \mathbb{P}[\zeta_i > b_i - g_i(x)] \right] \leq \prod_{i \in \mathcal{I}} \left[ 1 - \frac{d_i}{2(b_i - g_i(x))} \right],$$

and thus (11) holds if

$$\begin{cases} 
\sum_{i \in \mathcal{I}} \log \left[ 1 - \frac{d_i}{2(b_i - g_i(x))} \right] \geq \log \left[ 1 - \epsilon \right] \\
g_i(x) + 1 \leq b_i & i \notin \mathcal{I} \\
g_i(x) < b_i & i \in \mathcal{I}.
\end{cases}$$

We note that assuming $d \in [0, 1]^m$ is equivalent to assuming that $\mathcal{P}$ is nonempty. Moreover, we note that we cannot assume a support of $[-1, 1]$ and a mean of 0 without loss of generality, as this implies that $\zeta$ is symmetric. It is straightforward, however, to extend Theorem 3 and 4 to an ambiguity set with support $[-1, u]$ for some $u > 0$, which can be assumed without loss of generality. For the sake of brevity, we only state this generalized version of Theorem 3.

**Theorem 5.** Let $g : \mathbb{R}^n \mapsto \mathbb{R}$, $b \in \mathbb{R}$ and let $\zeta$ be an 1-dimensional random variable whose distribution lies in the ambiguity set

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

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

$$\inf_{P \in \mathcal{P}} \mathbb{P}[g(x) \leq b - \zeta] \geq 1 - \epsilon,$$

if and only if

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

Chance constraints with independent right-hand sides appear in, for example, inventory management, where the future demand is uncertain. In particular, traditional inventory control policies with a prescribed minimal service level fit this framework. We illustrate this application in Section 5.1.
A similar argument to the one made in Theorem 4 can be made to reformulate an ambiguous joint chance constraint with the same uncertain parameter on all their right-hand sides. For the sake of brevity, we do not include the proof of this theorem here, but it can be found in Appendix A.

**Theorem 6.** Let \( g_i : \mathbb{R}^n \to \mathbb{R} \) be convex and \( b_i \in \mathbb{R} \) for \( i = 1, \ldots, m \) and let \( \zeta \) be a 1-dimensional random variable whose distribution lies in the ambiguity set

\[
P = \{ \mathbb{P} : \mathbb{P}[\zeta \in [-1, 1]] = 1, \ E[\zeta] = 0, \ E[|\zeta|] = d \},
\]

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

\[
\inf_{\mathbb{P} \in P} \mathbb{P}\left[ g_i(x) \leq b_i - \zeta \ \forall i \right] \geq 1 - \epsilon,
\]

if and only if

\[
\min_i \{ b_i - g_i(x) \} \geq \min \left\{ 1, \frac{d}{2\epsilon} \right\}.
\]

We note that the minimum of concave functions is concave, and thus (16) is a convex constraint. Joint chance constraints with an identical uncertain parameter on their right-hand side find applications in, e.g., floodwater protection Postek et al. (2019), where it models the sea-level rise, and inventory management problems with multiple constraints that depend on the demand of a single product.

Providing a tractable reformulation when uncertainty is present in the left-hand side is significantly more complicated. Using our results, we provide a reformulation when all constraints are linear and the left-hand side coefficients affinely depend on \( \zeta \). As the proof of this theorem is highly similar to the earlier proofs in this section, we have included it only in Appendix A.

**Theorem 7.** Let \( \bar{a}, \hat{a} \in \mathbb{R}^n, b \in \mathbb{R} \) and \( \zeta \in \mathbb{R} \) be a random variable whose distribution lies in the ambiguity set

\[
P = \{ \mathbb{P} : \mathbb{P}[\zeta \in [-1, 1]] = 1, \ E[\zeta] = 0, \ E[|\zeta|] = d \},
\]

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

\[
\inf_{\mathbb{P} \in P} \mathbb{P}\left[ (\bar{a} + \zeta \hat{a})^\top x \leq b \right] \geq 1 - \epsilon,
\]

if and only if

\[
\bar{a}^\top x + \min \left\{ 1, \frac{d}{2\epsilon} \right\} \cdot |\hat{a}^\top x| \leq b.
\]

Interesting to note is that (18) has a linear representation. An example of constraints with left-hand side uncertainty dependent on a single uncertain parameter can be found
in radiation therapy (Ten Eikelder, 2018). We discuss this example in more detail in Section 5.2.

We now compare the result of Theorem 7 to one of the safe approximations derived by Postek et al. (2018):

\[
\begin{align*}
\tilde{a}^\top x - b &= u_0 + v_0 \quad \text{(19a)} \\
\tilde{a}^\top x &= u_1 + v_1 \quad \text{(19b)} \\
u_0 + |u_1| &\leq 0 \quad \text{(19c)} \\
v_0 + \sqrt{2 \log \left(\frac{1}{\epsilon}\right) \cdot \sigma \cdot |v_1|} &\leq 0, \quad \text{(19d)}
\end{align*}
\]

where \(u_0, u_1, v_0, v_1 \in \mathbb{R}\) are auxiliary variables and \(\sigma\) is given by

\[
\sigma = \sup_{t \in \mathbb{R}} \sqrt{\frac{2 \log \left(\frac{1}{\epsilon}d \cdot \cosh(t) + 1 - d\right)}{t^2}}.
\]

An important difference between (18) and (19) is the appearance of \(\epsilon\). More specifically, because of the logarithm in (19), its extra safety buffer does not grow as extremely as \(\epsilon\) becomes small. On the other hand, below a threshold, varying \(\epsilon\) does not influence the feasible region of (18), or the feasible region of (19) because it is optimal to choose \(v_1 = 0\).

Figure 4 compares the feasible region of a simple two-dimensional example:

\[
\inf_{P \in \mathcal{P}} \mathbb{P} \left[ \left( \begin{bmatrix} 4 \\ 2 \end{bmatrix} + \zeta \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)^\top x \leq 5 \right] \geq 1 - \epsilon \quad \text{where } d = \frac{1}{4}
\]

and we consider \(\epsilon = 0.2\) and \(\epsilon = 0.4\). In this numerical illustration, the feasible regions of (18) and (19) are identical for any \(\epsilon \leq 0.125\), which are exactly the values such that \(\min \{1, \frac{d}{2\epsilon} \} = 1\). We also remark that the formulation provided in (18) is an exact reformulation, while (19) is not. The latter is, however, applicable to a much wider selection of chance constraints, which we discuss in Section 4.2.

## 4 Sums of Random Variables

### 4.1 An Upper Bound for the Tail Probability

In this section we consider \(N\) random variables \(X_1, \ldots, X_N\) with known support, mean and mean absolute deviation and the tail probability of their sum, that is,

\[
\sup_{P \in \mathcal{F}} \mathbb{P} \left[ \sum_{i=1}^{N} X_i \geq \tau \right]. \quad (20)
\]
A Chebyshev inequality based on mean absolute deviation

Roos et al.

Figure 4: The feasible regions of our exact reformulation (18) in blue and the safe approximation (19) of Postek et al. (2018) in orange for \( \bar{a} = [4, 2]^\top, \hat{a} = [1, 1]^\top, b = 5, d = \frac{1}{4}, \) and two different values of \( \epsilon. \)

Here, \( \mathcal{F} \) is the multi-dimensional ambiguity set, i.e.,

\[
\mathcal{F} = \{ \mathbb{P} \mid \mathbb{P}[X_i \in [a_i, b_i]] = 1, \ \mathbb{E}_\mathbb{P}[X_i] = \mu_i, \ \mathbb{E}_\mathbb{P}[|X_i - \mu_i|] = d_i, \ i = 1, \ldots, N \}. \tag{21}
\]

Note that we do not make any assumptions with regard to (in)dependence or correlation between the random variables. To analyze (20) we define a new random variable \( Y = \sum_{i=1}^N X_i. \) Clearly, the support and mean of \( Y \) follow from (21):

\[
\mathbb{P}\left[ Y \in \left[ \sum_{i=1}^N a_i, \sum_{i=1}^N b_i \right] \right] = 1, \quad \mathbb{E}_\mathbb{P}[Y] = \sum_{i=1}^N \mu_i.
\]

Unfortunately, the mean absolute deviation is unknown and applying Theorem 1 is thus not straightforward. To ease notation, we shall denote the sums of \( a_i, \mu_i, b_i \) and \( d_i, \) by \( \bar{a}, \bar{b}, \bar{\mu} \) and \( \bar{d}, \) respectively. Theorem 8 presents a bound on (20) for any upper bound on the mean absolute deviation of \( Y. \)

**Theorem 8.** For any \( \hat{d} \) such that

\[
\mathbb{E}\left[ \left| \sum_{i=1}^N X_i - \bar{\mu} \right| \right] \leq \hat{d},
\]

it holds that

\[
\sup_{\mathbb{P} \in \mathcal{F}} \mathbb{P}\left[ \sum_{i=1}^N X_i \geq \tau \right] \leq \begin{cases} 
1 & \text{if } \tau \leq \bar{\mu} \\
\min\left\{ \frac{\hat{d}}{2(\tau - \bar{\mu})}, \frac{\bar{\mu} - \bar{a}}{\tau - \bar{b}} \right\} & \text{if } \tau \in (\bar{\mu}, \bar{b}].
\end{cases}
\]

**Proof.** We consider the following ambiguity set for the distribution of \( Y: \)

\[
\mathcal{G} = \left\{ \mathbb{P} \mid \mathbb{P}[Y \in [\bar{a}, \bar{b}]] = 1, \ \mathbb{E}_\mathbb{P}[Y] = \bar{\mu}, \ \mathbb{E}[|Y - \bar{\mu}|] \leq \hat{d} \right\}. \tag{23}
\]
It should be noted that (23) is not an ambiguity set of the form (1), because of the inequality for the mean absolute deviation. This also means that the tightest upper bound this approach can obtain for \( \tau \in [\bar{\mu}, \bar{a}] \) is 1, as the distribution with probability mass 1 on \( \bar{\mu} \) is an element of \( G \). For \( \tau \in (\bar{\mu}, \bar{b}] \), we can however infer from Theorem 1 that

\[
\sup_{P \in \mathcal{F}} \mathbb{P} \left[ \sum_{i=1}^{N} X_i \geq \tau \right] \leq \sup_{P \in G} \mathbb{P} \left[ Y \geq \tau \right]
\]

= \begin{cases} 
1 & \text{if } \tau \leq \bar{\mu} \\
\max_{d \in [0, \hat{d}]} \left\{ \min \left\{ \frac{d}{2(\tau - \bar{\mu})}, 1 - \frac{d}{2(\bar{\mu} - \bar{a})} \right\} \right\} & \text{if } \tau \in (\bar{\mu}, \bar{b}] 
\end{cases}

We will now simplify this expression by solving the maximization problem over \( d \) explicitly. To that end, we first note that the minimum is taken over two linear functions of \( d \), an increasing and a decreasing one. Therefore, the global maximum is at the intersection of these functions, and the optimal \( d \) is thus either \( \hat{d} \) or \( d^* \), where \( d^* \) is such that

\[
\frac{d^*}{2(\tau - \bar{\mu})} = 1 - \frac{d^*}{2(\bar{\mu} - \bar{a})};
\]

Solving this equation yields

\[
d^* = \frac{2(\tau - \bar{\mu})(\bar{\mu} - \bar{a})}{\tau - \bar{a}}.
\]

We remark that \( \hat{d} \) is the optimal solution when

\[
\frac{\hat{d}}{2(\tau - \bar{\mu})} < 1 - \frac{\hat{d}}{2(\bar{\mu} - \bar{a})},
\]

as this means that \( \hat{d} < d^* \). Therefore, we find that

\[
\max_{d \in [0, \hat{d}]} \left\{ \min \left\{ \frac{d}{2(\tau - \bar{\mu})}, 1 - \frac{d}{2(\bar{\mu} - \bar{a})} \right\} \right\} = \min \left\{ \frac{\hat{d}}{2(\tau - \bar{\mu})}, \frac{d^*}{2(\tau - \bar{\mu})} \right\} = \min \left\{ \frac{\hat{d}}{2(\tau - \bar{\mu})}, \frac{\bar{\mu} - \bar{a}}{\tau - \bar{a}} \right\}.
\]

The most obvious candidate for \( \hat{d} \) is given by the sum of all mean absolute deviations \( \bar{d} \), which is clearly an upper bound as

\[
\mathbb{E} \left[ \sum_{i=1}^{N} X_i - \bar{\mu} \right] = \mathbb{E} \left[ \sum_{i=1}^{N} X_i - \mu_i \right] \leq \sum_{i=1}^{N} \mathbb{E} \left[ |X_i - \mu_i| \right] = \bar{d}.
\]

This bound, however, is generally not tight. It is tight, for example, when \( \mu_i = 0 \) for \( i = 1, \ldots, N \) and \( d_i = d_j \) for all \( i \neq j \). Several other possible bounds that can be used are described by Postek et al. (2018).

Figure 5 shows the progression of (22) for a simple numerical example where \( N = 2 \), \( a_1 = a_2 = -1, \mu_1 = \mu_2 = 0, b_1 = b_2 = 1 \) and \( d_1 = d_2 = \frac{1}{2} \).
Figure 5: An example of the multi-dimensional bound given by (22) with the parameters \( N = 2, a_1 = a_2 = -1, \mu_1 = \mu_2 = 0, b_1 = b_2 = 1 \) and \( d_1 = d_2 = \frac{1}{2} \).

### 4.2 Chance Constraints

In this subsection we will apply the upper bound obtained in Theorem 8 for the sum of random variables to find a safe approximation to linear chance constraints with left-hand side uncertainty. As the proof is a direct application of Theorem 8 similar to those in Section 3, we have omitted it here. We refer the interested reader to Appendix A.

**Theorem 9.** Let \( \bar{a} \in \mathbb{R}^n, A \in \mathbb{R}^{n \times m}, b \in \mathbb{R} \) and let \( \zeta \) be an \( m \)-dimensional random variable whose distribution lies in the ambiguity set

\[
\{ P \mid P[\zeta_i \in [-1, 1]] = 1, \ E_P[\zeta_i] = 0, \ E_P[|\zeta_i|] = d_i, \ i = 1, \ldots, m \}. 
\]

For any \( \epsilon \in (0, \frac{1}{2}) \) and \( x \in \mathbb{R}^n \) it holds that

\[
\bar{a}^\top x + \sum_{i=1}^m \min \left\{ 1, \frac{d_i}{2\epsilon} \right\} \left| a_i^\top x \right| \leq b, \tag{24}
\]

implies

\[
\inf_{P \in \mathcal{F}} P \left[ (\bar{a} + A\zeta)^\top x \leq b \right] \geq 1 - \epsilon. \tag{25}
\]

The safe approximation derived in Theorem 9 can also be interpreted as a tractable robust counterpart, that is, it is equivalent to

\[
(\bar{a} + A\zeta)^\top x \leq b \quad \forall \zeta \in U,
\]

where \( U \) is the uncertainty set defined as

\[
U = \{ \zeta \in \mathbb{R}^m \mid \|\zeta\|_\infty \leq 1 \} \cap \left\{ \zeta \in \mathbb{R}^m \mid |\zeta_i| \leq \frac{d_i}{2\epsilon} \quad i = 1, \ldots, m \right\}.
\]
The first part of this set is simply the support of $\zeta$, and imposing that uncertainty set thus leads to a safe approximation for any $d$ and $\epsilon$. The second part, on the other hand, leads to some potential improvement, as it potentially shrinks the uncertainty set.

Based on this robust counterpart interpretation, the safe approximation we derive seems rather conservative, especially when compared to the one provided in Postek et al. (2018), where the uncertainty set is the former box intersected with an ellipsoid. The key difference, however, is that in Postek et al. (2018) the ambiguity set for $\zeta$ also assumes pairwise independence. This removes a significant number of pathological distributions from the ambiguity set, such as those where all $\zeta_i$ are perfectly correlated, i.e., in any scenario where $\zeta_1$ takes its worst possible value, so do $\zeta_2, \ldots, \zeta_m$. Because of the inclusion of those distributions in the ambiguity set, it includes distributions that have an unreasonably high probability at the vertices of the support. When pairwise independence is assumed, however, no such distributions are considered, and thus a much smaller uncertainty set $U$ can be used to find a safe approximation. In conclusion, chance constraints under ambiguity without any assumption on independence are extremely conservative and will often lead to safe approximations that simply account for the whole support.

5 Illustrations

5.1 Inventory Management

One of the basic optimization problems in inventory management is determining the control parameters of a control policy. Important in this optimization is the trade off between the cost of holding inventory and the costs of not being able to serve all customers due to a stockout. However, because stockout costs are usually difficult to estimate, companies often use a customer service level instead. A well-known customer service level is the cycle service level, which is the non-stockout probability in a replenishment cycle (Silver et al., 1998). For a standard base-stock control policy with base-stock level $s$, the cycle service level is $P[D \leq s]$, where $D$ is the lead-time demand. Hence, the lowest possible base-stock level that guarantees a cycle service of $1 - \epsilon$ can be obtained by solving:

$$\min_s s$$

s.t. $P[D > s] \leq \epsilon$,

with $P$ is the lead-time demand’s distribution.

Common distributions used in inventory management are the normal and gamma distributions, whose parameters would be estimated from available data (Silver et al., 1998). In this section we will analyze the worst- and best-case performance of the base-stock level resulting from these assumptions. As an illustration, we will assume the
demand can vary between $a = 0$ and $b = 100$, and has mean $\mu = 50$. Moreover, we will vary the maximum allowed probability of a stockout between 1% and 20%, and the mean absolute deviation between 5 and 20.

The theoretical optimal base-stock levels for the normal and gamma distributions with these characteristics are plotted in Figure 6. For $d = 10$, for example, the theoretically optimal base-stock levels for the normal and gamma distributions are given by $s \approx 70.62$ and $s \approx 72.37$, respectively. The worst-case probability that the demand exceeds these base-stock levels can easily be found by applying Theorem 1. Remarkably, because of the construction of the base-stock level for the normal distribution and the relation between the standard deviation and the mean absolute deviation for the normal distribution ($\sigma = \frac{d\sqrt{\pi/2}}{2}$), the worst-case probability of a stockout is the same for all values of $d$:

$$\sup_{p \in \mathcal{P}} P(D > s) = \frac{d}{2(s - \mu)} = \frac{d}{2(\mu + \sigma\Phi^{-1}(1 - \epsilon) - \mu)}$$

$$= \frac{d}{2 \frac{d}{\sqrt{\pi}} \Phi^{-1}(1 - \epsilon)} = \frac{\sqrt{\frac{2}{\pi}}}{2\Phi^{-1}(1 - \epsilon)} \approx 0.2425,$$

with $\Phi$ the cumulative distribution function of the standard normal distribution. For the gamma distribution, the worst-case probability does vary: it decreases slightly as $d$ increases, from approximately 0.2326 at $d = 5$ to 0.2075 at $d = 20$, indicating that assuming a gamma distribution results in better worst-case characteristics at high variability.

Figure 7 plots the worst-case probability of a stockout for different values of $\epsilon$. Here, the performance of the normal and gamma distribution is pretty similar, although the
gamma distribution performs slightly better for low values of $\epsilon$.

The results of Theorem 1 can of course also be used to find a base-stock level such that the worst-case probability of the demand exceeding it is limited. In this numerical illustration, however, for any base-stock level strictly smaller than $b = 100$, the worst-case probability of the demand exceeding it is at least 10%. Without an upper bound on the support, the base-stock level needed to guarantee a worst-case probability of stockout of 5% is in fact 150, over twice as high as those found by assuming a normal or gamma distribution. This shows that simply replacing the probability in the original problem by the worst-case probability can lead to overly conservative solutions. We therefore recommend using the worst-case probability as additional information regarding the possible effect of assuming an incorrect distribution, but not making it the sole criterion.

5.2 Radiation Therapy

In radiation therapy, the primary goal is to deliver ionizing radiation to a target volume while keeping the dose to all organs-at-risk (OARs) to a minimum. Traditionally, a single treatment plan would be devised to achieve these goals. In this section we will illustrate the theory we develop through an example in analyzing a treatment plan.

Mathematically, we consider a number of treatments, also known as treatment fractions, indexed by $t = 1, \ldots, N$, with an associated dose of radiation to the tumor denoted by $x_t$. Because cells can recover from radiation between treatment fractions, we consider
the biological effective dose (BED) model (Fowler, 2010), which defines the BED based on some dose \( x \) distributed over \( N \) fractions as

\[
BED(x, N) = \sum_{t=1}^{N} \left( x_t + \frac{x_t^2}{\alpha/\beta} \right),
\]

where \( \alpha \) and \( \beta \) are radiosensitivity parameters of the irradiated tissue. More specifically, \( \alpha \) captures the linear effect of dose, i.e., the radiosensitivity of the tissue, while \( \beta \) captures the quadratic effect, that is, it accounts for its repair capability. The ratio \( \alpha/\beta \) can then be interpreted as the tissue’s sensitivity to fractionation, where a low ratio indicates a high sensitivity to fractionation, i.e., the distribution of treatment over multiple fractions.

From here on out, we will denote the inverse of the \( \alpha/\beta \) ratio by \( \rho \) such that we have

\[
BED(x, N) = \sum_{t=1}^{N} \left( x_t + \rho x_t^2 \right).
\]

While there is an extensive body of research on the value of \( \rho \) for different tumor sites, it remains subject to significant uncertainty (Joiner and 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 \( \rho \) follows some well known distribution. We will therefore just assume a reasonable estimate of its support, mean and mean absolute deviation are available from a combination of limited data and expert knowledge, which we will again denote by \([a, b] = [\mu, \mu + d] \).

A frequently used objective in treatment planning is then to maximize the tumor BED. Under the uncertainty of \( \rho \), we are interested in guarantees for the tumor BED, i.e., the percentiles of its worst-case distribution. Mathematically, this means we are interested in solving

\[
\max_{\tau} \quad \tau \\
\text{s.t.} \quad \mathbb{P} \left[ \tau > \sum_{t=1}^{N} \left( x_t + \rho x_t^2 \right) \right] \leq \epsilon \quad \forall \mathbb{P} \in \mathcal{P},
\]

for some prescribed safety level \( \epsilon \), where \((x_1, \ldots, x_N)\) is the provided treatment plan.

For a numerical illustration, we use the setup of the numerical experiments performed by Ten Eikelder (2018). More specifically, we use the solutions to a static robust optimization problem maximizing the tumor BED subject to some constraints, where \( \rho \) was considered uncertain and in the interval \([0.1875, 0.3125]\), with a nominal value of 0.25. It thus makes sense to use an ambiguity set with support \([a, b] = [0.1875, 0.3125]\) and mean \( \mu = 0.25 \). Figure 8 shows the value of \( \tau \) for different values of the mean absolute deviation \( d \), averaged over 50 instances, for \( \epsilon = 0.1 \). Here, note that 125.95 Gy is the mean worst-case tumor BED, i.e., the one for \( \rho = a = 0.1875 \).
Figure 8: Highest 75% guaranteed tumor BED for varying mean absolute deviation, for the parameters \([a, b] = [0.1875, 0.3125]\) and \(\mu = 0.25\).

Figure 9 shows the value of \(\tau\) for different values of \(\epsilon\) while the mean absolute deviation is set to \(d = 0.0125\). It shows that with 90% confidence, we can guarantee nothing better than the absolute worst tumor BED, while at 75% confidence we can guarantee a tumor BED of approximately 138.11 Gy.

6 Conclusions and Recommendations

In this paper we present a tight upper and lower bound on the tail probability of a random variable given support, mean and mean absolute deviation. Additionally, we also derive an upper bound on the tail probability of the sum of random variables given this distributional information. The main result is related to the famous Chebyshev inequality, from which it differs through the use of support information as well as a different dispersion measure.

These fundamental results allow us to find exact reformulations of single chance constraints satisfying specific structural properties. Using the support information available, we are able to efficiently identify and incorporate situations where there is zero probability of violating the underlying constraint. This explains the difference between our exact reformulation and the level of conservatism of the currently available safe approximations to such constraints.

Moreover, we present safe approximations to joint ambiguous chance constraints and general single linear chance constraints with multiple uncertain parameters. We discuss, using a robust counterpart interpretation, how ambiguous chance constraints without independence assumption can be overly conservative and are often equivalent to simply
Figure 9: Highest $(100 \cdot \epsilon) \%$ guaranteed tumor BED for varying $\epsilon$, for the parameters $[a, b] = [0.1875, 0.3125]$, $\mu = 0.25$ and $d = 0.0125$.

requiring the constraint to be valid for any realization in the support.

Acknowledgments

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].

References


A Proofs

Proof of Theorem 6. Using the fact that every constraint features the same uncertain parameter $\zeta$ we find
\[
\inf_{P \in \mathcal{P}} \mathbb{P} \left[ g_i(x) \leq b_i - \zeta_i \ \forall i \right] = \inf_{P \in \mathcal{P}} \mathbb{P} \left[ \zeta \leq b_i - g_i(x) \ \forall i \right]
= \inf_{P \in \mathcal{P}} \mathbb{P} \left[ \zeta \leq \min_i \{b_i - g_i(x)\} \right]
= 1 - \sup_{P \in \mathcal{P}} \mathbb{P} \left[ \zeta > \min_i \{b_i - g_i(x)\} \right].
\]
We thus know that (15) is equivalent to
\[
\sup_{P \in \mathcal{P}} \mathbb{P} \left[ \zeta > \min_i \{b_i - g_i(x)\} \right] \leq \epsilon,
\]
to which we can apply Theorem 1 with $\tau = \min_i \{b_i - g_i(x)\}$. Since we know $\epsilon < \frac{1}{2}$, we once again find that it must hold that $\min_i \{b_i - g_i(x)\} > 0$ and thus by the same reasoning as in the proof of Theorem 4 we find that it must hold that
\[
\sup_{P \in \mathcal{P}} \mathbb{P} \left[ \zeta > \min_i \{b_i - g_i(x)\} \right] = \min \left\{ \frac{d}{\min_i \{b_i - g_i(x)\}}, \ 1 - \frac{d}{2} \right\} \text{ if } \min_i \{b_i - g_i(x)\} < 1
0 \quad \text{ if } \min_i \{b_i - g_i(x)\} \geq 1.
\]
Once again, it must hold that $\min \left\{ \frac{d}{\min_i \{b_i - g_i(x)\}}, \ 1 - \frac{d}{2} \right\} = \frac{d}{2 \min_i \{b_i - g_i(x)\}}$ and subsequently both cases can be combined. We thus find that (15) is equivalent to
\[
\min_i \{b_i - g_i(x)\} \geq \min \left\{ 1, \ \frac{d}{2\epsilon} \right\},
\]
which makes the condition $\min_i \{b_i - g_i(x)\} > 0$ redundant.

Proof of Theorem 7. Rewriting (17) we find
\[
\inf_{P \in \mathcal{P}} \mathbb{P} \left[ (\bar{a} + \zeta \hat{a})^\top x \leq b \right] = 1 - \sup_{P \in \mathcal{P}} \mathbb{P} \left[ (\bar{a} + \zeta \hat{a})^\top x \leq b \right]
= 1 - \sup_{P \in \mathcal{P}} \mathbb{P} \left[ \zeta \cdot \hat{a}^\top x > b - \bar{a}^\top x \right].
\]
Thus, clearly, (17) is equivalent to
\[
\sup_{P \in \mathcal{P}} \mathbb{P} \left[ \zeta \cdot \hat{a}^\top x > b - \bar{a}^\top x \right] \leq \epsilon,
\]
and since $\epsilon \in \left( 0, \frac{1}{2} \right)$, it must hold that $b - \bar{a}^\top x > 0$, because for any $b - \bar{a}^\top x \leq 0$, the supremum is at least $\frac{1}{2}$ by Theorem 1. Given that $b - \bar{a}^\top x > 0$ we thus find
\[
\sup_{P \in \mathcal{P}} \mathbb{P} \left[ \zeta \cdot \hat{a}^\top x > b - \bar{a}^\top x \right] = \begin{cases} 
\min \left\{ \frac{d|\hat{a}^\top x|}{2(b - \bar{a}^\top x)}, \ 1 - \frac{d}{2} \right\} & \text{if } b - \bar{a}^\top x < |\hat{a}^\top x| \\
0 & \text{if } b - \bar{a}^\top x \geq |\hat{a}^\top x|.
\end{cases}
\]
Since $1 - \frac{d}{2} \geq \frac{1}{2}$, it must hold that $\min \left\{ \frac{d \cdot |a^\top x|}{2(b - a^\top x)}, \, 1 - \frac{d}{2} \right\} = \frac{d \cdot |a^\top x|}{2(b - a^\top x)}$ for any $x$ that satisfies (27). Rewriting the case where $b - a^\top x < |a^\top x|$ we find

$$\frac{d \cdot |a^\top x|}{2(b - a^\top x)} \leq \epsilon \iff a^\top x + \frac{d}{2\epsilon} \cdot |a^\top x| \leq b.$$ 

We can thus combine both cases, such that (27) is equivalent to

$$a^\top x + \min \left\{ 1, \, \frac{d}{2\epsilon} \right\} \cdot |a^\top x| \leq b,$$

where it should be noted that this implies $b - a^\top x > 0$, which is therefore redundant.

**Proof of Theorem 9.** We first rewrite (25) as

$$\inf_{P \in F} P \left[ (a + A\zeta)^\top x \leq b \right] \geq 1 - \epsilon \iff \sup_{P \in F} P \left[ (a + A\zeta)^\top x > b \right] \leq \epsilon$$

$$\iff \sup_{P \in F} P \left[ x^\top A\zeta > b - a^\top x \right] \leq \epsilon. \tag{28}$$

Theorem 8 gives an upper bound for the left-hand side of this constraint, as $x^\top A\zeta$ is the sum of random variables with support $[-|x^\top a_i|, |x^\top a_i|]$, mean 0 and mean absolute deviation $d_i|x^\top a_i|$ for $i = 1, \ldots, m$, where $a_i$ is the $i$-th column of $A$. Plugging in the result of Theorem 8 we find

$$\sup_{P \in F} P \left[ x^\top A\zeta > b - a^\top x \right]$$

$$\leq \begin{cases} 1 & \text{if } b - a^\top x \leq 0 \\ \min \left\{ \frac{|A^\top x|^d}{2(b - a^\top x)}, \, \frac{\|A^\top x\|_1}{b - a^\top x + \|A^\top x\|_1} \right\} & \text{if } 0 < b - a^\top x < \|A^\top x\|_1 \\ 0 & \text{if } b - a^\top x \geq \|A^\top x\|_1. \end{cases}$$

From this, we find the following safe approximation to (28):

$$\begin{cases} b - a^\top x > 0 \\ \min \left\{ \frac{|A^\top x|^d}{2(b - a^\top x)}, \, \frac{\|A^\top x\|_1}{b - a^\top x + \|A^\top x\|_1} \right\} \leq \epsilon. \tag{29} \end{cases}$$

The second constraint means either of the fractions should be smaller than $\epsilon$. Some rewriting yields that

$$\frac{|A^\top x|^d}{2(b - a^\top x)} \leq \epsilon \iff a^\top x + \frac{1}{2\epsilon} |A^\top x|^d \leq b$$

$$\iff a^\top x + \sum_{i=1}^m \frac{d_i}{2\epsilon} |a_i^\top x| \leq b, \tag{30}$$

and

$$\frac{\|A^\top x\|_1}{b - a^\top x + \|A^\top x\|_1} \leq \epsilon \iff a^\top x + \left( \frac{1}{\epsilon} - 1 \right) \|A^\top x\|_1 \leq b$$

$$\iff a^\top x + \left( \frac{1}{\epsilon} - 1 \right) \sum_{i=1}^m |a_i^\top x| \leq b. \tag{31}$$
The constraint \( b - \bar{a}^\top x > 0 \) is implied by both, and can thus be disregarded. From the upper bound we also found that the probability of violation is zero if \( b - \bar{a}^\top x \geq \|A^\top x\|_1 \), i.e., this is also a sufficient condition for (25) to hold. This can be included in (31) by altering it as

\[
\bar{a}^\top x + \min \left\{ 1, \frac{1}{\epsilon} - 1 \right\} \cdot \sum_{i=1}^{m} |a_i^\top x| \leq b.
\]

Because \( \min \left\{ 1, \frac{1}{\epsilon} - 1 \right\} = 1 \) for any \( \epsilon \leq \frac{1}{2} \), we find that this is equivalent to

\[
\bar{a}^\top x + \sum_{i=1}^{m} |a_i^\top x| \leq b,
\]

which can be combined with (30) to

\[
\bar{a}^\top x + \sum_{i=1}^{m} \min \left\{ 1, \frac{d_i}{2\epsilon} \right\} |a_i^\top x| \leq b,
\]

which completes the proof.