Adjusted Distributionally Robust Bounds on Expected Loss Functions

Yasemin Merzifonluoğlu[1] and Joseph Geunes[2]

[1]Tilburg School of Economics and Management, Econometrics and Operations Research
PO Box 90153 5000 LE Tilburg, The Netherlands
[2]Texas A&M University, Department of Industrial & Systems Engineering, College Station, TX, USA

May 27, 2022

Abstract

Optimization problems in operations and finance often include a cost that is proportional to the expected amount by which a random variable exceeds some fixed quantity, known as the expected loss function. Representation of this function often leads to computational challenges, depending on the distribution of the random variable of interest. Moreover, in practice, a decision maker may possess limited information about this probability distribution, such as the mean and variance, but not the exact form of the associated probability density or distribution function. In such cases, a distributionally robust (DR) optimization approach seeks to minimize the maximum expected cost among all possible distributions that are consistent with the available information. Past research has recognized the overly conservative nature of this approach because it accounts for worst-case probability distributions that almost surely do not arise in practice. Motivated by this, we propose a DR approach that accounts for the worst-case performance with respect to a broad class of common continuous probability distributions, while producing solutions that are less conservative (and, therefore, less expensive, on average) than those produced by existing DR approaches in the literature. The methods we propose also permit approximation of the expected loss function for probability distributions under which exact representation of the function is difficult or impossible. Finally, we draw a connection between Scarf-type bounds from the literature, and mean-MAD (mean absolute deviation) bounds when MAD information is available in addition to variance.

Keywords: Distributionally Robust Optimization; Mean variance ambiguity; Mixture ambiguity; Newsvendor problem; Scarf's model.

1 Introduction

For a random variable X, the expected shortage, or loss function, is defined as the expected amount by which the random variable exceeds some predetermined fixed level (say, Q); that is, the expected loss function can be expressed mathematically as $\ell(Q) = E\left[(X-Q)^+\right]$, where $E[\cdot]$ denotes the expectation operator, and $(x)^+ = \max\{x, 0\}$. Expected loss functions are used quite often in finance, inventory theory, and other fields in operations management, where a decision maker would like to quantify the expected amount by which a financial loss exceeds some value, or demand for a resource exceeds available capacity. When an expected loss function arises within an optimization problem, computational and modeling obstacles may arise. The first difficulty occurs when, although the decision maker is able to characterize the true distribution of the random variable(s) under consideration, an explicit functional form for the expected loss function may not exist for the given distribution. Furthermore, in cases involving, for example, the convolution of random variables, the resulting distribution is sometimes extremely difficult to characterize or to work with analytically. Such cases require an ability to approximate the expected cost of the associated optimization problem as closely as possible. The second difficulty occurs when the decision maker is not able to characterize the 'true' distribution from among all conceivable distributions. In practice, the decision maker may only be able to estimate some basic parameters of the distribution, e.g., a few low-order moments.

Much of the operations literature has approached these computational and modeling obstacles using a normal distribution approximation. The rationale behind this distribution choice can be explained as follows. Firstly, when aggregate demand is comprised of individual demands, each of which follows a normal distribution, then this aggregate demand also follows a normal distribution. Secondly, under a normal distribution, an analytical expression exists for the expected overflow that is computationally straightforward to evaluate. Lastly, the sum of a number of independent and identically distributed demands tends toward a normal distribution as the number of such demands increases due to the Central Limit Theorem, regardless of the underlying distribution.

On the other hand, a normal approximation may not provide the desired results in some situations where the underlying distributions are not closed under addition. (A distribution is closed under addition if, for any two independent random variables X and Y with distributions belonging to a distribution family $\mathcal{F}$, the distribution of $X + Y$ also belongs to $\mathcal{F}$; examples of such distributions include the normal, Poisson, chi-squared, and gamma distribution with a common rate parameter.) Moreover, demand quantities in operations are typically positive random variables, while the normal distribution allows negative values. Applying a normal distribution assumption may thus be prohibitive when the coefficient of variation is high and/or the distribution is not symmetric. The use of expected loss functions is also crucial for service operations management models, where a normal distribution generally does not provide the best fit. For instance, several studies have concluded that the lognormal distribution fits actual surgery-duration data better than a normal distribution (see references [4], [13], [20], [21], [22], [23]) due to its positive skew. Furthermore, in a service setting, where classical queuing models are often adopted (such as in stochastic machine scheduling), inter-event times are often modeled using an exponential distribution.

We take a distributionally robust (DR) approach, assuming the decision maker is only able to estimate the first two moments of the distribution, i.e., the mean and variance. This leads to an *ambiguity set* of possible distributions, or what is effectively an infinite set of distributions that may be consistent with the available information. Under such circumstances, a solution approach designed for a particular distribution may lead to very poor performance under the true demand distribution. A DR approach seeks a solution that achieves the best possible performance in view of the worst-case distribution within the ambiguity set (Hanasusanto [8]). The DR approach we take also facilitates approximation of the expected loss function in cases where the probability distribution is known, but its exact representation leads to intractable optimization problems.

Expected loss functions play an important role in newsvendor type problems within operations. Since the seminal work of Scarf [19], researchers have shown interest in a DR version of the newsvendor problem. Given only mean and variance information, the distribution that minimizes the worst-case expected performance is a discrete distribution with two point masses that depends on the order quantity (Scarf [19]). The associated unnatural discrete distribution is generally inconsistent with the available data in most practical settings, and the dependence of the demand distribution on the order quantity implies that the likelihood of such a demand distribution arising in practice is practically zero. Thus, the resulting model is likely to provide overly conservative solutions. As a result, Scarf's model has not been widely adopted by practitioners, as it is considered to be overly pessimistic by many researchers (de Klerk et al. [6]).

In this paper we aim to mitigate the overly conservative nature of well-known DR bounds for expected loss functions in the literature by proposing effective alternative approaches. The first of these approaches limits the ambiguity set to certain classes of mixture distributions, and uses a point-wise maximum of the expected loss function under each component of the mixture distribution as an upper bound on the expected loss function. The second approach, motivated by the relationship between Scarf's bound and an underlying class of implied probability distributions, uses a slight generalization of Scarf's bound to reduce the gap between Scarf's original bound and the bound provided by the mixture of distributions. We also provide lower bounds on the loss function when the variable is defined on a finite range of values. Lastly, we draw a connection between Scarf-type bounds and bounds obtained when only mean and mean absolute deviation (MAD) information are available. The primary contributions of this paper are hence threefold: (I) suggesting tighter and analytically and computationally tractable DR loss functions for varying classes of ambiguity sets; (II) introducing a "standardized loss function" concept for various probability distributions; (III) suggesting new and novel MAD-based bounds for symmetric distributions, and comparing these with existing bounds in the literature and (IV) providing computational performance analyses and evidence of the effectiveness of our proposed solution methods.

Section 2 provides DR bounds for expected loss functions based on previous literature. We also illustrate an interesting relationship between the bound provided by Scarf [19] and the well-known Student's t-distribution. Section 3 introduces *standardized loss functions*, derived from past development of this idea for the analysis of normally distributed random variables. We consider the standardized loss functions implied by past work on DR bounds for the expected loss function. Section 4 proposes using a generalized t-distribution to provide tighter upper bounds that apply to a fairly broad and general class of probability distributions. Section 5 considers a class of continuous random variables that follow a mixture density, and provides bounds on the associated loss functions. In Section 6, we focus our attention on symmetric distributions and compare our adjusted bounds with existing MAD-based bounds in the literature. Section 7 discusses the application of our proposed approaches to the DR newsvendor problem, while Section 8 provides a summary of a set of computational tests intended to characterize the performance of these approaches. Section 9 provides concluding remarks and discusses potential directions for future related work.

2 DR Loss Functions

We consider a random variable X with mean μ and variance v. When this is the only information available about the associated random variable (and its associated probability distribution), we will refer to this as *mean-variance ambiguity.* The expected loss function for this random variable assuming an inventory or capacity level Q is defined as $\ell(Q) = E\left[(X-Q)^+\right]$, which determines the expected amount by which the random variable X exceeds Q. This expected loss function is a fundamental component of the newsvendor objective function, characterizing the expected number of unsatisfied customer demands, as we discuss later is Section 7. Defining $\mathcal{D}_{[a,b]}(\mu, v)$ as the set of all random variables with finite lower and upper bounds a and b, respectively (with $a < b$), expected value μ, and variance v, Theorem 2.1 in the monograph of Karlin and Studden [11] provides bounds for $E\left[\min\{X, Q\}\right]$ for any $X \in \mathcal{D}_{[a,b]}(\mu, v)$. Kamburowski [10] summarized these lower and upper bounds, together with the underlying implied worst- and best-case probability distributions. Lemmas 2.1 and 2.2, presented in the next two subsections, provide corresponding lower and upper bounds from past literature on the expected loss function, $E[(X-Q)^{+}]$, for $X \in \mathcal{D}_{[a,b]}(\mu, v)$.

In the remainder of this section, we discuss bounds on the expected loss function for the most general case, where the range of the distribution $[a, b]$ is known and finite; one may set $a = 0$ and $b = \infty$ for nonnegative but unbounded random variables, or $a = -\infty$ and $b = \infty$ for random variables that are unbounded from above and below. We refer to the bounds in this section (Lemma 2.1 and Lemma 2.2) as DR bounds. Note that Scarf [19] first proposed an upper bound on $E[(X-Q)^{+}]$ for unbounded variables. Because of this, we sometimes refer to these bounds as *Scarf-type* or *mean-variance* bounds.

When presenting results, we use the subscripts $[+]$ and $[-]$, respectively, for upper and lower bounds (and the associated worst- and best-case distributions), respectively, when the range of the random variable is a bounded interval (for example, $[+]$ ($[-]$) denotes an upper (lower) bound for a bounded distribution). We replace $[+]$ with $+$ (and $[-]$ with $-$) when providing corresponding bounds for a random variable with a range that is unbounded, but has the given fixed values of mean and variance. Moreover, we use the superscript *DR* to refer to the *Distributionally Robust* bounds suggested in the lemmas that follow. We use $F(x)$ to denote the the cumulative distribution function (CDF) and $f(x)$ to denote the density function (PDF) for a random variable X.

2.1 DR Upper Bounds on the Loss Function

Lemma 2.1 (Upper Bound, [10]) *For a random variable X with mean μ and variance v defined on the interval $[a,b]$, with $-\infty < a < \mu < b < \infty$, $\underset{X\in\mathcal{D}_{[a,b]}(\mu,v)}{Max} E[(X-Q)^{+}] = \ell_{[+]}^{DR}(Q)$, where $\ell_{[+]}^{DR}$ is the convex function defined by*

$$\ell_{[+]}^{DR}(Q) = \begin{cases} -(Q-\mu) + (Q-a)\frac{v}{(\mu-a)^2+v}, & a \le Q \le \frac{1}{2}\left(a+\mu-\frac{v}{a-\mu}\right), \\ \ell_{+}^{DR}(Q) = \frac{1}{2}\left(\sqrt{v+(Q-\mu)^2} - (Q-\mu)\right), & \frac{1}{2}\left(a+\mu-\frac{v}{a-\mu}\right) \le Q \le \frac{1}{2}\left(b+\mu-\frac{v}{b-\mu}\right), \\ \frac{v(b-Q)}{v+(b-\mu)^2}, & \frac{1}{2}\left(b+\mu-\frac{v}{b-\mu}\right) \le Q \le b. \end{cases} \quad (1)$$

This corresponds to a worst-case distribution with the CDF:

$$F_{[+]}^{DR}(x) = \begin{cases} \frac{v}{(\mu-a)^2+v}, & a \le x < \frac{1}{2}\left(a+\mu-\frac{v}{a-\mu}\right), \\ F_{+}^{DR}(x) = \frac{1}{2}\left(1+\frac{x-\mu}{\sqrt{(x-\mu)^2+v}}\right), & \frac{1}{2}\left(a+\mu-\frac{v}{a-\mu}\right) \le x \le \frac{1}{2}\left(b+\mu-\frac{v}{b-\mu}\right), \\ \frac{(b-\mu)^2}{(b-\mu)^2+v}, & \frac{1}{2}\left(b+\mu-\frac{v}{b-\mu}\right) < x < b, \\ 1, & x = b. \end{cases} \tag{2}$$

When the bounds of the random variable are not finite ($a = -\infty$ and $b = \infty$), Scarf's original model [19] characterizes a worst-case, two-point discrete demand distribution that depends on the quantity Q, under which the associated bound on the expected loss function, $\ell_{+}^{DR}(Q)$, is tight. This approach has been criticized because the associated worst-case distribution is discrete with two support points that depend on the value of Q. On the other hand, the worst-case distribution $F_{+}^{DR}(x)$ associated with $\ell_{+}^{DR}(Q)$ is continuous and can be derived using the property $\ell'(Q) = F(Q) - 1$ for any valid continuous distribution with a differentiable CDF, F. The implied random variable, which we denote by X_{+}^{DR}, corresponds to a three-parameter Student's t-distribution with location parameter μ, scale parameter $\sqrt{\frac{v}{2}}$, and two degrees of freedom (see, e.g., [9]). As a result, it corresponds to a valid probability distribution with mean μ and variance ∞. Müller and Stoyan [15] were the first to suggest using this distribution for real-valued random variables with known first and second moments (in their Theorem 1.10.7). This result was also observed by Das et al. [5]. We note that the Student's t-distribution is also a maximum entropy distribution (see [16]). (Section 3.2 will present other such "maximum entropy distributions," where entropy serves as a measure of the degree of uncertainty implied by the distribution.)

When the random distribution has a finite range, $-\infty < a < b < \infty$, the worst-case distribution of Lemma 2.1 is a valid cumulative distribution function with a point mass of weight $\frac{v}{(\mu-a)^2+v}$ at $x = a$ and a point mass of weight $\frac{v}{(b-\mu)^2+v}$ at $x = b$. The random variable associated with $F_{[+]}^{DR}$ can be expressed using the following mixture of discrete and continuous terms:

$$X_{[+]}^{DR} = \begin{cases} a, & a \le X_{+}^{DR} < \frac{1}{2}\left(a+\mu-\frac{v}{a-\mu}\right), \\ X_{+}^{DR}, & \frac{1}{2}\left(a+\mu-\frac{v}{a-\mu}\right) \le X_{+}^{DR} \le \frac{1}{2}\left(b+\mu-\frac{v}{b-\mu}\right), \\ b, & \frac{1}{2}\left(b+\mu-\frac{v}{b-\mu}\right) < X_{+}^{DR} \le b. \end{cases} \tag{3}$$

2.2 DR Lower Bounds on the Loss Function

The following lemma provides the best possible lower bounds on the expected loss function.

Lemma 2.2 (Lower Bound, [10]) *For any random variable X with mean μ and variance v defined on the interval $[a, b]$ with $a < \mu < b$,*

$$\underset{X \in \mathcal{D}_{[a,b]}(\mu,v)}{Min} E[(X-Q)^+] = \ell_{[-]}^{DR}(Q) \text{ where } \ell_{[-]}^{DR} \text{ is defined by}$$

$$\ell_{[-]}^{DR}(Q) = \begin{cases} \mu - Q, & a \le Q \le \left(\mu - \frac{v}{b-\mu}\right), \\ \mu - \frac{(\mu-a)Q+(b-\mu)\mu-v}{b-a}, & \left(\mu - \frac{v}{b-\mu}\right) \le Q \le \left(\mu - \frac{v}{a-\mu}\right), \\ 0, & \left(\mu - \frac{v}{a-\mu}\right) \le Q \le b. \end{cases}$$

The implied two-point best-case distribution can then be written as:

$$X_{[-]}^{DR} = \begin{cases} \mu - \frac{v}{b-\mu}, & w.p. \ \frac{b-\mu}{b-a}, \\ \mu - \frac{v}{a-\mu}, & w.p. \ \frac{\mu-a}{b-a}. \end{cases} \tag{4}$$

Note that the lower bound in Lemma 2.2 provides useful information only when the random variable's range is known and finite. Table 1 summarizes the upper and lower bounds from Lemmas 2.1 and 2.2.

		DR Loss Function	Distribution
	Range for Q	$\ell_{[+]}^{DR}$	$F_{[+]}^{DR}(Q)$
Upper bound	$\left[a, \frac{1}{2}\left(a+\mu-\frac{v}{a-\mu}\right)\right)$	$-(Q-\mu)+(Q-a)\frac{v}{(\mu-a)^2+v}$	$\frac{v}{(\mu-a)^2+v}$
	$\left[\frac{1}{2}\left(a+\mu-\frac{v}{a-\mu}\right), \frac{1}{2}\left(b+\mu-\frac{v}{b-\mu}\right)\right]$	$\ell_{+}^{DR}(Q) = \frac{1}{2}\left(\sqrt{v+(Q-\mu)^2}-(Q-\mu)\right)$	$F_{+}^{DR}(x) = \frac{1}{2}\left(1+\frac{x-\mu}{\sqrt{(x-\mu)^2+v}}\right)$
	$\left(\frac{1}{2}\left(b+\mu-\frac{v}{b-\mu}\right), b\right)$	$\frac{v(b-Q)}{v+(b-\mu)^2}$	$\frac{(b-\mu)^2}{(b-\mu)^2+v}$
	b	0	1
	Range for Q	$\ell_{[-]}^{DR}$	$F_{[-]}^{DR}(Q)$
Lower bound	$\left[a, \left(\mu-\frac{v}{b-\mu}\right)\right)$	$\mu - Q$	0
	$\left[\left(\mu-\frac{v}{b-\mu}\right), \left(\mu-\frac{v}{a-\mu}\right)\right)$	$\mu - \frac{(\mu-a)Q+(b-\mu)\mu-v}{b-a}$	$\frac{b-\mu}{b-a}$
	$\left[\left(\mu-\frac{v}{a-\mu}\right), b\right]$	0	1

Table 1: DR loss functions based on lower and upper bounds from existing literature.

3 Standardized Loss Functions

It is sometimes useful to standardize a random variable X by subtracting its mean and dividing by the standard deviation, i.e., $Z = \frac{X-\mu}{\sqrt{v}}$; the resulting random variable Z has expected value 0 and variance 1. For a location-scale family of probability distributions, given the density function $f(x)$ at $X = x$, the density function of Z at $Z = z$, denoted as $\phi(z)$, satisfies $f(x) = \frac{1}{\sqrt{v}}\phi(z)$, with $F(x) = \Phi(z)$, where $z = \frac{x-\mu}{\sqrt{v}}$ and $\Phi(z) = \int_{-\infty}^{z} \phi(u)du$. For any quantity Q, we define the standardized value of Q as $z_Q = \frac{Q-\mu}{\sqrt{v}}$, and the standardized loss function as $L(z_Q) = \frac{\ell(Q)}{\sqrt{v}}$.

In Section 3.1, we suggest DR standardized loss functions, using the worst- and best-case DR distributions discussed in the previous section. Section 3.2, then introduces standardized loss functions for a set of well-known and commonly applied continuous distributions.

3.1 DR Standardized Loss Functions

We next wish to derive an implied set of *DR standardized loss functions*. Characterizing standardized loss functions is especially valuable for certain operations planning models. For instance, the reader may refer to prior work on selective newsvendor problems [24] or stochastic knapsack problems [14]. The modeling and solution approaches used in those studies rely on the fact that the expected loss function can be expressed in the form $L_n(z)\sqrt{v}$, where $L_n(z)$ corresponds to a standardized normal loss function. The definition of DR standardized loss functions enables developing DR optimization approaches for these and other problem classes that take advantage of the problem structure resulting from using the standardized normal loss function.

Table 2 introduces standardized versions of the DR loss functions associated with Lemmas 2.1 and 2.2, i.e., $L_{[+]}^{DR}(z_Q) = \frac{\ell_{[+]}^{DR}(Q)}{\sqrt{v}}$ and $L_{[-]}^{DR}(z_Q) = \frac{\ell_{[-]}^{DR}(Q)}{\sqrt{v}}$. This table also provides the worst and best-case DR standardized CDFs where $F_{[+]}^{DR}(x) = \Phi_{[+]}^{DR}\left(\frac{x-\mu}{\sqrt{v}}\right) = \Phi_{[+]}^{DR}(z)$ and $F_{[-]}^{DR}(x) = \Phi_{[-]}^{DR}\left(\frac{x-\mu}{\sqrt{v}}\right) = \Phi_{[-]}^{DR}(z)$.

The worst-case standard distribution, $\Phi_{[+]}^{DR}$, contains a point mass with probability $\frac{1}{1+z_a^2}$ at $z = z_a$ and a point mass with probability $= \frac{1}{1+z_b^2}$ at $z = z_b$. Letting Z_{+}^{DR} denote the random variable with CDF Φ_{+}^{DR}, we define a corresponding censored random variable with a mixture of discrete and continuous terms as follows:

$$Z_{[+]}^{DR} = \begin{cases} z_a, & z_a \leq Z_{+}^{DR} < \frac{1}{2}\left(z_a - \frac{1}{z_a}\right), \\ Z_{+}^{DR}, & \frac{1}{2}\left(z_a - \frac{1}{z_a}\right) \leq Z_{+}^{DR} \leq \frac{1}{2}\left(z_b - \frac{1}{z_b}\right), \\ z_b, & \frac{1}{2}\left(z_b - \frac{1}{z_b}\right) < Z_{+}^{DR} \leq z_b. \end{cases} \tag{5}$$

Equation (5) provides the standardized form of the random variable $X_{[+]}^{DR}$ as defined by Equation (3). The best-case distribution in Lemma 2.2 then implies the following standardized distribution corresponding to $X_{[-]}^{DR}$, which was introduced in Equation (4).

$$Z_{[-]}^{DR} = \begin{cases} -\frac{1}{z_b}, & w.p.\ \frac{z_b}{z_b - z_a}, \\ -\frac{1}{z_a}, & w.p.\ -\frac{z_a}{z_b - z_a}. \end{cases} \tag{6}$$

Table 2 provides general results for standardized random variables bounded between a and b.

		Standardized DR Loss Function	Standardized Distribution
	Range for z_Q	$L_{[+]}^{DR}(z_Q)$	$\Phi_{[+]}^{DR}(z)$
Upper bound	$\left[z_a, \frac{1}{2}\left(z_a - \frac{1}{z_a}\right)\right)$	$-\frac{\frac{1}{z_a}+z_Q}{\left(\frac{1}{z_a}\right)^2+1}$	$\frac{1}{1+{z_0}^2}$
	$\left[\frac{1}{2}\left(z_a - \frac{1}{z_a}\right), \frac{1}{2}\left(z_b - \frac{1}{z_b}\right)\right]$	$L_{+}^{DR}(z_Q) = \frac{1}{2}\left(\sqrt{1+z_Q^2} - z_Q\right)$	$\Phi_{+}^{DR}(z) = \frac{1}{2}\left(1 + \frac{z}{\sqrt{z^2+1}}\right)$
	$\left[\frac{1}{2}\left(z_b - \frac{1}{z_b}\right), z_b\right)$	$\frac{(z_b - z_Q)}{1+{z_b}^2}$	$\frac{{z_b}^2}{1+{z_b}^2}$
	z_b	0	1
	Range for z_Q	$L_{[-]}^{DR}(z_Q)$	$\Phi_{[-]}^{DR}(z)$
Lower bound	$\left[z_a, -\frac{1}{z_b}\right)$	$-z_Q$	0
	$\left[-\frac{1}{z_b}, -\frac{1}{z_a}\right)$	$\frac{1+z_Q z_a}{z_b - z_a}$	$\frac{z_b}{z_b - z_a}$
	$\left[-\frac{1}{z_a}, z_b\right]$	0	1

Table 2: Standardized versions of DR loss functions.

(As we noted in Section 2, for unbounded distributions, we can obtain corresponding results by taking limits as z_a and/or z_b go to $-\infty$ and ∞, respectively.)

3.2 Standardized Loss Functions for a set of Continuous Random Variables

In this section, our goal is to create a tractable standardized loss function for each member of a manageable set of probability distributions with a broad range of possible properties, including a range of possible coefficient of variation (*cv*) values, distribution shapes, and possible value ranges (i.e., bounds). For example, while the normal distribution is symmetric (and "bell-shaped") and defined on the entire real line, the uniform distribution is symmetric, flat, and has finite upper and lower bounds. In addition, the gamma, Pareto, and lognormal distributions are nonnegative, and, depending on the parameter values, can provide a very large number of shapes, skewness, and tail thickness properties. The Pareto and lognormal distributions are characterized as *heavy-tailed distributions*, i.e., distributions whose tail is not bound by an exponential function (see, e.g., Bryson [3]), while the normal and gamma distributions are not heavy-tailed.

We also note that the uniform, normal, lognormal, and gamma distributions are maximum entropy distributions, implying that these distributions maximize uncertainty under various constraints on (or knowledge of) a distribution's parameter values (see, e.g., [12]). Thus, for example, if all we know about a distribution is that it is continuous and bounded between a and b (with $b > a$), then the uniform distribution provides the maximum entropy. If we know the distribution is nonnegative and has a known positive mean, then the exponential distribution (a special case of the gamma distribution) maximizes entropy. For a distribution with known mean and variance that may take any real value, a normal distribution maximizes entropy (for a random variable that must take positive values, where the expected value and variance of the natural log are known, the lognormal distribution maximizes entropy). If a random variable is positive with known mean μ and expected value of $\ln(\mu)$ equal to $\frac{\Gamma'(\mu)}{\Gamma(\mu)}$ (where $\Gamma(\mu)$ denotes the gamma function, i.e., $\Gamma(\mu) = \int_0^\infty x^{\mu-1} e^{-x} dx$), then the gamma distribution maximizes entropy.

Table 3 characterizes properties of these continuous distributions, namely normal, gamma, Pareto, lognormal, and uniform. Table 4 provides the functional forms of corresponding standardized expected loss functions for each of these distributions. In some cases, such as the normal, exponential, and uniform, the resulting standardized loss function is independent of specific distribution parameters. When the random variable takes positive values, i.e., $a = 0, b = \infty$, the coefficient of variation (cv), denoted as $\delta = \frac{\sqrt{v}}{\mu} = -\frac{1}{z_0}$ (which depends on the mean and standard deviation), is required in defining some of these functions.

	Normal(m, s)	**Gamma**(α, β)	**Pareto**(x_m, α)	**Lognormal**(m, s)	**Uniform**(l, u)
μ	m	$\frac{\alpha}{\beta}$	$\frac{\alpha x_m}{\alpha-1}$	$e^{m+\frac{s^2}{2}}$	$\frac{u+l}{2}$
$\sqrt{v}$	s	$\frac{\sqrt{\alpha}}{\beta}$	$\frac{x_m}{\alpha-1}\sqrt{\frac{\alpha}{\alpha-2}}$	$\sqrt{(e^{s^2}-1)(e^{m+\frac{s^2}{2}})^2}$	$\frac{u-l}{2\sqrt{3}}$
δ	$\frac{s}{m}$	$\frac{1}{\sqrt{\alpha}}$	$\frac{1}{\alpha}\sqrt{\frac{\alpha}{\alpha-2}}$	$\sqrt{e^{s^2}-1}$	$\frac{u-l}{(u+l)\sqrt{3}}$
$[z_a, z_b]$	$[-\infty, \infty]$	$[-\frac{1}{\delta}, \infty]$	$\left[-\sqrt{\frac{\sqrt{1+\frac{1}{\delta^2}}-1}{\sqrt{\frac{1}{\delta^2}+1}+1}}, \infty\right]$	$[-\frac{1}{\delta}, \infty]$	$[-\sqrt{3}, \sqrt{3}]$

Table 3: Properties of a set of well-known and broadly applicable continuous distributions.

Distribution, i	Standardized Loss Function, $L_i(z_Q)$
Normal (m, s)	$\phi_n(z_Q) - z_Q(1 - \Phi_n(z_Q))$
Exponential (β)	$e^{(1+z_Q\sqrt{1})}$
Gamma (α, β)	$\frac{1}{\delta}\left(1 - \frac{1}{\Gamma(\frac{1}{\delta^2})}\gamma\left(\frac{1}{\delta^2}+1, \frac{1}{\delta^2}+1+z_Q\sqrt{\frac{1}{\delta}+1}\right)\right) + \left(\sqrt{\frac{1}{\delta}}+z_Q\right)\left(1 - \frac{1}{\Gamma(\frac{1}{\delta^2})\gamma(\frac{1}{\delta^2}, \frac{1}{\delta^2}+\frac{1}{\delta}z_Q)}\right)$
Lognormal (m, s)	$\frac{1}{\delta}\left(\Phi\left(\frac{\sqrt{\ln(\delta^2+1)}}{2} - \frac{2\ln(1+z_Q\delta)}{\sqrt{l\ell(\delta^2+1)}}\right) - (1+z_Q\delta)\left(1 - \Phi\left(\frac{\sqrt{l\ell(\delta^2+1)}}{2} + \frac{2\ln(1+z_Q\delta)}{\sqrt{l\ell(\delta^2+1)}}\right)\right)\right)$
Uniform (l, u)	$\frac{\sqrt{3}}{4}\left(1 - \frac{z_Q}{\sqrt{3}}\right)^2$
Pareto (x_m, α)	$\frac{1}{\delta+\sqrt{1+\delta^2}}\left(\frac{1}{\delta+\sqrt{1+\delta^2}}\sqrt{\frac{1+\delta^2}{1+\delta^2 z_Q}}\right)^{\sqrt{1+\frac{1}{\delta^2}}}$

Table 4: Standardized loss functions for a set of continuous distributions.

The standardized DR loss functions, $L^{DR}_{[+]}(z_Q)$ and $L^{DR}_{[-]}(z_Q)$ provide valid upper and lower bounds for the loss functions in Table 4. For the distribution set under consideration, $L^{DR}_{[-]}(z_Q)$ only provides a useful lower bound on the standardized loss function of the uniform distribution, as this is the only one of these distributions with a finite range. In Section 5, we will propose tighter bounds on the loss function for the distributions discussed in this section (and on any distribution that is comprised of a weighted combination, or mixture, of these distributions). However, the following section first discusses useful generalization of the DR loss functions we have discussed thus far, which can lead to improved bounds and approximations for some classes of distributions.

4 Adjusted DR Loss Functions

The goal of this section is to provide stricter bounds on the loss functions discussed in the previous sections. We present the corresponding bounds, which we will refer to as *adjusted* DR upper and lower bounds, the next two subsections, respectively.

4.1 Adjusted DR Upper Bounds on Loss Functions

The key idea behind our adjusted upper bounds for loss functions is to use a smaller *scale parameter* for the t-distribution than the one suggested by the worst-case distribution, as we next discuss.

4.1.1 Adjusted Worst-case Distribution

We derive an adjusted distribution that replaces the variance v in Lemma 2.1 with the term κv, with $0 < \kappa \leq 1$. We use the superscript κ to denote the bound type (replacing the superscript DR in the loss functions and worst-case distributions in Table 1, the results of which are shown in Table 5). The resulting distribution corresponds to a censored random variable with a mixture of discrete and continuous terms as follows:

$$X_{[+]}^{\kappa} = \begin{cases} a, & a \leq X_{+}^{\kappa} < \frac{1}{2}\left(a + \mu - \frac{\kappa v}{a-\mu}\right), \\ X_{+}^{\kappa}, & \frac{1}{2}\left(a + \mu - \frac{\kappa v}{a-\mu}\right) \leq X_{+}^{\kappa} \leq \frac{1}{2}\left(b + \mu - \frac{\kappa v}{b-\mu}\right), \\ b, & \frac{1}{2}\left(b + \mu - \frac{\kappa v}{b-\mu}\right) < X_{+}^{\kappa} \leq b. \end{cases} \tag{7}$$

When the random variable is unbounded, we then have the worst-case distribution for the random variable X_{+}^{κ}, which is equivalent to the random variable X_{+}^{DR} when the variance equals κv. The random variable $X_{[+]}^{\kappa}$ in (7) follows a three-parameter Student's t-distribution with location parameter μ, scale parameter $\sqrt{\frac{\kappa v}{2}}$ and two degrees of freedom. The associated cumulative distribution function, $F_{[+]}^{\kappa}$, is provided in Table 5. This random variable continues to have mean μ and variance ∞, as does the t-distribution associated with the worst-case distribution in the previous section. We also make the following observations:

		Adjusted DR Loss Function	Adjusted Distribution
	Range for Q	$\ell^{\kappa}_{[+]}$	$F^{\kappa}_{[+]}(Q)$
Upper bound	$\left[a, \frac{1}{2}\left(a+\mu-\frac{\kappa v}{a-\mu}\right)\right)$	$-(Q-\mu)+(Q-a)\frac{\kappa v}{(\mu-a)^2+\kappa v}$	$\frac{\kappa v}{(\mu-a)^2+\kappa v}$
	$\left[\frac{1}{2}\left(a+\mu-\frac{\kappa v}{a-\mu}\right), \frac{1}{2}\left(b+\mu-\frac{\kappa v}{b-\mu}\right)\right]$	$\ell^{\kappa}_{+}(Q)=\frac{1}{2}\left(\sqrt{\kappa v+(Q-\mu)^2}-(Q-\mu)\right)$	$F^{\kappa}_{+}(x)=\frac{1}{2}\left(1+\frac{x-\mu}{\sqrt{(x-\mu)^2+\kappa v}}\right)$
	$\left(\frac{1}{2}\left(b+\mu-\frac{\kappa v}{b-\mu}\right), b\right)$	$\frac{\kappa v(b-Q)}{\kappa v+(b-\mu)^2}$	$\frac{(b-\mu)^2}{(b-\mu)^2+\kappa v}$
	b	0	1
	Range for Q	$\ell^{\tau}_{[-]}$	$F^{\tau}_{[-]}(Q)$
Lower bound	$\left[a, \left(\mu-\frac{\tau v}{b-\mu}\right)\right)$	$\mu-Q$	0
	$\left[\left(\mu-\frac{\tau v}{b-\mu}\right), \left(\mu-\frac{\tau v}{a-\mu}\right)\right)$	$\mu-\frac{(\mu-a)Q+(b-\mu)\mu-\tau v}{b-a}$	$\frac{b-\mu}{b-a}$
	$\left[\left(\mu-\frac{\tau v}{a-\mu}\right), b\right]$	0	1

Table 5: Adjusted DR loss functions.

- When $\kappa = 2$, F^{κ}_{+} corresponds to a three-parameter Student's t-distribution with location parameter μ, scale parameter $\sqrt{v} > 0$ and two degrees of freedom.

- When $\kappa = 1$, F^{κ}_{+} corresponds to F^{DR}_{+} (the DR bound setting discussed in Section 2).

We next consider the associated standardized variable, $Z^{\kappa}_{[+]} = \frac{X^{\kappa}_{[+]}-\mu}{\sqrt{v}}$ obeying the cumulative distribution $\Phi^{\kappa}_{[+]}(z)$ shown in Table 6, which provides the standardized version of the results in Table 5. This distribution corresponds to a censored random variable with a mixture of discrete and continuous terms as follows:

$$Z^{\kappa}_{[+]} = \begin{cases} z_a, & z_a \leq Z^{\kappa}_{+} < \frac{1}{2}\left(z_a - \frac{\kappa}{z_a}\right), \\ Z^{\kappa}_{+}, & \frac{1}{2}\left(z_a - \frac{\kappa}{z_a}\right) \leq Z^{\kappa}_{+} \leq \frac{1}{2}\left(z_b - \frac{\kappa}{z_b}\right), \\ z_b, & \frac{1}{2}\left(z_b - \frac{\kappa}{z_b}\right) < Z^{\kappa}_{+} \leq z_b. \end{cases} \tag{8}$$

Note that Z^{κ}_{+} follows the distribution $\Phi^{\kappa}_{+}(z)$ where $F^{\kappa}_{+}(x) = \Phi^{\kappa}_{+}\left(\frac{x-\mu}{\sqrt{v}}\right) = \Phi^{\kappa}_{+}(z)$. For this standardized random variable, we also note similar observations:

- When $\kappa = 2$, Φ^{κ}_{+} corresponds to a standard Student's t-distribution.

- When $\kappa = 1$, Φ^{κ}_{+} corresponds to Φ^{DR}_{+} as defined in Table 2.

We will refer to an approach that uses the DR approach with a parameter κ value less than one ($\kappa < 1$) as an *adjusted DR* approach.

		Standardized Adjusted DR Loss Function	Standardized Adjusted Distribution
	Range for z_Q	$L^{\kappa}_{[+]}(z_Q)$	$\Phi^{\kappa}_{[+]}(z)$
Upper bound	$\left[z_a, \frac{1}{2}\left(z_a - \frac{\kappa}{z_a}\right)\right)$	$-\frac{\frac{\kappa}{z_a}+z_Q}{\kappa\left(\frac{1}{z_a}\right)^2+1}$	$\frac{\kappa}{\kappa+{z_0}^2}$
	$\left[\frac{1}{2}\left(z_a - \frac{\kappa}{z_a}\right), \frac{1}{2}\left(z_b - \frac{\kappa}{z_b}\right)\right]$	$L^{\kappa}_{+}(z_Q) = \frac{1}{2}\left(\sqrt{\kappa + z_Q^2} - z_Q\right)$	$\Phi^{\kappa}_{+}(z) = \frac{1}{2}\left(1 + \frac{z}{\sqrt{z^2+\kappa}}\right)$
	$\left(\frac{1}{2}\left(z_b - \frac{\kappa}{z_b}\right), z_b\right)$	$\frac{\kappa(z_b-z_Q)}{\kappa+{z_b}^2}$	$\frac{{z_b}^2}{\kappa+{z_b}^2}$
	z_b	0	1
	Range for z_Q	$L^{\tau}_{[-]}(z_Q)$	$\Phi^{\tau}_{[-]}(z)$
Lower bound	$\left[z_a, -\frac{\tau}{z_b}\right)$	$-z_Q$	0
	$\left[-\frac{\tau}{z_b}, -\frac{\tau}{z_a}\right)$	$\frac{\tau+z_Q z_a}{z_b-z_a}$	$\frac{z_b}{z_b-z_a}$
	$\left[-\frac{\tau}{z_a}, z_b\right]$	0	1

Table 6: Standardized adjusted DR loss functions.

4.1.2 Optimal Parameter Selection

We next consider the adjusted DR loss function, $\ell^{\kappa}_{[+]}(Q) = E\left[(X^{\kappa}_{[+]} - Q)^+\right]$, as shown in Table 5. Letting $z_Q = \frac{Q-\mu}{\sqrt{v}}$, Table 6 provides an adjusted standardized DR loss function $L^{\kappa}_{[+]}(z_Q) = \frac{\ell^{\kappa}_{[+]}(Q)}{\sqrt{v}}$. It is straightforward to show that $L^{\kappa}_{[+]}(z_Q) \leq L^{DR}_{[+]}(z_Q)$ for all z_Q when $\kappa \leq 1$, with strict inequality holding when $\kappa < 1$. Given a set of possible probability distributions with given mean and variance values (i.e., a common ambiguity set); we can then search for an appropriate value of $0 < \kappa \leq 1$ such that the resulting bound on the expected loss function remains valid for the set under consideration.

To provide stricter upper bounds, we seek the smallest possible κ value such that $L_i(z_Q) \leq L^{\kappa}_{[+]}(z_Q)$ for a distribution i in the ambiguity set. We first show that when using the value of κ satisfying $L_i(z_Q) \leq L^{\kappa}_{+}(z_Q)$ for a given distribution, the resulting function $L^{\kappa}_{[+]}(z_Q)$ provides a valid upper bound on the value of the loss function of the corresponding distribution (say, distribution i). This is clearly true for the middle interval because $L^{\kappa}_{[+]}(z_Q) = L^{\kappa}_{+}(z_Q)$ for z_Q on the corresponding interval, and $L^{\kappa}_{+}(z_Q) \geq L_i(z_Q)$ for all z_Q by construction. We can also show that (a) $L^{\kappa}_{+}(z_Q)$ is a convex function, (b) $L^{\kappa}_{[+]}(z_Q)$ is a continuous function, (c) $L^{\kappa}_{[+]}(z_Q) = L_i(z_Q)$ at $z_Q = z_a$ and at $z_Q = z_b$, and (d) $L^{\kappa}_{[+]}(z_Q)$ is linear in z_Q on the first and third intervals. The convexity of $L_i(z_Q)$ along with the above properties implies that $L^{\kappa}_{[+]}(z_Q)$ provides an upper bound on $L_i(z_Q)$ for $z_Q \in [z_a, z_b]$ (a detailed proof is provided in the appendix).

We therefore limit ourselves to the middle interval and seek the smallest possible κ value such that $L_i(z_Q) \leq L^{\kappa}_{+}(z_Q)$ for a possible distribution i in the ambiguity set. This is equivalent to

$$\kappa \geq g_i(z_Q, \delta) = 4L_i(z_Q)(z + L_i(z_Q)) = 4L_i(z_Q)\tilde{L}_i(z_Q), \tag{9}$$

where $\tilde{L}_i(z_Q) = z + L_i(z_Q)$ is the standardized version of the *expected leftover*, i.e., $\tilde{L}_i(z_Q) = \frac{E\left[(Q-X)^+\right]}{\sqrt{v}}$. Thus, determining the smallest value of κ that satisfies (9) requires maximizing the product of the standardized loss and standardized leftover functions. The function $g_i(z_Q, \delta)$ is neither concave nor convex in general in the variable z_Q (for distributions such as the uniform and normal, where g_i is independent of δ, and with a slight abuse of notation, we will write $g_i(z)$, suppressing the dependence of z on Q). Figure 1 illustrates this function for standardized gamma, normal, Pareto, uniform, and lognormal distributions (where we set $\delta = 1$ for gamma, Pareto and lognormal distributions for illustrative purposes).

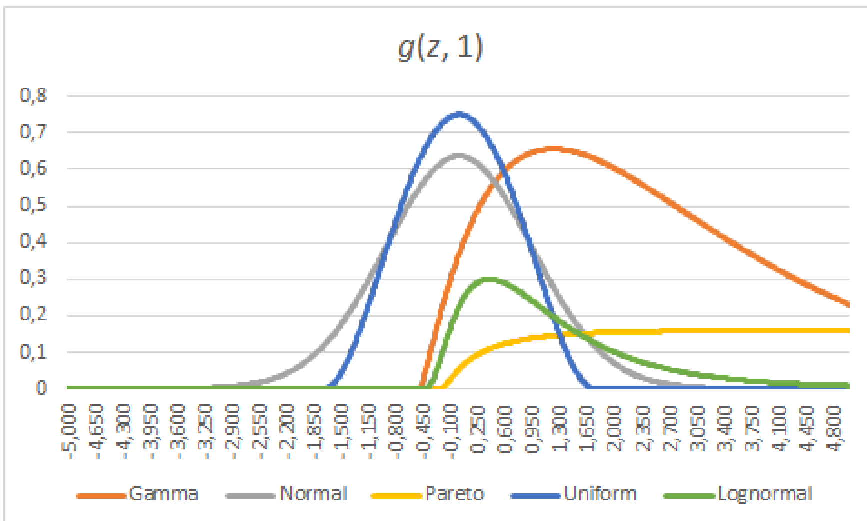


Figure 1: Illustration of $g_i(z_Q, \delta)$ when $\delta = 1$ for various distributions.

The first-order condition for $g_i(z)$ can be written as $L_i(z)\Phi_i(z) = (1 - \Phi_i(z))\tilde{L}_i(z)$, which is equivalent to $\frac{E\left[(u-z)^+\right]}{1-\Phi_i(z)} = \frac{E\left[(z-u)^+\right]}{\Phi_i(z)}$, and can be interpreted as requiring that the conditional expected shortage equals the conditional expected leftover. Furthermore, this condition is equivalent to $z = \frac{E[u|u \geq z] + E[u|u \leq z]}{2}$. Observe that for symmetric distributions, this condition is satisfied at $z = 0$. Under a symmetric distribution, at $z = 0$, $L(z) = \tilde{L}(z)$, so that $\frac{\tilde{L}(z)}{\tilde{L}(z)+L(z)} = \frac{1}{2}$. The following lemmas, the proofs of which are provided in the appendix, characterize properties of $g_i(z_Q, \delta)$.

Lemma 4.1 *For any continuous probability distribution, $g_i(z_Q, \delta) \geq 0$ on its domain, while for any continuous symmetric distribution, $g_i(z_Q, \delta)$ has a stationary point at $z_{0.5} = 0$, where $z_{0.5}$ denotes the 50^{th} percentile of the distribution.*

For any symmetric probability distribution that has a corresponding increasing failure rate (IFR) function[1], we can also show the following lemma analytically (we will say that any distribution

[1]In reliability theory, the failure rate for a distribution with pdf $\phi(z)$ and CDF $\Phi(z)$ is given by $\frac{\phi(z)}{1-\Phi(z)}$.

with this property is an IFR distribution). This class of IFR distributions includes the normal, continuous uniform, symmetric triangular, logistic, Laplace (double exponential), and beta (with parameters $\alpha = \beta \geq 0.8$) distributions, among others.

Lemma 4.2 *For any IFR symmetric distribution, $g_i(z_Q, \delta)$ has a global maximum at $z_Q = 0$.*

Note that the IFR condition for a symmetric distribution in Lemmar 4.2 provides a *sufficient* condition for ensuring that a global maximum of $g_i(z_Q, \delta)$ exists at $z_Q = 0$. For example, the beta distribution with $\alpha = \beta < 0.8$ is symmetric but is not IFR, and we are able to show numerically that $g_{beta}(z_Q, \delta)$ is nevertheless maximized at $z_Q = 0$. We next state two additional lemmas that can provide general guidance on the selection of an appropriate κ value. Henceforth, let d denote the mean absolute deviation (MAD) of a distribution, i.e. $d = E\left[|X - \mu|\right]$.

Lemma 4.3 *For any continuous and symmetric IFR distribution, the adjusted DR loss function with $\kappa = \frac{d^2}{v}$ provides a valid upper bound, and this bound is tight at $z_Q = 0$.*

Lemma 4.4 *For any continuous and asymmetric distribution, there exists a valid and tight adjusted DR loss function using κ where $\frac{d^2}{v} < \kappa \leq 1$.*

Lemma 4.3 provides a fairly surprising result that can be quite powerful when the MAD is known for a symmetric IFR distribution. In subsequent sections, we will focus on particular ambiguity sets and analyse the implications of these lemmas in greater detail. Before doing so, we first provide adjusted DR lower bounds for loss functions.

4.2 Adjusted DR Lower Bounds for Loss Functions

To provide stricter lower bounds on the loss function, one may simply utilize a *larger* variance term in the DR lower bound. We therefore consider the DR lower bounds provided by Lemma 2.2 with variance τv where $\tau \geq 1$. The implied best-case distribution for this bound is given by

$$X^{\tau}_{[-]} = \begin{cases} \mu - \frac{\tau v}{b-\mu}, & w.p.\ \frac{b-\mu}{b-a}, \\ \mu - \frac{\tau v}{a-\mu}, & w.p.\ \frac{\mu-a}{b-a}. \end{cases} \tag{10}$$

Table 6 provides a formula for the standardized loss function for this distribution $L^{\tau}_{[-]}(z_Q) = E\left[\left(X^{\tau}_{[-]} - Q\right)^+\right]$. It is easy to see that $L^{\tau}_{[-]}(z_Q) \geq L^{DR}_{[-]}(z_Q)$ when $\tau \geq 1$, with strict inequality holding when $\tau > 1$. We next search for an appropriate value of τ such that the resulting bound on the expected loss function is still valid for the ambiguity set under consideration.

Considering the three intervals for $L^{\tau}_{[-]}(z_Q)$ in Table 6, we first note that the value of τ has no effect on the first interval, as $-z_Q \leq L_i(z_Q)$ for any distribution i. This is because $L_i(z_Q) + z_Q \geq 0$ for all z_Q, as this quantity denotes the expected leftover. We then require $L^{\tau}_{[-]}(z_Q) \leq L_i(z_Q)$ for the middle interval. This implies that the constant τ must satisfy the relationship

$$\tau \leq t_i(z_Q) = L_i(z_Q)(z_b - z_a) - z_Q z_a = L_i(z_Q) z_b - \tilde{L}_i(z_Q) z_a. \tag{11}$$

It is straightforward to show that $t_i(z_Q)$ is convex and is minimized at z_Q where $\Phi_i(z_Q) = \frac{z_b}{z_b - z_a}$. For symmetric distributions, this function is minimized when $\Phi_i(z_Q) = 0.5$ (or $z_Q = 0$) as $z_b = -z_a = \frac{b-\mu}{\sqrt{v}} = \frac{b-a}{2\sqrt{v}}$. We then have $\tau = 2L_i(0) z_b = \frac{d}{\sqrt{v}} z_b = \frac{d(b-a)}{2v}$. We can also show that $\tau \geq 2\sqrt{\kappa^*}$ where $\kappa^* = \frac{d^2}{v}$ (see Appendix A.4 for proof of this result). The following lemmas formalize these results, which lead to a tighter lower bound for the loss function for any symmetric distribution when the MAD is known in addition to the variance.

Lemma 4.5 *For any continuous and symmetric distribution, the adjusted lower bound for the loss function using* $\tau = \frac{d(b-a)}{2v}$ *is valid and tight at* $z_Q = 0$.

Lemma 4.6 *For any continuous distribution, there exists a valid and tight lower bound for the loss function using* τ *where* $1 \leq \tau \leq \frac{d(b-a)}{2v}$.

5 DR Loss Functions for a Mixture Ambiguity Set

In this section, we continue to assume that X has mean μ and variance v. We also assume this random variable is continuous and has a mixture density. That is, the probability density function can be expressed as a convex combination of a set of other density functions (i.e., a weighted sum, with non-negative weights that sum to one). In particular, suppose we have n probability density functions, $h_1(x), h_2(x), ..., h_n(x)$, each with the same expected value and variance i.e., $\mu_i = \mu$ and $v_i = v$ for $i = 1, \ldots, n$. If X follows a mixture distribution, then its density, $h_w(x) = \sum_{i=1}^n w_i h_i(x)$, where $\sum_{i=1}^n w_i = 1$. Note that X has the same mean and variance as each of its mixture components.

For practical purposes, we limit ourselves to the set of distributions introduced previously, i.e., $n = 5$, with $i =$ *normal, uniform, gamma, Pareto, lognormal.* This set of distributions provides a wide range of potential skewness and kurtosis values, and thus, distribution shapes. Although we have a finite number of mixture components within this set of distributions, we assume that the weights are not known to the decision maker. Thus, the number of combinations of distributions and associated weights encompasses an infinite number of different distribution shapes. In the next section, we introduce the standardized loss functions for the random distributions in this set.

5.1 Maximum of Loss Functions for a Mixture Ambiguity Set

For a given z_Q, we define $L_{max}(z_Q) = \max_{i=1,...,n} L_i(z_Q)$, where $L_{max}(z_Q)$ is defined by the upper envelope of the loss functions $L_i(z_Q), i = 1, ..., n$. Each $L_i(z_Q)$ is convex and nonincreasing in z_Q. As the maximum among a set of convex and nonincreasing functions, $L_{max}(z_Q)$ is also convex and nonincreasing. Next, observe that the expected loss function for any weighted distribution satisfies

$$\begin{aligned}\ell(Q) &= E\left[(X-Q)^+\right] = \int_Q^\infty (x-Q)\ell(x)dx = \sum_{i=1}^n w_i \int_Q^\infty (x-Q)h_i(x)dx = \sum_{i=1}^n w_i n_i(Q) \\ &= \sqrt{v}\sum_{i=1}^n w_i L_i(z_Q) \leq \sqrt{v} L_{max}(z_Q)\end{aligned}$$

Thus, $L_{max}(z_Q)$ provides a convex function that serves as an upper bound on the loss function value for each of the distributions included within the mixture, *as well as for any of the infinite collection of weighted distributions that can be constructed from this underlying set of distributions.* If we have sufficient confidence that the actual distribution can be closely approximated using the mixture distribution described, we can use $L_{max}(z_Q) < L_{[+]}^{DR}(z_Q)$ as a less conservative (but valid) bound on the worst-case value of the expected loss function. Figure 2 illustrates the bound provided by $L_{max}(z_Q)$ (the dotted curve labeled "maximum"), as well as the apparent gap between this bound and Scarf's DR bound (labeled DR), assuming a *cv* equal to one. Thus, we can argue that if the set of underlying distributions is sufficiently broad, the resulting bounding function is quite distributionally robust, while at the same time being considerably less conservative than the common approach found in the existing literature.

We may choose to exclude certain distributions from the mixture if they are inconsistent with the available contextual data. That is, if we know the mean μ and variance v of the underlying distribution, then this implies a fixed value of the distribution's *cv* of $\delta = \frac{\sqrt{v}}{\mu}$. A uniform distribution on the interval $[l, u]$ with $0 \leq l < u$, for example, implies $\delta = \frac{u-l}{\sqrt{3}(u+l)} \leq \frac{1}{\sqrt{3}}$. Similarly, a normal distribution with $\delta > \frac{1}{3}$ implies a nonnegligible probability of a negative demand value. Thus, for example, if we observe a value of $\delta > \frac{1}{\sqrt{3}}$, we may exclude the normal and uniform distributions from the mixture because they are inconsistent with the data. Alternatively, we may choose to add distributions that fit with the available data. For example, if $\delta = 1$, we can directly use a

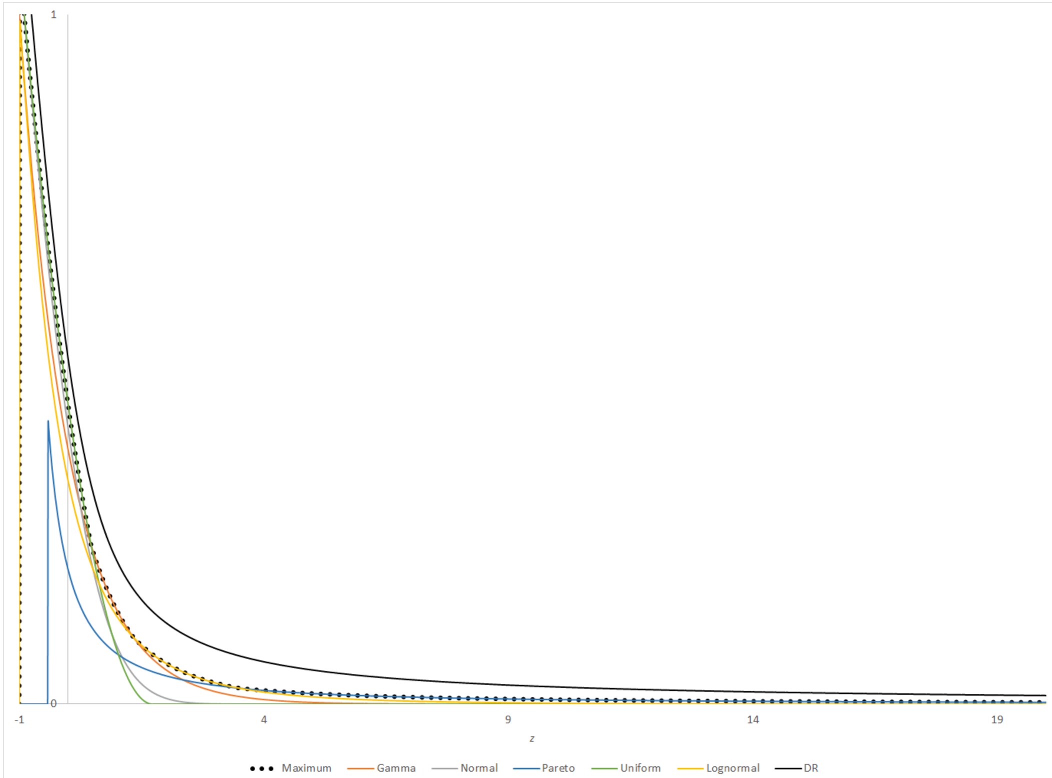


Figure 2: Maximum (upper envelope) of Standardized Loss Functions when $\delta = 1$.

standardized version of the exponential distribution, which has the form $L(z_Q) = e^{-z-1}$.

5.2 Adjusted Loss Functions for a Mixture Ambiguity Set

In this section, we provide adjusted DR bounds for the mixture ambiguity set based on the approach described in Section 4. Given the distributions in the mixture ambiguity set, one may calculate a corresponding κ_i value for each individual distribution. For the mixture distributions, the adjusted DR loss function with $\kappa = \arg\max_i\{\kappa_i\}, i = 1, ..., n$ provides a valid bound for the corresponding loss function for the mixture.

For each of the distributions shown in Figure 1, the maximum of $g(z_Q, \delta)$ occurs at a value $z_Q = z^*$ where the partial derivative is 0, i.e., $\frac{z^*+L_i(z^*)}{z^*+2L_i(z^*)} = F_i(z^*)$, where $\kappa_i = 4L_i(z^*)\tilde{L}_i(z^*)$. Lemma 4.3 implies that for a normal distribution, $\kappa_n = \frac{2}{\pi} = 0.6366$ (for the normal distribution, $d^2 = \frac{2v}{\pi}$). Similarly, for the continuous uniform and symmetric triangular distributions, we can show that $\kappa_u = \frac{3}{4}$ and $\kappa_t = \frac{2}{3}$, respectively. For each of the other distributions considered in the mixture, we performed a search among the z values at cv levels between and including 0 and 20, and determined that $\kappa = 0.75$ provides a valid bound for each of these distributions. When

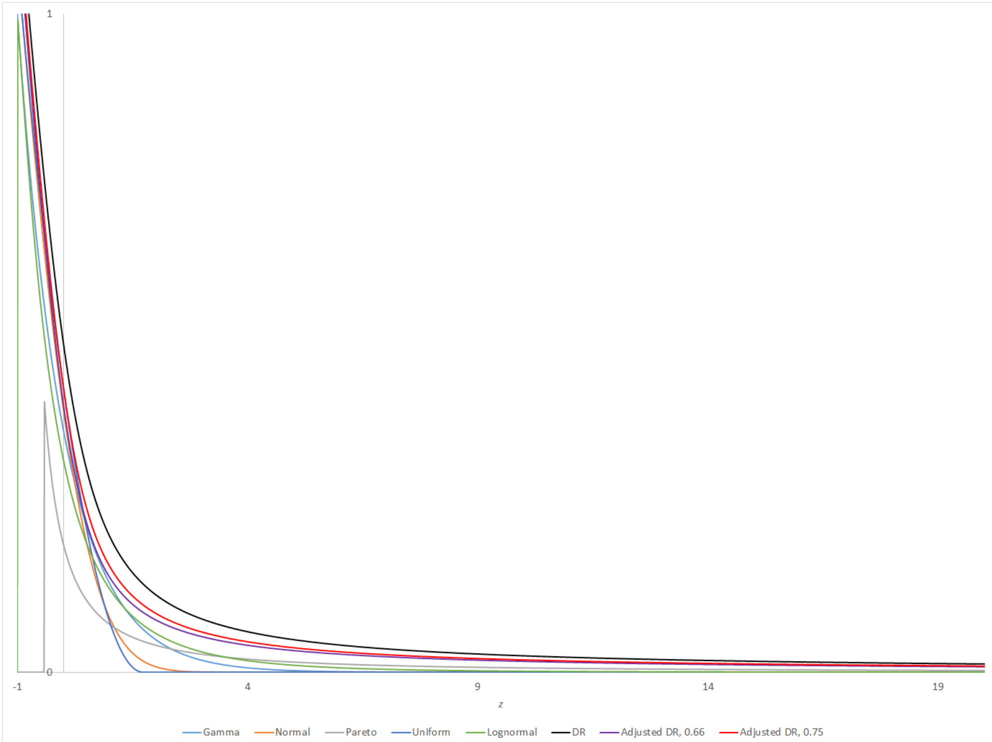


Figure 3: Standardized loss function $L(z_Q)$, when $\delta = 1$.

the continuous uniform and symmetric triangular distributions are not part of the mixture set, we found that $\kappa = 0.663$ is valid for the remaining distributions we considered, where this bound is tight for a gamma distribution. Figure 3 illustrates various loss functions when $\delta = 1$. In addition to showing the loss function for distributions in the mixture set, the figure also illustrates the DR loss function, as well as the adjusted bounds when using $\kappa = 0.75$ and $\kappa = 0.663$. Observe that a noticeable gap exists between the DR bound and the adjusted bounds at standardized z values less than or equal to 9 when $\kappa = 0.75$ (and this gap is larger when $\kappa = 0.663$).

Lemma 4.3 allows us to analytically determine κ_i for distributions i that are continuous, symmetric, and IFR, assuming the MAD and variance are known, using $\kappa = \frac{d^2}{v}$. Note that the required magnitude of κ is related to the tail properties of the distribution. For example, the beta distribution when $\alpha = \beta$ is a short-tailed, symmetric distribution that nicely illustrates a continuum of (nondecreasing) values of the ratio $\kappa = \frac{d^2}{v}$ between $\frac{2}{\pi}$ and 1, as the value at which $\alpha = \beta$ decreases from ∞ to 0.[2] Thus, for a symmetric beta distribution on an interval with a fixed width, as the weight of the distribution increasingly shifts from the center of the distribution to the tails, the

[2]When $\alpha \to \infty$, the beta distribution converges to a normal distribution, while when $\alpha \to 1$, it converges to a uniform distribution; when $\alpha \to 0$, it converges to a Bernoulli(0.5) distribution. For the normal, uniform, and Bernoulli(0.5) distributions, the ratio $\frac{d^2}{v}$ equals $\frac{2}{\pi} \approx 0.6366$, 0.75 and 1, respectively.

required value of κ increases. The logistic and Laplace (double exponential) distributions provide examples of symmetric IFR distributions with longer tails (relative to the normal) for which the ratio $\frac{d^2}{v}$ equals 0.58 and 0.5, respectively. (Thus, we see that longer-tailed distributions permit lower upper bounds on the loss function.)

When the distribution is asymmetric, we showed that κ must exceed $\frac{d^2}{v}$. Note also, however, that the ratio $\frac{d^2}{v}$ itself decreases as the distribution becomes more skewed.[3] For example, for a fixed cv of $\delta = 1$, we observed that the values of κ permitted for the lognormal and Pareto distributions are much smaller than for the gamma distribution (see Figure 1, where $\kappa < 0.3$ for these distributions while κ is 0.663 for the gamma). Both the lognormal and Pareto distributions have heavier tails on the right side compared to the gamma distribution. Therefore, it is possible to use smaller κ values when data implies right-skewed, heavy tailed distributions.

6 DR Loss Functions for Symmetric IFR Distributions

Lemma 4.3 showed that the adjusted DR upper bound of the loss function is tight at $z_Q = 0$ for symmetric IFR distributions when $\kappa = \frac{d^2}{v}$, where d denotes the mean absolute deviation (MAD). Similarly, Lemma 4.5 states that the adjusted DR lower bound provides a tight lower bound on the loss function of a symmetric distribution when $\tau = \frac{d(b-a)}{2v}$. Therefore, we have shown that when MAD information is available, in addition to mean, variance, and range information, the existing *mean-variance*-based upper bounds in the literature (given by Lemma 2.1) can substantially be improved for symmetric IFR distributions, while the lower bounds (given by Lemma 2.2) can be improved for symmetric bounded distributions.

In Section 6.1, we show that an adjusted DR upper bound can be provided for symmetric IFR distributions when we have MAD information instead of variance, while a corresponding adjusted DR lower bound can be provided for symmetric distributions. That is, our bounds can be written under symmetry and *mean-MAD ambiguity*, defined as an ambiguity set in which we only know the mean μ, mean absolute deviation d, and range $[a, b]$ of the underlying distribution. In Section 6.2 we then compare the resulting bounds with available *mean-MAD* bounds in the literature.

[3]For any distribution on $[a, b]$, $\frac{2v}{(b-a)} \leq d \leq 2\sqrt{v\beta(1-\beta)} \leq \sqrt{v}$ where $\beta = P(x \geq \mu)$ [17]. For highly asymmetric distributions the value of β gets closer to 0 or 1, lowering the upper bound on d.

6.1 Adjusted DR Loss Functions for Symmetric IFR Distributions

For symmetric IFR distributions (symmetric distributions), the adjusted DR upper (lower) bounds under mean-MAD ambiguity are given in Table 7, which is obtained by substituting d^2 for κv ($\frac{d(b-a)}{2}$ for τv) in Table 5, due to Lemmas 4.3 and 4.5. These bounds can also be obtained by using d^2 $\left(\frac{d(b-a)}{2}\right)$ in place of v in the mean-variance upper (lower) bounds given by Lemma 2.1 (Lemma 2.2). We let $\ell^{\kappa*}_{[+]}$ and $\ell^{\tau*}_{[-]}$ denote these upper and lower bounds, respectively. Note that due to symmetry, we may further simplify the entries in the table by letting $\omega = 2(b-\mu) = 2(\mu-a) = b-a$.

Interestingly, these bounds are solely based on mean-MAD information. Therefore, when the decision maker has mean, MAD, range and symmetry information, the lower bound provided by Table 7 is tight, while the the upper bound is valid and tight for symmetric IFR distributions. In the next section, our objective is to compare our adjusted DR bounds with existing mean-MAD bounds in the literature. As our upper bound (Lemma 4.3) and lower bound (Lemma 4.5) results are provided for symmetric IFR and symmetrics distributions, respectively, we limit the ambiguity set by allowing only *symmetric* IFR distributions with a given mean μ, range $[a,b]$ and MAD d when discussing upper bounds, and symmetric distributions when discussing the lower bounds.

		Adjusted DR Loss Function	Adjusted Distribution
	Range for Q	$\ell^{\kappa*}_{[+]}$	$F^{\kappa*}_{[+]}(Q)$
Upper bound	$\left[a, \frac{1}{2}\left(a+\mu-\frac{d^2}{a-\mu}\right)\right)$	$-(Q-\mu)+(Q-a)\frac{d^2}{(\mu-a)^2+d^2}$	$\frac{d^2}{(\mu-a)^2+d^2}$
	$\left[\frac{1}{2}\left(a+\mu-\frac{d^2}{a-\mu}\right), \frac{1}{2}\left(b+\mu-\frac{d^2}{b-\mu}\right)\right]$	$\ell^{\kappa}_{+}(Q) = \frac{1}{2}\left(\sqrt{d^2+(Q-\mu)^2}-(Q-\mu)\right)$	$F^{\kappa}_{+}(x) = \frac{1}{2}\left(1+\frac{x-\mu}{\sqrt{(x-\mu)^2+d^2}}\right)$
	$\left(\frac{1}{2}\left(b+\mu-\frac{d^2}{b-\mu}\right), b\right)$	$\frac{d^2(b-Q)}{d^2+(b-\mu)^2}$	$\frac{(b-\mu)^2}{(b-\mu)^2+d^2}$
	b	0	1
	Range for Q	$\ell^{\tau*}_{[-]}$	$F^{\tau*}_{[-]}(Q)$
Lower bound	$\left[a, \left(\mu-\frac{\frac{d(b-a)}{2}}{b-\mu}\right)\right)$	$\mu-Q$	0
	$\left[\left(\mu-\frac{\frac{d(b-a)}{2}}{b-\mu}\right), \left(\mu-\frac{\frac{d(b-a)}{2}}{a-\mu}\right)\right)$	$\mu-\frac{(\mu-a)Q+(b-\mu)\mu-\frac{d(b-a)}{2}}{b-a}$	$\frac{b-\mu}{b-a}$
	$\left[\left(\mu-\frac{\frac{d(b-a)}{2}}{a-\mu}\right), b\right]$	0	1

Table 7: DR loss function upper bounds for symmetric IFR distributions, and lower bounds for symmetric distributions, with MAD information.

6.2 DR Loss Functions for Mean-MAD Ambiguity Sets

We next discuss bounds provided by prior literature when dispersion information is in the form of mean absolute deviation, i.e., a *mean-MAD ambiguity set*. Ben-Tal and Hochman [1] developed tight upper and lower bounds on the expected value of a convex function of a random variable (not necessarily symmetric) for such ambiguity sets (see Lemmas A.1 and A.2 in the Appendix).

Several researchers later utilized these bounds to solve DR optimization problems (see for example [17] and [25]). As $(X-Q)^+$ is a convex function of the random variable X, one can also derive corresponding bounds on the expected loss function, $E\left[(X-Q)^+\right]$ (see Appendix A.5 and A.6).

6.2.1 Loss Function Upper Bound for Mean-MAD Ambiguity Sets

The upper bound for $E\left[(X-Q)^+\right]$ for the *mean-MAD* ambiguity set can be formulated below by utilizing key results from Ben-Tal and Hochman [1]. The reader is referred to Appendix A.5 for the derivation of this upper bound, which can be written as

$$\ell_{[+]}^{MAD}(Q) = \begin{cases} (\mu - Q)\left(1 - \frac{2d}{\omega}\right) + (b-Q)\frac{d}{\omega}, & a \le Q \le \mu \\ (b-Q)\frac{d}{\omega}, & \mu \le Q \le b. \end{cases} \tag{12}$$

We next compare this upper bounding function with the adjusted DR upper bound ($\ell_{[+]}^{\kappa*}$) provided in Table 7. Note that both functions are equal and tight at three points, namely a (taking the value of $\mu - a$), $Q = \mu$ (taking the value of $\frac{d}{2}$), and $Q = b$ (taking the value of 0). Although these two upper bounding functions coincide at these three points, we can show that our suggested adjusted DR bound improves over the mean-MAD bound (12). One can see this by noting that the mean-MAD bound and adjusted DR bound start and end at the same points and are equal at $Q = \mu$. Since the middle section of the adjusted DR bound is a strictly convex function, this bound is tighter than the mean-MAD bound, which is a piece-wise linear convex function.

Note that the mean-MAD upper bound above is based on a discrete distribution with point masses at a, μ, and b. Using the adjusted DR bound, we showed that the corresponding probability distribution implied by this bound is a mixture of discrete points (having mass at a and b) and a continuous function between these points, given by (7). This greatly improves not only the bound provided, but also the applicability and utility of the bounding function as an approximation for symmetric IFR loss functions for optimization problems in operations. Furthermore, this allows us to utilize a MAD-based bound even when upper and lower bound (a, b) information is not available.

6.2.2 Loss Function Lower Bounds for Mean-MAD Ambiguity Sets

A lower bound for $E\left[(X-Q)^+\right]$ for the *mean-MAD* ambiguity set can also be derived based on the results of Ben-Tal and Hochman [1]. In order to obtain a useful MAD-based lower bound, we require information on the value of β such that $P(X \ge \mu) = \beta$ for the ambiguity set ([1] and [17]); clearly this value is $\beta = 0.5$ for any symmetric distribution. The reader is referred to Appendix

A.6 for explicit derivation of the corresponding lower bounding function, which can be written as

$$\ell_{[-]}^{MAD}(Q) = \begin{cases} \mu - Q, & a \le Q \le \mu - d, \\ \frac{\mu+d-Q}{2}, & \mu - d \le Q \le \mu + d, \\ 0 & \mu + d \le Q \le b. \end{cases} \tag{13}$$

Note that this bound is the same as our adjusted lower bound suggested in Table 7. Therefore, our adjusted bound does not provide any improvement over the existing *mean-MAD* lower bound from prior literature, even in the symmetric distribution case.

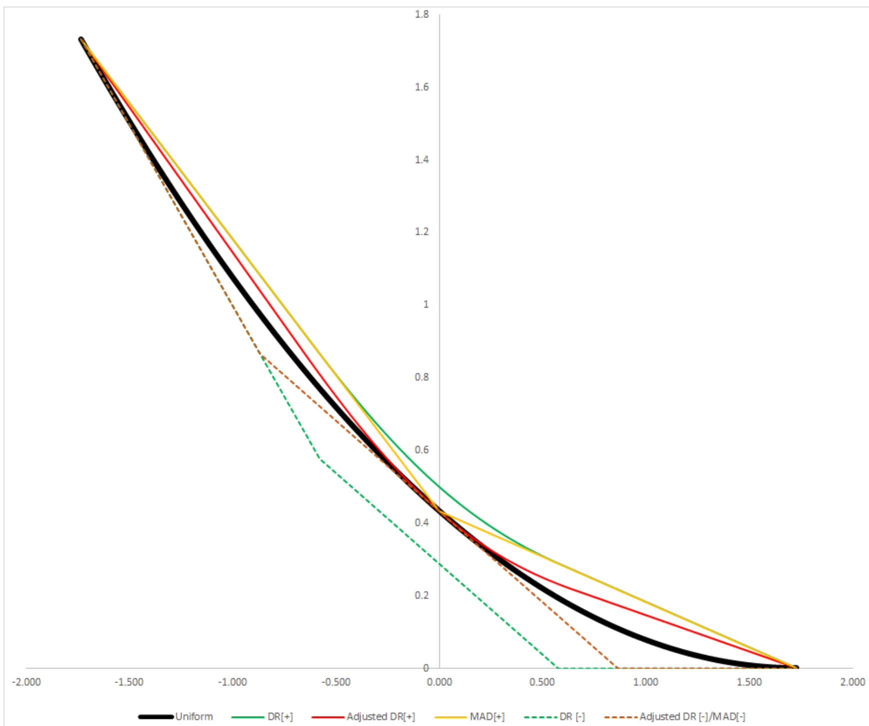


Figure 4: Upper and lower bounds for standard loss function of the uniform distribution.

When variance information is available, one may also formulate a standardized version of the MAD-based bounds above. Figure 4 illustrates these standardized bounds for the continuous uniform distribution on $[a, b]$ (see Appendix A.8 for the explicit definitions). For the uniform distribution, $\kappa = 0.75$ and $\tau = 1.5$, as $v = \frac{(b-a)^2}{12}$ and $d = \frac{b-a}{4}$. In the figure, *Uniform* denotes the standardized loss function for the uniform distribution ($L_{uniform}(z_Q)$), DR[+] denotes the original mean-variance upper bound $\left(L_{[+]}^{DR}(z_Q)\right)$, Adjusted DR[+] denotes the adjusted DR bound $\left(L_{[+]}^{\kappa*}(z_Q)\right)$ and MAD[+] denotes the mean-MAD bound $\left(L_{[+]}^{MAD}(z_Q)\right)$. This figure also illustrates lower bounds for the standard uniform loss function: DR[−] provides the original *mean-variance* lower bound $\left(L_{[-]}^{DR}(z_Q)\right)$, while Adjusted DR[−]/MAD[−] denotes the adjusted DR lower bound $\left(L_{[-]}^{\tau*}(z_Q)\right)$ and mean-MAD lower bound $\left(L_{[-]}^{MAD}(z_Q)\right)$, which are equivalent for symmetric distributions. As Figure 4 shows, when the decision maker has MAD, symmetry, and IFR information, the loss function can be closely approximated, especially in the neighborhood of the mean.

7 DR Newsvendor Problem

This section revisits the classical newsvendor problem when only mean and variance information are known, and a DR solution is preferred. Consider a firm that wishes to determine a procurement quantity Q at a unit cost of c prior to observing random demand X. Suppose that any unsatisfied demand incurs a unit cost of p. The order quantity that minimizes expected cost for a given distribution is characterized as $Q = F^{-1}(\rho)$, where F denotes the cumulative distribution function for X and ρ denotes the newsvendor critical fractile $\left(\rho = \frac{p-c}{p}\right)$.[4]

When the demand distribution (F) is not specified, but its expected value and variance, μ and σ^2, respectively, are known, this problem is referred to as the Distributionally Robust Newsvendor Problem (DRNP). Section 7.1 first formulates the DRNP, and then discusses the corresponding optimal order quantity under various assumptions on range of the underlying demand distribution. In Section 7.2, we will also provide the formulation and the solution method for the newsvendor problem with adjusted loss functions. Section 7.3 then considers the implications associated with using various DR approaches to bounding the expected loss function.

7.1 DRNP Model and Optimal Order Quantity

7.1.1 Worst-Case Model

We next provide a general formulation, which we refer to as ($\text{DRNP}_{[+]}$), where we utilize upper bounds on the loss function and assume that demand has a finite range $[a, b]$.

$$(\text{DRNP}_{[+]}) \quad \underset{a \le Q \le b}{\text{Minimize}} \ \underset{X \in \mathcal{D}_{[a,b]}(\mu, v)}{\text{Max}} pE\left[(X - Q)^+\right] + cQ. \tag{14}$$

[4] For both the minimization of expected cost under a known distribution and under the DR formulation below, an additional unit cost for leftover units at the end of the selling period can be accommodated via appropriate redefinition of the unit cost terms p and c, and with the subtraction of a constant from the objective function.

The worst-case expected loss term in the objective function can be formulated based on Lemma 2.1. We have the following optimal solution for the resulting problem.

$$Q = \begin{cases} a, & \text{if } \sqrt{\frac{\rho}{1-\rho}} < \frac{\sqrt{v}}{\mu - a}, \\ Q^* = \mu + \frac{1}{2}\left(\sqrt{\frac{\rho}{1-\rho}} - \sqrt{\frac{1-\rho}{\rho}}\right)\sqrt{v}, & \text{if } \frac{\sqrt{v}}{\mu-a} \leq \sqrt{\frac{\rho}{1-\rho}} \leq \frac{b-\mu}{\sqrt{v}}, \\ b, & \text{if } \sqrt{\frac{\rho}{1-\rho}} > \frac{b-\mu}{\sqrt{v}}. \end{cases} \tag{15}$$

When no information about the distribution is known beyond the mean and variance ($a = -\infty, b = \infty$), Scarf [19] derives the above expression on the middle interval for solving (14), i.e., $Q = Q^*$, while when demand is assumed to be nonnegative, the optimal solution under mean-variance ambiguity (based on Lemma 2.1, setting $a = 0, b = \infty$) becomes $Q = Q^*$ if $\frac{\sqrt{v}}{\mu} \leq \sqrt{\frac{\rho}{1-\rho}}$ and $Q = 0$ otherwise. This solution is also known as *Scarf's ordering rule.* Gallego and Moon [7] derived this solution for the case of nonnegative demand, although they did not explicitly employ the bound provided by the lemma. Rather, they suggested this condition based on the fact that an order of size zero leads to an expected profit equal to zero. They also considered various problem extensions, including a second order opportunity and fixed order costs.

7.2 Adjusted DRNP Model and Optimal Order Quantity

Under our adjusted DR approach with mixture ambiguity, we replace $\mathcal{D}_{[a,b]}(\mu, v)$ in the objective function of (14) with $\mathcal{S}^{[a,b]}(\mu, v) \subset \mathcal{D}_{[a,b]}(\mu, v)$, where $\mathcal{S}^{[a,b]}(\mu, v)$ denotes the set of all random variables with mean μ, variance v, and range $[a, b]$, which can be represented by an ambiguity set where the adjusted loss function bounds we have developed apply. For instance, this set might correspond to a mixture distribution containing some defined set of probability distributions.

7.2.1 Worst-case Model

To minimize the maximum expected cost, we substitute the adjusted worst-case expected loss function, $\ell^{\kappa}_{[+]}$, in the objective function of (14). It is straightforward to show that the following rule provides an optimal solution for this version of the DR newsvendor problem. (The appendix provides derivations of optimal solutions when using the DR and adjusted DR versions of the expected loss function in the newsvendor model.) In practical terms, solving the adjusted DRNP involves replacing v with κv in each of the foregoing equations for the optimal order quantity, i.e.,

$$Q = \begin{cases} a, & \text{if } \sqrt{\frac{\rho}{1-\rho}} < \frac{\sqrt{\kappa v}}{\mu - a}, \\ \mu + \frac{1}{2}\left(\sqrt{\frac{\rho}{1-\rho}} - \sqrt{\frac{1-\rho}{\rho}}\right)\sqrt{\kappa v}, & \text{if } \frac{\sqrt{\kappa v}}{\mu - a} \leq \sqrt{\frac{\rho}{1-\rho}} \leq \frac{b-\mu}{\sqrt{\kappa v}}, \\ b, & \text{if } \sqrt{\frac{\rho}{1-\rho}} > \frac{b-\mu}{\sqrt{\kappa v}}. \end{cases} \tag{16}$$

7.3 Properties of Adjusted DRNP Solution

In this section, we discuss implications associated with using different forms of DR loss function approximations for the newsvendor problem. For the sake of simplicity and following the operations literature, we assume that we solve a newsvendor problem with nonnegative demand. This is a widely adopted approach for most operations models. We therefore study the solution in which $a = 0$ and $b = \infty$, where $Q = \mu + \frac{1}{2}\left(\sqrt{\frac{\rho}{1-\rho}} - \sqrt{\frac{1-\rho}{\rho}}\right)\sqrt{\kappa v}$ when $\frac{\sqrt{\kappa v}}{\mu} \leq \sqrt{\frac{\rho}{1-\rho}}$, and $Q = 0$ otherwise.

Observe that the only difference between Scarf's DR ordering rule in Equation (15), and the ordering rule when the adjusted DR approach, Equation (16), involves replacing v with κv in (15), which is equivalent to reducing the absolute value of the associated safety stock by a factor of $(1-\sqrt{\kappa}) \times 100\%$ relative to Scarf's DR ordering rule (when $\rho > 0.5$, this corresponds to a reduction in the positive value of safety stock, or a reduction in total stock; when $\rho < 0.5$, it corresponds to a reduction in the amount of negative safety stock, or an increase in total stock). Recall that the smallest valid value of κ we observed among the distributions we considered in the mixture was $\kappa = 0.75$, which corresponded to the bound obtained for the uniform distribution. When $\kappa = 0.75$, this corresponds to a 13.4% reduction in the safety stock level prescribed by Scarf's ordering rule. When the uniform distribution can be excluded, we arrived at a value of $\kappa = 0.663$, which corresponds to an 18.58% reduction. Note that this percentage change in safety stock corresponds to an equivalent percentage change in the corresponding value of z_Q.

When the *cv* is relatively low and the critical fractile ρ is neither too high nor too low, the DR, adjusted DR, and normal distribution models suggest order quantities that are quite close to one another, as the corresponding amount of safety stock in such cases is relative small. Figure 5 shows optimal z_Q values under the normal, DR, and adjusted DR models. As the figure illustrates, the optimal z_Q values are extremely close to one another for each of the approaches when $\rho \in [0.2, 0.8]$.

Figure 6 indicates that when the optimal service level reaches a sufficiently high value ($\rho > 0.9$), Scarf's DR approach begins to suggest increasingly higher order quantities relative to the adjusted DR approaches. In practice, under either the DR or adjusted DR approach, the decision maker

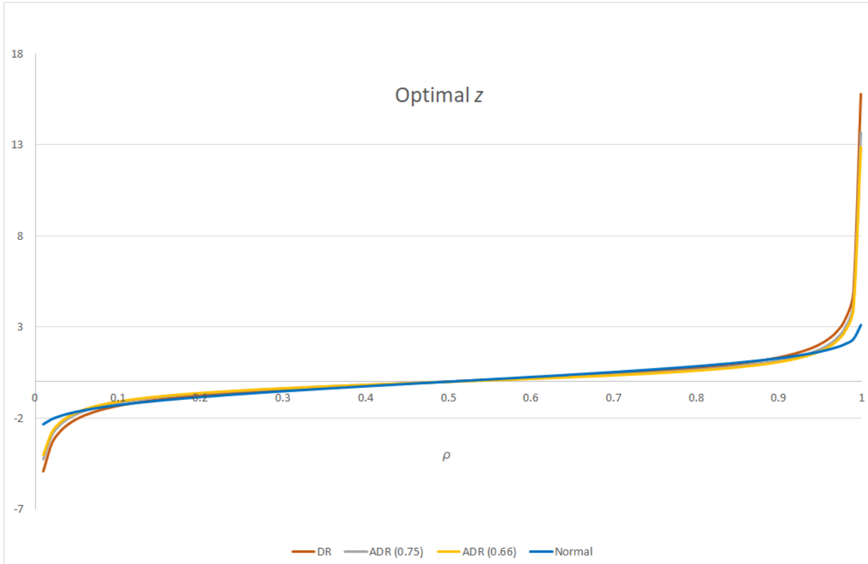


Figure 5: Optimal z versus ρ.

may choose to employ an artificial upper bound, b, on the demand distribution, in order to avoid unreasonably high order quantities that may be prescribed by the models.

When demand has a relatively high cv and/or the optimal service level is relatively low (to be more specific, when $\sqrt{\frac{\rho}{1-\rho}} < \frac{\sqrt{v}}{\mu} = -\frac{1}{z_0}$), the DR ordering rule may choose to order nothing, allowing demand to go unsatisfied, which may result in an overly conservative approach. Utilizing the adjusted DR approach decreases the possibility of such solutions that order nothing by adjusting this condition to $\sqrt{\frac{\rho}{1-\rho}} < \frac{\sqrt{\kappa v}}{\mu}$. Figure 7 illustrates the maximum possible cv value for a given value of ρ that leads to a positive order quantity at optimality. As the figure illustrates, the DR approach is more conservative, leading to a zero order quantity under lower levels of relative uncertainty than the adjusted DR approaches. Similarly, when the implied service level is relatively high, the DR approach sets the order quantity to the upper limit of b when $\sqrt{\frac{\rho}{1-\rho}} > z_b$. The adjusted DR approach reduces the likelihood of such cases by adjusting this condition to $\sqrt{\frac{\rho}{1-\rho}} > \frac{z_b}{\sqrt{\kappa}}$.

8 Computational Tests

This section summarizes the results of computational tests that characterize the performance of the adjusted DR bound under various underlying distribution assumptions. All tests were implemented on an Intel® Core™ i5-8250U CPU, 1.80 GHz processor with 8 GB RAM.

Within our mixture of distributions for computational testing purposes, we considered the normal, gamma, lognormal, and Pareto distributions, as well as a mixture of these distributions

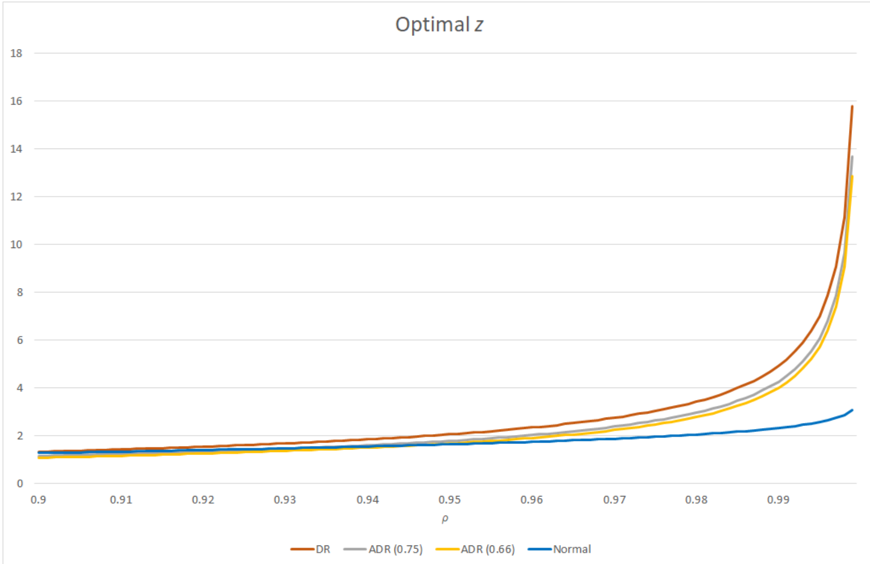


Figure 6: Optimal z versus ρ when $\rho > 0.9$.

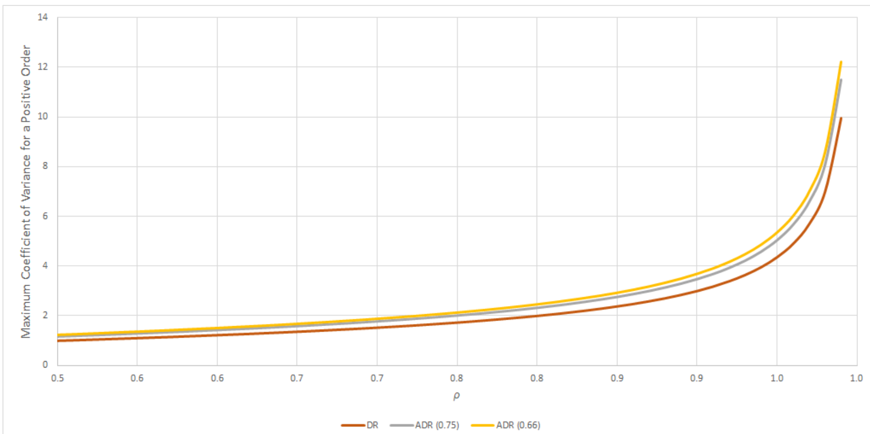


Figure 7: Positive Order Quantity Threshold for $\frac{\sqrt{v}}{\mu}$ given ρ.

with equal weights. We set the standard deviation of demand at 30 and considered four levels of cv equal to 0.3, 1, and 2, by varying the value of the expected demand. We assumed a unit purchase cost of $c = 1$, and varied the range of overflow penalty values from 2 to 100. This implies that we therefore tested newsvendor solutions with optimal service levels varying between 50% and 99%.

With the service level (SL) range between 50% and 99%, we solved 27 instances for each random distribution (including the mixture of distributions). For each instance, we first solved the problem assuming the known demand distribution. We then solved the same instance utilizing the DR bound on the expected loss function, the adjusted DR bound with $\kappa = 0.75$ (ADR(0.75)), the adjusted DR bound with $\kappa = 0.663$ (ADR(0.66)), and a normal approximation approach (i.e., using a standard normal loss function). We assume that demand is unbounded and utilized the order quantity

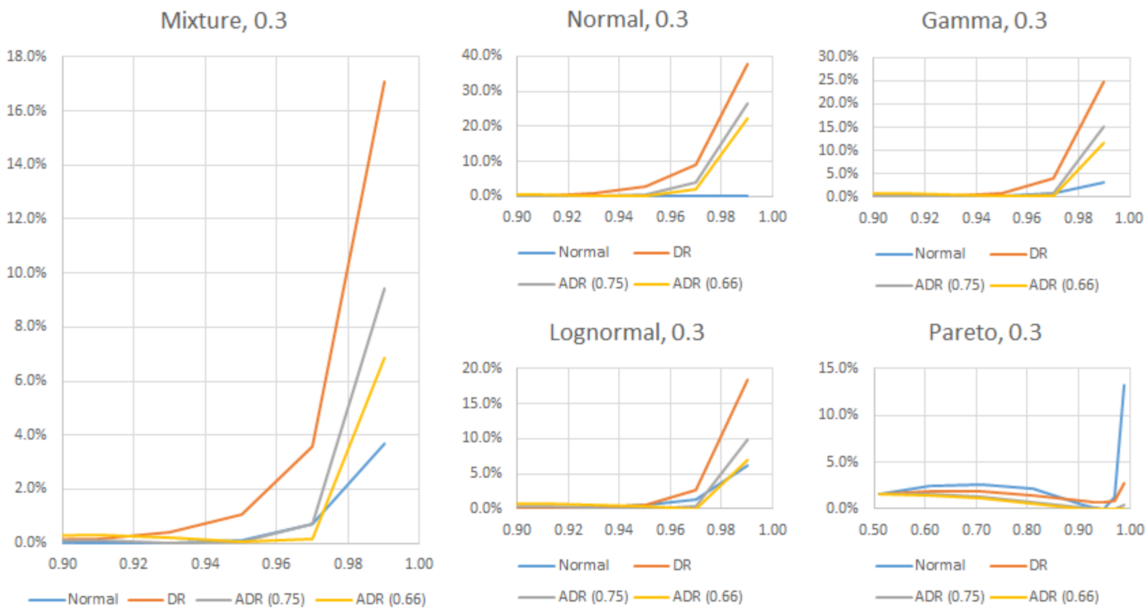
$$Q = \mu + \frac{1}{2}\left(\sqrt{\frac{\rho}{1-\rho}} - \sqrt{\frac{1-\rho}{\rho}}\right)\sqrt{\kappa v},$$

with $\kappa = 1$, 0.75, and 0.663 for the DR, ADR(0.75) and ADR(0.66) cases, respectively.

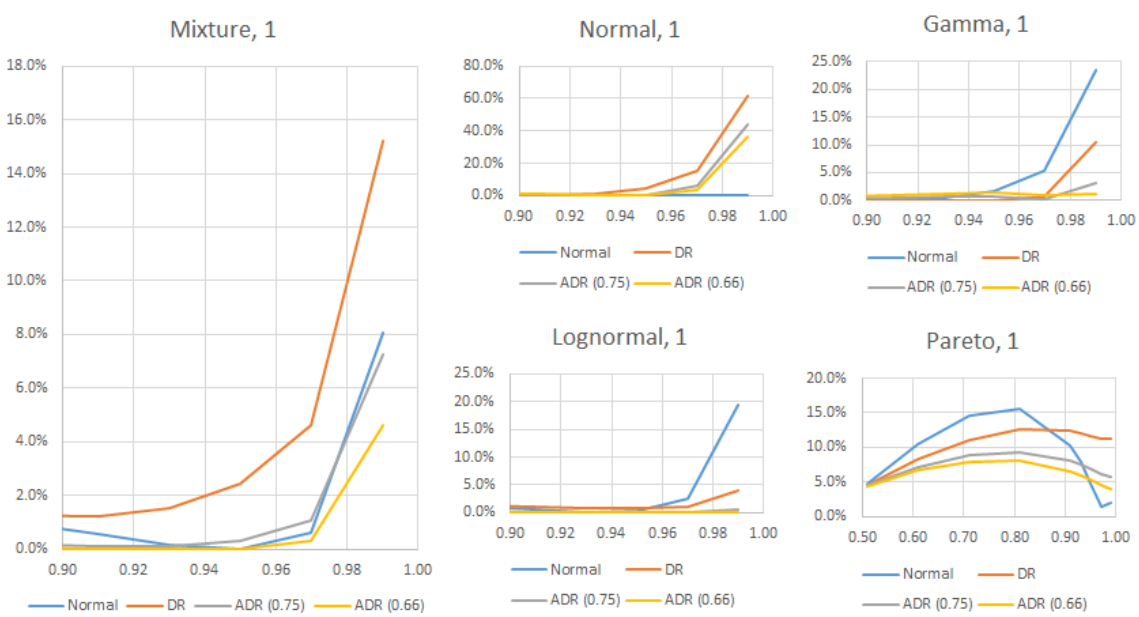
Tables 8, 9, and 10 in the Appendix summarize the optimality gaps for these instances. Specifically, for any instance, we let Π^* denote the optimal expected newsvendor cost for the given (known) underlying distribution, and let Π_j denote the expected cost (assuming the underlying distribution) of the newsvendor quantity given by approach j, where j is either DR, ADR(0.75), ADR(0.66), or normal. Entries in this table indicate values of the gap $\frac{\Pi_j - \Pi^*}{\Pi^*} \times 100\%$.

Figure 8 illustrates the performance of each solution approach for cv values of 0.3 and 1, and service levels (ρ) higher than 90%. Our results indicate that solving the newsvendor model assuming a normal distribution is an acceptable approach for the ambiguity set under consideration when the cv is low (for instance, less than 0.3) and the service level is not too high. This result is in line with past literature. More specifically, a normal approximation performs better when the underlying distribution is gamma. Note that a smaller cv for the gamma distribution implies a larger shape parameter, and as the shape parameter increases, the gamma distribution tends toward the normal distribution. Similarly, for the lognormal distribution, it is known that the error associated with using a normal approximation decreases as variance decreases. On the other hand, the performance of the DR approaches (particularly in the case of ADR(0.66)) provides the highest quality solutions relative to a normal distribution when the underlying distribution is Pareto.

When the cv and/or optimal service level increase, the DR approaches improve substantially. For instance, when the cv is 1 or more, the DR approaches perform well when the weight of the

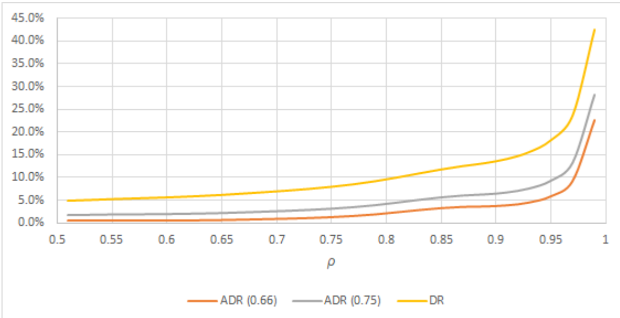


(a) $\frac{\sqrt{v}}{\mu} = 0.3$

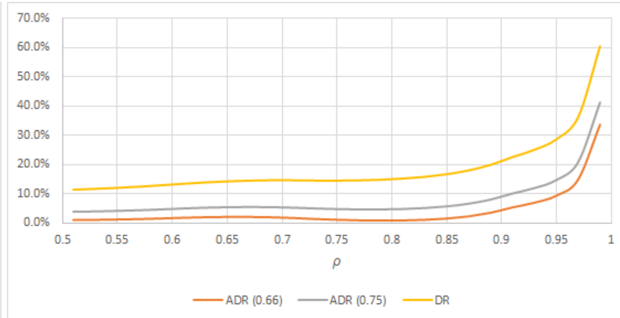


(b) $\frac{\sqrt{v}}{\mu} = 1$

Figure 8: Newsvendor optimality gaps versus ρ for specific distributions and DR approaches.



(a) $\frac{\sqrt{v}}{\mu} = 0.3$



(b) $\frac{\sqrt{v}}{\mu} = 1$

Figure 9: Optimality gaps for DR newsvendor problem solutions versus ρ with $\mathcal{S}(\mu, v) \subset \mathcal{D}_{[0,\infty]}(\mu, v)$.

normal distribution is relatively small within the mixture. According to our results, the adjusted DR approach performs considerably more favorably than the DR approach in such settings. The reader may observe this result in Figure 8(b) and Tables 9 and 10. This is mainly due to the fact that our adjusted loss functions more closely approximate the expected loss function when compared to a normal loss function or the traditional DR loss function for the distributions in the mixture (except for the cases, of course, in which the underlying distribution is actually normal).

We also consider the relative performance of DR approaches for solving the DRNP with mixture ambiguity (problem (14) with $X \in \mathcal{S}^{[a,b]}(\mu, v) \subset \mathcal{D}_{[0,\infty]}(\mu, v)$). Given the components of the mixture distribution, the worst-case expected loss term in the objective function can be written as $E\left[(X-Q)^+\right] = L_{max}(z_Q)\sqrt{v}$, where $L_{max}(z_Q)$ is defined by the upper envelope of the set of loss functions $L_i(z_Q)$, $i =$ *normal, gamma, lognormal* and *Pareto*. We let $\Pi^{mixture}$ denote the optimal objective function value of DRNP when using $L_{max}(z_Q)\sqrt{v}$. We also let Π^j denote the optimal objective function of DRNP model (14) with approach j, where j corresponds to DR, ADR(0.75) and ADR(0.66). Figure 9 and the entries in Table 11 in the Appendix illustrate values of the gap $\frac{\Pi^j - \Pi^{mixture}}{\Pi^{mixture}} \times 100\%$. As both cv and service level values increase, the gap between the classical DR approach (developed for mean-variance ambiguity) and the adjusted DR approaches we propose increases. The figures illustrate that a substantial premium may be incurred due to the overly conservative nature of the classical DR approach. If, for example, in addition to the mean and variance being known, we observe that the distribution has an approximately lognormal or gamma distributed shape, the adjusted DR methods can provide highly robust solutions that do not rely on the high cost of overly conservative bounds that are valid for all possible distributions.

On the other hand, our adjusted DR approaches (especially ADR(0.66)) provide a much closer estimate of the worst-case cost under mixture ambiguity compared to the classical DR approach. Although the bounds provided by the adjusted DR approaches explicitly consider the set of distributions in the mixture, the set of underlying distributions is quite general and includes well-known maximum entropy distributions as mentioned earlier. Moreover, this approach can be extended to include additional distributions that are amenable to standardization of the loss functions.

In certain practical settings, we may not have the knowledge of the exact mixture of distributions, but may still estimate the MAD (d) and distribution shape characteristics based on (limited) observations. The parameters required for the adjusted DR approach can then be estimated. In our numerical tests, we sampled 40 data points from the mixture distribution (with $\delta = 1$). Based on these samples, we then estimated $\frac{v^2}{d} = 0.63$ and $\frac{d(b-a)}{2v} = 1.64$. This leads to permissible intervals for κ and τ of $0.63 \leq \kappa \leq 1$ and $1 \leq \tau \leq 1.64$. Alternatively, one may also solve the optimization problems given in Equation 9 and 11 to determine κ and τ for the sample under consideration (see Appendix A.9 for further explanation of this approach). Given 40 observations, this approach resulted in $\kappa^* = 0.69$ and $\tau^* = 1.48$. Note that adopting an adjusted DR approach with these parameters would result in a considerable improvement over the traditional DR approach under our mixture distribution assumptions. Of course, the availability of additional data points would improve the accuracy of such an estimation approach. However, in such cases, the need for a DR optimization approach decreases as we may more directly and accurately estimate the loss function or fit a corresponding distribution.

9 Conclusion

The literature on the newsvendor problem largely focuses on cases in which the demand distribution is known (and often when it is assumed to follow a normal distribution). When the decision maker can only estimate a few low-order moments, but is not able to characterize the 'true' distribution, a DR approach may be adopted. Existing DR models in the literature based on moment ambiguity sets have been characterized as overly-conservative by various researchers. In this study, we suggest an alternative and practical DR approach wherein the ambiguity set is defined using a mixture distribution. We mitigate the degree of conservatism by suggesting stricter upper bounds on the loss function that apply to a broad class of probability distributions. We have shown that our suggested DR approach performs favorably compared to the traditional bound suggested by Scarf

[19] for a wide range of settings, and especially those in which the random variable originates from a right skewed distribution with a high *cv*. For such settings, a normal distribution approximation often leads to unfavorable results.

We also showed that when the distribution is known to be symmetric and IFR, our adjusted loss functions can be specified based on only mean-MAD information. Such bounds are tight at the mean, and stricter compared to mean-MAD bounds available in the literature. Furthermore, due to the continuous portion of the worst-case distribution, the optimal newsvendor solution is no longer based on a distribution with a few discrete points.

Furthermore, we introduced a standard loss function concept for various distributions as well as for application in DR settings. For a wide range of commonly applicable demand distributions, we were able to provide closed-form expressions for the standard loss functions using only the *cv* (independently from specific distribution parameters). Future research may formalize this result and may consider a wider range of continuous random distributions to include within the mixture set. Moreover, another interesting research direction lies in providing bounds that apply to the set of all continuous random distributions.

In addition to the classical newsvendor problem, this research may have valuable implications for numerous complex planning problems involving expected loss terms, where limited distribution information is available. When solutions are sought that account for unlikely but highly negative outcomes, the suggested adjusted DR approaches can be adopted in many practical settings.

References

[1] A. Ben-Tal and E. Hochman. More bounds on the expectation of a convex function of a random variable. *Journal of Applied Probability*, 9(4):803–812, 1972.

[2] R. Bhatia and C. Davis. A better bound on the variance. *The american mathematical monthly*, 107(4):353–357, 2000.

[3] M.C. Bryson. Heavy-tailed distributions: properties and tests. *Technometrics*, 16(1):61–68, 1974.

[4] S. Choi and W.E. Wilhelm. An analysis of sequencing surgeries with durations that follow the lognormal, gamma, or normal distribution. *IIE Transactions on Healthcare Systems Engineering*, 2(2):156–171, 2012.

[5] B. Das, A. Dhara, and K. Natarajan. On the heavy-tail behavior of the distributionally robust newsvendor. *Operations Research*, 69(4):1077–1099, 2021.

[6] E. de Klerk, D. Kuhn, and K. Postek. Distributionally robust optimization with polynomial densities: theory, models and algorithms. *Mathematical Programming*, 181:65–296, 2020.

[7] G. Gallego and I. Moon. The distribution free newsboy problem: Review and extensions. *Journal of the Operational Research Society*, 44(8):825–834, 1993.

[8] G.A. Hanasusanto, S.W. Wallace, and S. Zymler. Distributionally robust multi-item newsvendor problems with multimodal demand distributions. *Mathematical Programming*, 152(1):1–32, 2015.

[9] S. Jackman. *Bayesian analysis for the social sciences*, volume 846. John Wiley & Sons, 2009.

[10] J. Kamburowski. The distribution-free newsboy problem under the worst-case and best-case scenarios. *European Journal of Operational Research*, 237(1):106–112, 2014.

[11] S. Karlin and W. J. Studden. *Tchebycheff systems: With applications in analysis and statistics*, volume 15. Interscience Publishers, 1966.

[12] J.H.C. Lisman and M.C.A. Van Zuylen. Note on the generation of most probable frequency distributions. *Statistica Neerlandica*, 26(1):19–23, 1972.

[13] J.H. May, D.P. Strum, and L.G. Vargas. Fitting the lognormal distribution to surgical procedure times. *Decision Sciences*, 31(1):129–148, 2000.

[14] Y. Merzifonluoğlu, J. Geunes, and H.E. Romeijn. The static stochastic knapsack problem with normally distributed item sizes. *Mathematical Programming*, 134(2):459–489, 2012.

[15] A. Müller and D. Stoyan. *Comparison methods for stochastic models and risks*, volume 389. Wiley, 2002.

[16] S.Y. Park and A.K. Bera. Maximum entropy autoregressive conditional heteroskedasticity model. *Journal of Econometrics*, 150(2):219–230, 2009.

[17] K. Postek, A. Ben-Tal, D. den Hertog, and B. Melenberg. Robust optimization with ambiguous stochastic constraints under mean and dispersion information. *Operations Research*, 66(3):814–833, 2018.

[18] H. Rinne. *The hazard rate: Theory and inference.* Justus-Liebig-Universität Giessen: Giessen, Germany, 2014.

[19] H. Scarf. A min-max solution to an inventory problem. In *Studies in Mathematical Theory of Inventory and Production.* Stanford University Press, 1958.

[20] P.S. Stepaniak, C. Heij, G.H.H. Mannaerts, M. de Quelerij, and G. de Vries. Modeling procedure and surgical times for current procedural terminology-anesthesia-surgeon combinations and evaluation in terms of case-duration prediction and operating room efficiency: A multicenter study. *Anesthesia Analgesia*, 109(4):1232–1245, 2009.

[21] D.P. Strum, J.H. May, A.R. Sampson, L.G. Vargas, and W.E. Spangler. Estimating times of surgeries with two component procedures: Comparison of the lognormal and normal models. *The Journal of the American Society of Anesthesiologists*, 98(1):232–240, 2003.

[22] D.P. Strum, J.H. May, and L.G. Vargas. Modeling the uncertainty of surgical procedure times: Comparison of log-normal and normal models. *The Journal of the American Society of Anesthesiologists*, 92(4):1160–1167, 2000.

[23] D.P. Strum, A.R. Sampson, J.H. May, and L.G. Vargas. Surgeon and type of anesthesia predict variability in surgical procedure times. *The Journal of the American Society of Anesthesiologists*, 92(5):1454–1466, 2000.

[24] K. Taaffe, J. Geunes, and H.E. Romeijn. Target market selection and marketing effort under uncertainty: The selective newsvendor. *European Journal of Operational Research*, 189(3):987–1003, 2008.

[25] W. van Eekelen, D. den Hertog, and J.S.H. van Leeuwaarden. Mad dispersion measure makes extremal queue analysis simple. *INFORMS Journal on Computing*, Forthcoming, `https://doi.org/10.1287/ijoc.2021.1130`.

A Appendix

A.1 Characterization of structure of $L^{\kappa}_{[+]}(z_Q)$

In this section, we analyze the properties of the following function:

$$L^{\kappa}_{[+]}(z_Q) = E(Z^{\kappa}_{[+]} - z_Q)^+ = \begin{cases} -\frac{\frac{\kappa}{z_a}+z_Q}{\kappa\left(\frac{1}{z_a}\right)^2+1}, & z_a \leq z_Q \leq \frac{1}{2}\left(z_a - \frac{\kappa}{z_a}\right), \\ L^{\kappa}_{+}(z_Q), & \frac{1}{2}\left(z_a - \frac{\kappa}{z_a}\right) < z_Q \leq \frac{1}{2}\left(z_b - \frac{\kappa}{z_b}\right), \\ \frac{\kappa(z_b - z_Q)}{\kappa + {z_b}^2}, & \frac{1}{2}\left(z_b - \frac{\kappa}{z_b}\right) < z_Q \leq z_b. \end{cases} \tag{A.1}$$

When $z_Q = z_a$, then $L^{\kappa}_{[+]}(z_Q) = -z_a$. At this point $L^{\kappa}_{[+]}(z_Q) = L^{DR}_{[+]}(z_a) = L_i(z_a) = -z_a$ for any distribution i bounded below by a. Similarly, when $z_Q = z_b$,then $L^{\kappa}_{[+]}(z_Q) = L^{DR}_{[+]}(z_b) = L_i(z_b) = 0$ for any distribution i bounded above by b.

We next show that this function is continuous. Observe that at $z_Q = \frac{1}{2}\left(z_a - \frac{\kappa}{z_a}\right)$,

$$L^{\kappa}_{+}(z_Q) = \frac{1}{2}\left(\sqrt{\kappa + \left(\frac{1}{2}\left(z_a - \frac{\kappa}{z_a}\right)\right)^2} - \left(\frac{1}{2}\left(z_a - \frac{\kappa}{z_a}\right)\right)\right) = -\frac{z_a}{2},$$

where equality holds because $\sqrt{\kappa + \left(\frac{1}{2}\left(z_a - \frac{\kappa}{z_a}\right)\right)^2} = -\left(\frac{1}{2}\left(z_a + \frac{\kappa}{z_a}\right)\right)$.

At $z_Q = \frac{1}{2}\left(z_a - \frac{\kappa}{z_a}\right)$,

$$L^{\kappa}_{[+]}\left(\frac{1}{2}\left(z_a - \frac{\kappa}{z_a}\right)\right) = -\frac{\frac{\kappa}{z_a} + \left(\frac{1}{2}\left(z_a - \frac{\kappa}{z_a}\right)\right)}{\kappa\left(\frac{1}{z_a}\right)^2 + 1} = -\frac{z_a}{2}.$$

Similarly, when $z_Q = \frac{1}{2}\left(z_b - \frac{\kappa}{z_b}\right)$,

$$L^{\kappa}_{+}\left(\frac{1}{2}\left(z_b - \frac{\kappa}{z_b}\right)\right) = \frac{1}{2}\left(\sqrt{\kappa + \left(\frac{1}{2}\left(z_b - \frac{\kappa}{z_b}\right)\right)^2} - \left(\frac{1}{2}\left(z_b - \frac{\kappa}{z_b}\right)\right)\right) = \frac{\kappa}{2z_b},$$

while

$$\frac{\kappa\left(z_b - \left(\frac{1}{2}\left(z_b - \frac{\kappa}{z_b}\right)\right)\right)}{\kappa + {z_b}^2} = \frac{\kappa}{2z_b}.$$

This concludes the proof.

A.2 Proof of Lemma 4.1

Because $L_i(z) = \tilde{L}_i(z) - z$, the objective, $g_i(z_Q, \delta) = 4L_i(z)(z + L_i(z))$, is equivalent to $g_i(z_Q, \delta) = 4L_i(z)\tilde{L}_i(z)$. It is well-known that both $L_i(z)$ and $\tilde{L}_i(z)$, which correspond to the expected overstock and expected understock quantities, are nonnegative for all $z \in \mathbb{R}$, implying that $g_i(z_Q, \delta) \geq 0$ for all $z \in \mathbb{R}$. In addition, we can write $\frac{dg_i(z_Q,\delta)}{dz_Q} = 4L_i(z_Q)\Phi_i(z_Q) - 4\tilde{L}_i(z_Q)(1 - \Phi_i(z_Q))$. For a symmetric distribution, we have $\Phi_i(z_Q) = 0.5$ when $z_Q = 0$, and at this point the first-order derivative can then be written as $\frac{dg_i(z_Q,\delta)}{dz_Q} = 2L_i(0) - 2\tilde{L}_i(0)$. Because $L_i(0) = \tilde{L}_i(0)$, a stationary point exists for $g_i(z_Q, \delta)$ at $z_Q = 0$ for a symmetric distribution.

A.3 Proof of Lemma 4.2

We can write $g_i(z_Q, \delta)$ as $g_i(z)$, for convenience. For distributions such that $L_i(z)$ is independent of δ, we can write $g_i(z) = 4L_i(z)\tilde{L}_i(z)$, which has a derivative of $g_i'(z) = 4(L_i(z)\Phi_i(z) - (1 - \Phi_i(z))\tilde{L}_i(z))$. As a result, the stationary point condition for $g_i(z)$ is equivalent to $L_i(z)\Phi_i(z) = (1 - \Phi_i(z))\tilde{L}_i(z)$. This condition is equivalent to $\frac{E\left[(u-z)^+\right]}{1-\Phi_i(z)} = \frac{E\left[(z-u)^+\right]}{\Phi_i(z)}$, i.e., the conditional expected shortage equals the conditional expected leftover. This necessary stationary point condition for a maximizing solution can be written as $R_1(z) = R_1(-z)$ for a symmetric distribution, where $R_1(z) = \frac{E\left[(u-z)^+\right]}{1-\Phi_i(z)}$ is equivalent to the mean residual life (MRL) function in reliability analysis, and is continuously defined for all $z \in [z_a, z_b]$ (with $0 \in (z_a, z_b)$).

The stationary point condition $R_1(z) = R_1(-z)$ requires the MRL function to equal its mirror image about the vertical axis, which necessarily occurs at $z = 0$. It is straightforward to show that $R_1'(z) = R_1(z)\theta(z)$–1, where $\theta(z) = \phi(z)/(1 - \Phi(z))$ corresponds to the failure rate (FR) function of a distribution. It is well known that an increasing failure rate (IFR) distribution implies a decreasing mean residual life (DMRL) function, where the terms increasing and decreasing are typically used in the weak (monotonic) sense in the reliability literature ([18]). Because the MRL function is nonincreasing, it can only coincide with its mirror image about the vertical axis on a single contiguous interval that contains $z = 0$ (this interval may have zero width, thus containing only $z = 0$). Thus, the stationary point condition is satisfied only on a single contiguous interval containing 0.

Because the objective function $(4L(z)\tilde{L}(z))$ equals 0 at both z_a and z_b, and is positive for all $z \in (z_a, z_b)$ (and $0 \in (z_a, z_b)$), each point on the interval at which the stationary point condition is satisfied must correspond to a maximizing point of the objective function (if this is not the case, then some other stationary point solution must exist outside this interval, which cannot occur as the intersection of a nonincreasing function and its mirror image about the vertical axis). Because $z = 0$ is contained within this interval, the objective function is maximized at $z = 0$ for a symmetric distribution that is IFR.

A.4 Proof of MAD Lemmas

We showed that an optimal stationary point solution exists for $g_i(z_Q, \delta)$ at $z_Q = 0$ for a symmetric IFR distribution. We next show that $\frac{d^2}{v} = 4L_i(0)\tilde{L}_i(0)$.

Proof of Lemma 4.3

$$\begin{aligned} d &= \int_{-\infty}^{\infty} |x-\mu| f(x)dx \\ &= \int_{-\infty}^{\mu} (\mu - x)f(x)dx + \int_{\mu}^{\infty} (x-\mu)f(x)dx \\ &= \int_{-\infty}^{\infty} (\mu - x)f(x)dx - \int_{\mu}^{\infty} (\mu - x)f(x)dx + \int_{\mu}^{\infty} (x-\mu)f(x)dx \\ &= \int_{-\infty}^{\infty} (\mu - x)f(x)dx + 2\int_{\mu}^{\infty} (x-\mu)f(x)dx \\ &= 2\int_{\mu}^{\infty} (x-\mu)f(x)dx. \end{aligned}$$

Therefore $\frac{d}{2} = \int_{-\infty}^{\mu}(\mu - x)f(x)dx = \int_{\mu}^{\infty}(x-\mu)f(x)dx$. Because $\kappa = 4L_i(0)\tilde{L}_i(0)$,

$$\begin{aligned} \kappa v &= 4\left(L_i(0)\sqrt{v}\right)\left(\tilde{L}_i(0)\sqrt{v}\right) \\ &= 4E\left[(X-\mu)^+\right]E\left[(\mu - X)^+\right] \\ &= 4\int_{\mu}^{\infty}(x-\mu)f(x)dx \cdot \int_{-\infty}^{\mu}(\mu - x)f(x)d(x) \\ &= \left(2\int_{\mu}^{\infty}(x-\mu)f(x)dx\right)^2 \\ &= d^2. \end{aligned}$$

Proof of Lemma 4.4 The stationary point condition for $g_i(z)$ is defined as $L_i(z)\Phi_i(z) = (1 - \Phi_i(z))\tilde{L}_i(z)$ in Lemma 4.1. We now consider an asymmetric distribution. For such distributions $L_i(0) = \tilde{L}_i(0)$ but $\Phi_i(0) \neq 0.5$. Therefore, $z = 0$ is not a stationary point, indicating that there exists a z^* value where $g_i(z^*) > g_i(0) = 4L_i(0)\tilde{L}_i(0) = \frac{d^2}{v}$. Therefore, a valid value of κ for the adjusted DR upper bound requires $\kappa > \frac{d^2}{v}$. We know from Scarf [19] that $\kappa = 1$ provides a valid bound on any distribution, implying that a value of κ such that $\frac{d^2}{v} < \kappa \leq 1$ exists such that the adjusted DR bound is valid.

Proof of Lemma 4.5 The value of τ must satisfy the relationship $\tau \leq t_i(z_Q) = L_i(z_Q)(z_b - z_a) - z_Q z_a = L_i(z_Q)(z_b) - \tilde{L}(z_Q)(z_b)$. Note that $t_i(z_Q)$ is convex, as the second derivative is positive $(\phi_i(z_Q)(z_b - z_a) > 0)$. This function is thus minimized when the first derivative is equal to 0, i.e., $\Phi_i(z^*) = \frac{z_b}{z_b - z_a}$ or $z^* = \Phi_i{}^{-1}\left(\frac{z_b}{z_b - z_a}\right)$. This implies $\tau = L_i\left(\Phi_i{}^{-1}\left(\frac{z_b}{z_b-z_a}\right)\right) z_b - \tilde{L}\left(\Phi_i{}^{-1}\left(\frac{z_b}{z_b-z_a}\right)\right) z_a$. For symmetric distributions, we have $z_a = -z_b$ and $z_Q = 0$ is optimal. This results in $\tau = 2z_b L_i(0) = \sqrt{\kappa} z_b$.

Relationship between τ and κ: For symmetric IFR distributions, we have $\kappa = \frac{d^2}{v}$ and $\tau = \frac{d(b-a)}{2v}$ (see Lemmas 4.3 and 4.5). Note that $\tau = \frac{d}{\sqrt{v}} \frac{\omega}{\sqrt{v}} = \sqrt{\kappa} \frac{\omega}{\sqrt{v}}$ where $\omega = b - a$. For any distribution, we have $v \leq (\mu - a)(b - \mu)$ (see [2]). For symmetric distributions, we then have $v \leq \frac{\omega^2}{4}$, implying $\frac{\omega}{\sqrt{v}} \geq 2$. Combining these results, we have $\tau \geq 2\sqrt{\kappa}$.

Proof of Lemma 4.6 It is known that $1 \leq \frac{d(b-a)}{2v}$ for any distribution [17]. As we are minimizing $t_i(z_Q)$, we also have $\tau = t_i(z^*) \leq t_i(0) = L_i(0)(z_b - z_a) = \frac{d}{2\sqrt{v}} \frac{b-a}{\sqrt{v}} = \frac{d(b-a)}{2v}$. For the DR setting we have $\tau = 1$, and therefore, for the adjusted bound setting $\tau \geq 1$.

A.5 Upper Bound given by Mean-MAD Ambiguity Set

We let $\mathcal{D}_{[a,b]}(\mu, d)$ denote the set of all random variables with finite lower and upper bounds $[a, b]$ (with $a < b$), expected value μ, and mean absolute deviation d where $d = E\left[|D - \mu|\right]$.

Lemma A.1 (Upper Bound, [1]) *For any random variable X with mean μ and mean absolute deviation d defined on the interval $[a, b]$, with $-\infty < a < \mu < b < \infty$ and any convex function $f(x)$* $Max_{X \in \mathcal{D}_{[a,b]}(\mu,d)} E\left[f(x)\right] = p_1 f(a) + p_2 f(\mu) + p_3 f(b),$
where
$p_1 = \frac{d}{2(\mu-a)}$; $p_2 = 1 - \frac{d}{2(\mu-a)} - \frac{d}{2(b-\mu)}$ *and* $p_3 = \frac{d}{2(b-\mu)}$.

Upper Bound on the Loss Function

First note that $(X - Q)^+$ is a convex function in X. Therefore, for any $a \leq Q \leq b$, we have the following upper bound for the loss function, i.e., $E\left[(X - Q)^+\right]$.

$$Max_{X \in \mathcal{D}_{[a,b]}(\mu,d)} E\left[(X - Q)^+\right] = \begin{cases} (\mu - Q)\left(1 - \frac{d}{2(\mu-a)} - \frac{d}{2(b-\mu)}\right) + (b - Q)\frac{d}{2(b-\mu)}, & a \leq Q \leq \mu, \\ (b - Q)\frac{d}{2(b-\mu)}, & \mu \leq Q \leq b, \end{cases}$$

Symmetric Distributions

Letting $\omega = 2(b-\mu) = 2(\mu-a) = b-a$, the worst-case distribution and the corresponding upper bound for $E\left[(X-Q)^+\right]$ for this *mean-MAD* ambiguity set are given below, respectively.

$$X_{[+]}^{MAD} = \begin{cases} a, & w.p.\ \frac{d}{\omega}, \\ \mu, & w.p.\ 1-\frac{2d}{\omega}, \\ b, & w.p.\ \frac{d}{\omega}. \end{cases}$$

$$\ell_{[+]}^{MAD}(Q) = \begin{cases} (\mu-Q)\left(1-\frac{2d}{\omega}\right) + (b-Q)\frac{d}{\omega}, & a \le Q \le \mu, \\ (b-Q)\frac{d}{\omega}, & \mu \le Q \le b. \end{cases}$$

A.6 Lower Bound given by Mean-MAD Ambiguity Set

We define $\mathcal{D}_{[a,b]}(\mu, d, \beta)$ as the set of all random variables with finite lower and upper bounds $[a, b]$ (with $a < b$), expected value μ, mean absolute deviation d where $d = E\left[|D-\mu|\right]$ and $\beta = P(D \ge \mu)$.

Lemma A.2 (Lower Bound, [1]) *For any random variable X with mean μ and mean absolute deviation d defined on the interval $[a, b]$, with $-\infty < a < \mu < b < \infty$, $\beta = P(X \ge \mu)$ and any convex function $f(x)$*
$Min_{X\in\mathcal{D}_{[a,b]}(\mu,d,\beta)} E\left[f(x)\right] = p_1 f\left(\mu+\frac{d}{2\beta}\right) + p_2 f\left(\mu-\frac{d}{2(1-\beta)}\right)$ *where*
$p_1 = \beta$; $p_2 = 1-\beta$.

Lower Bound on the Loss Function

$(X-Q)^+$ is a convex function in X. Therefore, for any $a \le Q \le b$, we have the following lower bound for the loss function, i.e., $E\left[(X-Q)^+\right]$.

$$Min_{X\in\mathcal{D}_{[a,b]}(\mu,d,\beta)} E\left[(X-Q)^+\right] = \begin{cases} \mu-Q, & a \le Q \le \mu-\frac{d}{2(1-\beta)}, \\ \left(\mu+\frac{d}{2\beta}-Q\right)\beta, & \mu-\frac{d}{2(1-\beta)} \le Q \le \mu+\frac{d}{2\beta}, \\ 0, & \mu+\frac{d}{2\beta} \le Q \le b. \end{cases}$$

Symmetric Distributions

Letting $\omega = 2(b-\mu) = 2(\mu-a) = b-a$ and $\beta = 0.5$, the best-case distribution and the corresponding lower bound for $E\left[(X-Q)^+\right]$ for this *mean-MAD* ambiguity set are given below, respectively.

$$X_{[-]}^{MAD} = \begin{cases} \mu - d, & w.p.\ \frac{1}{2}, \\ \mu + d, & w.p.\ \frac{1}{2}. \end{cases}$$

$$\ell_{[-]}^{MAD}(Q) = \begin{cases} \mu - Q, & a \le Q \le \mu - d, \\ \frac{\mu+d-Q}{2}, & \mu - d \le Q \le \mu + d, \\ 0, & \mu + d \le Q \le b. \end{cases}$$

A.7 Comparison of Lower Bounds

Lemma 2.2 with variance τv where $\tau \ge 1$ results the following adjusted DR lower bound

$$L_{[-]}^{\tau}(z_Q) = \begin{cases} -z_Q, & z_Q \in \left[z_a, -\frac{\tau}{z_b}\right), \\ \frac{\tau + z_Q z_a}{z_b - z_a}, & z_Q \in \left[-\frac{\tau}{z_b}, -\frac{\tau}{z_a}\right), \\ 0, & z_Q \in \left[-\frac{\tau}{z_a}, z_b\right]. \end{cases}$$

It is known that $\frac{d(b-a)}{2v} \ge 1$ for any distribution [17]. Since $\frac{d(b-a)}{2v} = \frac{d}{\sqrt{v}} z_b$, we have $\tau = \frac{d(b-a)}{2v} \ge 1$ for a symmetric distribution. Therefore, this is a tighter upper bound compared to the original DR upper bound given by Lemma 2.1. Letting $z_b = -z_a$ and $\tau = \frac{d}{\sqrt{v}} z_b = \sqrt{\kappa} z_b$; we can rewrite this bound as

$$L_{[-]}^{\tau}(z_Q) = L_{[-]}^{MAD}(z_Q) = \begin{cases} -z_Q, & z_a \le z \le -\sqrt{\kappa}, \\ \frac{\sqrt{\kappa} - z_Q}{2}, & -\sqrt{\kappa} \le z \le \sqrt{\kappa}\ . \\ 0, & \sqrt{\kappa} \le z_Q \le z_b \end{cases}$$

Therefore, for symmetric distributions, the adjusted mean-variance bound with $\tau = \frac{d}{\sqrt{v}} z_b = \sqrt{\kappa} z_b$ results in the same lower bound provided by mean-MAD ambiguity.

A.8 Standardized DR Loss Functions for Mean-MAD Ambiguity Set

In this section, we assume that both variance and MAD information are available. Therefore, we know the ratios $\kappa = \frac{d^2}{v}$ and $\tau = \frac{d(b-a)}{2v}$. One may then formulate standardized versions of the above

MAD-based bounds, i.e., $L_{[+]}^{MAD}(z_Q) = \frac{\ell_{[+]}^{MAD}(Q)}{\sqrt{v}}$ and $L_{[-]}^{MAD}(z_Q) = \frac{\ell_{[-]}^{MAD}(Q)}{\sqrt{v}}$ where $z_Q = \frac{Q-\mu}{\sqrt{v}}$, as

$$L_{[+]}^{MAD}(z_Q) = \begin{cases} \left(1 - \frac{z_Q}{z_a}\right)\frac{\sqrt{\kappa}}{2} - z_Q, & \text{for } z_a \leq z_Q \leq 0 \\ \frac{\sqrt{\kappa}(z_b - z_Q)}{2z_b}, & \text{for } 0 \leq z_Q \leq z_b. \end{cases} \tag{A.2}$$

$$L_{[-]}^{MAD}(z_Q) = \begin{cases} -z_Q, & z_a \leq z \leq -\sqrt{\kappa}, \\ \frac{\sqrt{\kappa} - z_Q}{2}, & -\sqrt{\kappa} \leq z \leq \sqrt{\kappa}, \\ 0, & \sqrt{\kappa} \leq z_Q \leq z_b. \end{cases} \tag{A.3}$$

Let $L_{[+]}^{\kappa *}(z_Q)$ indicate the adjusted DR upper bound, i.e., $L_{[+]}^{\kappa}(z_Q)$ with $\kappa = \frac{d^2}{v}$. As in the non-standardized setting, this bound is tighter than the mean-MAD bound given by Equation (A.2). Noting $z_b = -z_a$, the lower bound provided by (A.3) coincides with the adjusted DR lower bound, $L_{[-]}^{\tau *}(z_Q)$, which is $L_{[-]}^{\tau}(z_Q)$ at $\tau = \frac{d(b-a)}{2v}$.

A.9 Adjusted DR Bounds Based on a Sample

To determine the parameters for the adjusted DR bounds, we solve the optimization problems provided by Equations (9) and (11). This is equivalent to solving the following problems:

$$\kappa v = \text{Max } 4E\left[(X_i - Q)^+\right] \times E\left[(Q - X_i)^+\right],$$
$$\tau v = \text{Min } bE\left[(X_i - Q)^+\right] + aE\left[(Q - X_i)^+\right].$$

When we do not know the distribution, but have sample observations, we can solve these problems for the sample under consideration. We can then compute the adjusted DR bounds provided in Table 5 using the estimates of $\widehat{\kappa v}$ and $\widehat{\tau v}$ (one may also estimate a, b, and μ from the data).

A.10 Proof of DR Newsvendor Solution

In this section, we derive the optimal solution for the model

$$(\text{DRNP}_{[+]}) \text{ Minimize}_{a \leq Q \leq b} \; p\ell_{[+]}^{DR}(Q, \mu, v) + cQ. \tag{A.4}$$

This problem can be written as

$$(\text{DRNP}_{[+]})(z_Q)\text{Minimize}_{z_a \le z_Q \le z_b}\ pL^{DR}_{[+]}(z_Q)\sqrt{v} + c(\mu + z_Q\sqrt{v}), \tag{A.5}$$

We can solve this problem by considering the three subproblems below, based on the values z_Q can take.

$$\text{DRNP}_{[+]}\text{I}(z_Q): \ \text{Minimize}_{z_a \le z_Q \le \frac{1}{2}\left(z_a - \frac{1}{z_a}\right)} \ -p\left(\frac{\frac{1}{z_a} + z_Q}{1 + \frac{1}{z_a}^2}\right)\sqrt{v} + c(\mu + z_Q\sqrt{v}).$$

$$\text{DRNP}_{[+]}\text{II}(z_Q): \ \text{Minimize}_{\frac{1}{2}\left(z_a - \frac{1}{z_a}\right) < z_Q \le \frac{1}{2}\left(z_b - \frac{1}{z_b}\right)} \ \frac{p}{2}\left(\sqrt{\left(1 + z_Q^2\right)} - z_Q\right)\sqrt{v} + c(\mu + z_Q\sqrt{v}).$$

$$\text{DRNP}_{[+]}\text{III}(z_Q): \ \text{Minimize}_{\frac{1}{2}\left(z_b - \frac{1}{z_b}\right) < z_Q \le z_b} \ p\left(\frac{z_b - z_Q}{1 + {z_b}^2}\right)\sqrt{v} + c(\mu + z_Q\sqrt{v}).$$

Omitting the constants from the objective functions, the solutions for these models are equivalent to those obtained by solving:

$$\text{DRNP}_{[+]}\text{I}'(z_Q): \ \text{Minimize}_{z_a \le z_Q \le \frac{1}{2}\left(z_a - \frac{1}{z_a}\right)} \ \left(c - \frac{p}{1 + \frac{1}{z_a}^2}\right)z_Q.$$

$$\text{DRNP}_{[+]}\text{II}'(z_Q): \ \text{Minimize}_{\frac{1}{2}\left(z_a - \frac{1}{z_a}\right) < z_Q \le \frac{1}{2}\left(z_b - \frac{1}{z_b}\right)} \ \frac{p}{2}\left(\sqrt{\left(1 + z_Q^2\right)} - z_Q\right) + cz_Q.$$

$$\text{DRNP}_{[+]}\text{III}'(z_Q): \ \text{Minimize}_{\frac{1}{2}\left(z_b - \frac{1}{z_b}\right) < z_Q \le z_b} \ \left(c - \frac{p}{1 + {z_b}^2}\right)z_Q.$$

- For $\text{DRNP}_{[+]}\text{I}'(z_Q)$, when $-\frac{1}{z_a} \le \sqrt{\frac{p-c}{c}} = \sqrt{\frac{\rho}{1-\rho}}$, then the coefficient of z_Q is negative. In this case, $z_Q = \frac{1}{2}\left(z_a - \frac{1}{z_a}\right)$ is optimal for this problem. Otherwise, $z_Q = z_a$ is optimal.

- For $\text{DRNP}_{[+]}\text{II}'(z_Q)$, the first order condition suggests $z_Q^* = \frac{1}{2}\left(\sqrt{\frac{\rho}{1-\rho}} - \sqrt{\frac{1-\rho}{\rho}}\right)$. When $-\frac{1}{z_a} \le \sqrt{\frac{\rho}{1-\rho}} \le z_b$ then $\frac{1}{2}\left(z_a - \frac{1}{z_a}\right) \le z_Q^* \le \frac{1}{2}\left(z_b - \frac{1}{z_b}\right)$, and z_Q^* is optimal for this problem. When $-\frac{1}{z_a} > \sqrt{\frac{\rho}{1-\rho}}$, $\left(z_a - \frac{1}{z_a}\right)$ is optimal for this problem. When $z_b < \sqrt{\frac{\rho}{1-\rho}}$, then $z_Q = \frac{1}{2}\left(z_b - \frac{1}{z_b}\right)$ is optimal for this problem.

- For $\text{DRNP}_{[+]}\text{III}'(z_Q)$, when $z_b \le \sqrt{\frac{p-c}{c}} = \sqrt{\frac{\rho}{1-\rho}}$; then the coefficient of z_Q is negative. In this case $z_Q = z_b$ is optimal for this problem. Otherwise $z_Q = \frac{1}{2}\left(z_b - \frac{1}{z_b}\right)$ is optimal.

Combining the solutions of these three models, we have the following result:

- If $\sqrt{\frac{\rho}{1-\rho}} < -\frac{1}{z_a}$, then $z_Q = z_a$ solves $\text{DRNP}_{[+]}\text{I}'(z_Q)$, $z_Q = \frac{1}{2}\left(z_a - \frac{1}{z_a}\right)$ solves $\text{DRNP}_{[+]}\text{II}'(z_Q)$, and $z_Q = \frac{1}{2}\left(z_b - \frac{1}{z_b}\right)$ solves DRNP$_{[+]}$III$'(z_Q)$. Because the objective function is continuous and the end points are included in the optimization problem for each subproblem, the optimal objective function value of DRNP$_{[+]}$I$'(z_Q)$ is smaller than the optimal objective function value of DRNP$_{[+]}$II$'(z_Q)$. Similarly, the optimal objective function value of DRNP$_{[+]}$II$'(z_Q)$is less than or equal to that of DRNP$_{[+]}$III$'(z_Q)$. Therefore $z_Q = z_a$.

- If $\sqrt{\frac{\rho}{1-\rho}} > z_b$, then $z_Q = \frac{1}{2}\left(z_a - \frac{1}{z_a}\right)$ solves DRNP$_{[+]}$I$'(z_Q)$, $z_Q = \frac{1}{2}\left(z_b - \frac{1}{z_b}\right)$ solves DRNP$_{[+]}$II$'(z_Q)$, and $z_Q = z_b$ solves DRNP$_{[+]}$III$'(z_Q)$. When this is the case, DRNP$_{[+]}$III$'(z_Q)$ provides an optimal solution for this problem at $z_Q = z_b$.

- If $-\frac{1}{z_a} < \sqrt{\frac{\rho}{1-\rho}} < z_b$, then $z_Q = \frac{1}{2}\left(z_a - \frac{1}{z_a}\right)$ solves DRNP$_{[+]}$I$'(z_Q)$, $z_Q = \frac{1}{2}\left(\sqrt{\frac{\rho}{1-\rho}} - \sqrt{\frac{1-\rho}{\rho}}\right)$ solves DRNP$_{[+]}$II$'(z_Q)$, and $z_Q = \frac{1}{2}\left(z_b - \frac{1}{z_b}\right)$ solves DRNP$_{[+]}$III$'(z_Q)$. When this is the case, DRNP$_{[+]}$III$'(z_Q)$ provides an optimal solution for this problem at $z_Q = \frac{1}{2}\left(\sqrt{\frac{\rho}{1-\rho}} - \sqrt{\frac{1-\rho}{\rho}}\right)$.

The following therefore characterizes an optimal solution, and this completes the proof.

$$
\begin{aligned}
z_Q &= z_a, \text{ if } \sqrt{\frac{\rho}{1-\rho}} < -\frac{1}{z_a}, \\
z_Q &= \frac{1}{2}\left(\sqrt{\frac{\rho}{1-\rho}} - \sqrt{\frac{1-\rho}{\rho}}\right), \text{ if } -\frac{1}{z_a} \le \sqrt{\frac{\rho}{1-\rho}} \le z_b, \\
z_Q &= z_b, \text{ if } \sqrt{\frac{\rho}{1-\rho}} > z_b.
\end{aligned}
$$

A.11 Proof of Adjusted DR Newsvendor Solution

It is easy to see that the adjusted DR approach simply solves the DR problem with a smaller variance, i.e., with $v' = \kappa v$. We can then define $z'_a = \frac{z_a}{\sqrt{\kappa}}$ and $z'_b = \frac{z_b}{\sqrt{\kappa}}$. The adjusted DR solution in this case then becomes:

$$
\begin{aligned}
z_Q &= z'_a, \text{ if } \sqrt{\frac{\rho}{1-\rho}} < -\frac{1}{z'_a}, \\
z_Q &= \frac{1}{2}\left(\sqrt{\frac{\rho}{1-\rho}} - \sqrt{\frac{1-\rho}{\rho}}\right), \text{ if } -\frac{1}{z'_a} \le \sqrt{\frac{\rho}{1-\rho}} \le z'_b, \\
z_Q &= z'_b, \text{ if } \sqrt{\frac{\rho}{1-\rho}} > z'_b.
\end{aligned}
$$

Mixture				
	Normal	**DR**	**ADR (0.75)**	**ADR (0.66)**
0.51	0.4%	0.4%	0.4%	0.4%
0.61	0.5%	0.3%	0.2%	0.2%
0.71	0.4%	0.1%	0.0%	0.0%
0.81	0.2%	0.0%	0.0%	0.1%
0.91	0.0%	0.2%	0.1%	0.3%
0.93	0.0%	0.4%	0.0%	0.2%
0.95	0.1%	0.1%	0.0%	0.0%
0.97	0.7%	3.6%	0.7%	0.2%
0.99	3.7%	17.1%	9.5%	6.9%

	Normal				**Gamma**			
	Normal	**DR**	**ADR (0.75)**	**ADR (0.66)**	**Normal**	**DR**	**ADR (0.75)**	**ADR (0.66)**
0.51	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.1%	0.1%
0.61	0.0%	0.0%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%
0.71	0.0%	0.1%	0.3%	0.4%	0.1%	0.0%	0.1%	0.1%
0.81	0.0%	0.1%	0.6%	0.9%	0.0%	0.0%	0.3%	0.5%
0.91	0.0%	0.1%	0.2%	0.6%	0.1%	0.0%	0.4%	0.8%
0.93	0.0%	0.7%	0.0%	0.2%	0.1%	0.2%	0.2%	0.6%
0.95	0.0%	2.7%	0.4%	0.0%	0.3%	0.9%	0.0%	0.2%
0.97	0.0%	9.2%	3.8%	2.1%	0.9%	4.1%	0.9%	0.2%
0.99	0.0%	37.7%	26.7%	22.4%	3.2%	24.9%	15.2%	11.6%
	Lognormal				**Pareto**			
	Normal	**DR**	**ADR (0.75)**	**ADR (0.66)**	**Normal**	**DR**	**ADR (0.75)**	**ADR (0.66)**
0.51	0.2%	0.2%	0.2%	0.2%	1.6%	1.6%	1.6%	1.6%
0.61	0.2%	0.1%	0.0%	0.0%	2.4%	1.9%	1.6%	1.5%
0.71	0.2%	0.0%	0.0%	0.0%	2.7%	1.8%	1.3%	1.1%
0.81	0.1%	0.0%	0.2%	0.3%	2.1%	1.4%	0.8%	0.6%
0.91	0.0%	0.0%	0.3%	0.7%	0.5%	0.9%	0.2%	0.0%
0.93	0.2%	0.1%	0.2%	0.6%	0.1%	0.8%	0.1%	0.0%
0.95	0.5%	0.5%	0.0%	0.3%	0.0%	0.7%	0.0%	0.0%
0.97	1.4%	2.6%	0.3%	0.0%	1.3%	0.9%	0.0%	0.0%
0.99	6.2%	18.4%	9.9%	7.0%	13.3%	2.7%	0.4%	0.0%

Table 8: Optimality gap versus ρ ($\rho > 0.5$) when $\frac{\sqrt{v}}{\mu} = 0.3$.

The optimal ordering rule can then be written as follows:

$$
\begin{aligned}
Q &= a, \text{ if } \sqrt{\frac{\rho}{1-\rho}} < \frac{\sqrt{\kappa v}}{\mu}, \\
Q &= \mu + \frac{1}{2}\left(\sqrt{\frac{\rho}{1-\rho}} - \sqrt{\frac{1-\rho}{\rho}}\right)\sqrt{\kappa v}, \text{ if } \frac{\sqrt{\kappa v}}{\mu} \leq \sqrt{\frac{\rho}{1-\rho}} \leq \frac{b-\mu}{\sqrt{\kappa v}}, \\
Q &= b, \text{ if } \sqrt{\frac{\rho}{1-\rho}} > \frac{b-\mu}{\sqrt{\kappa v}}.
\end{aligned}
$$

Mixture								
	Normal	**DR**	**ADR (0.75)**	**ADR (0.66)**				
0.51	2.5%	2.4%	2.4%	2.3%				
0.61	3.6%	2.6%	2.1%	2.0%				
0.71	3.4%	2.0%	1.3%	1.0%				
0.81	2.4%	1.4%	0.5%	0.2%				
0.91	0.5%	1.2%	0.1%	0.0%				
0.93	0.2%	1.5%	0.1%	0.0%				
0.95	0.0%	2.4%	0.3%	0.0%				
0.97	0.6%	4.6%	1.1%	0.3%				
0.99	8.1%	15.2%	7.3%	4.6%				
Normal					**Gamma**			
	Normal	**DR**	**ADR (0.75)**	**ADR (0.66)**	**Normal**	**DR**	**ADR (0.75)**	**ADR (0.66)**
0.51	0.0%	0.0%	0.0%	0.0%	2.6%	2.5%	2.4%	2.4%
0.61	0.0%	0.1%	0.2%	0.2%	2.6%	1.9%	1.5%	1.4%
0.71	0.0%	0.2%	0.6%	0.9%	2.0%	1.1%	0.6%	0.4%
0.81	0.0%	0.2%	1.2%	1.7%	0.8%	0.3%	0.0%	0.0%
0.91	0.0%	0.3%	0.3%	1.1%	0.1%	0.0%	0.4%	0.9%
0.93	0.0%	1.3%	0.0%	0.4%	0.5%	0.0%	0.6%	1.2%
0.95	0.0%	4.7%	0.6%	0.0%	1.7%	0.1%	0.6%	1.4%
0.97	0.0%	15.8%	6.5%	3.6%	5.4%	0.6%	0.2%	0.8%
0.99	0.0%	61.6%	43.7%	36.7%	23.4%	10.5%	3.2%	1.3%
Lognormal					**Pareto**			
	Normal	**DR**	**ADR (0.75)**	**ADR (0.66)**	**Normal**	**DR**	**ADR (0.75)**	**ADR (0.66)**
0.51	3.3%	3.2%	3.2%	3.2%	4.8%	4.6%	4.5%	4.5%
0.61	4.5%	3.4%	2.9%	2.7%	10.4%	8.3%	7.2%	6.7%
0.71	4.6%	3.0%	2.1%	1.7%	14.6%	11.0%	8.8%	8.0%
0.81	3.2%	2.1%	1.0%	0.7%	15.5%	12.6%	9.3%	8.1%
0.91	0.5%	1.0%	0.1%	0.0%	10.3%	12.4%	8.1%	6.6%
0.93	0.0%	0.8%	0.0%	0.0%	7.9%	12.1%	7.6%	6.1%
0.95	0.2%	0.8%	0.0%	0.1%	4.9%	11.7%	7.0%	5.4%
0.97	2.5%	1.0%	0.0%	0.2%	1.4%	11.2%	6.3%	4.6%
0.99	19.5%	3.9%	0.6%	0.1%	2.1%	11.3%	5.7%	3.9%

Table 9: Optimality gap versus ρ ($\rho > 0.5$) when $\frac{\sqrt{v}}{\mu} = 1$.

	Mixture							
	Normal	**DR**	**ADR (0.75)**	**ADR (0.66)**				
0.51	5.2%	5.0%	4.9%	4.8%				
0.61	9.8%	7.6%	6.5%	6.1%				
0.71	10.8%	7.7%	5.8%	5.2%				
0.81	8.4%	6.1%	3.8%	3.0%				
0.91	3.0%	4.5%	1.7%	0.9%				
0.93	1.8%	4.9%	1.6%	0.8%				
0.95	0.8%	6.1%	2.2%	1.1%				
0.97	0.0%	8.4%	3.2%	1.7%				
0.99	4.2%	13.0%	6.1%	4.0%				
	Normal				**Gamma**			
	Normal	**DR**	**ADR (0.75)**	**ADR (0.66)**	**Normal**	**DR**	**ADR (0.75)**	**ADR (0.66)**
0.51	0.0%	0.0%	0.0%	0.0%	13.1%	12.8%	12.7%	12.7%
0.61	0.0%	0.1%	0.2%	0.3%	14.7%	12.5%	11.3%	10.9%
0.71	0.0%	0.3%	0.8%	1.1%	12.3%	9.5%	7.8%	7.1%
0.81	0.0%	0.3%	1.5%	2.2%	6.9%	5.2%	3.4%	2.8%
0.91	0.0%	0.3%	0.4%	1.3%	0.5%	1.1%	0.1%	0.0%
0.93	0.0%	1.6%	0.0%	0.4%	0.0%	0.5%	0.0%	0.1%
0.95	0.0%	5.6%	0.8%	0.1%	0.7%	0.2%	0.2%	0.6%
0.97	0.0%	18.7%	7.7%	4.3%	5.3%	0.1%	0.6%	1.4%
0.99	0.0%	71.4%	50.5%	42.5%	33.1%	1.3%	0.1%	0.7%
	Lognormal				**Pareto**			
	Normal	**DR**	**ADR (0.75)**	**ADR (0.66)**	**Normal**	**DR**	**ADR (0.75)**	**ADR (0.66)**
0.51	9.3%	9.0%	8.8%	8.8%	6.9%	6.5%	6.3%	6.2%
0.61	15.2%	12.4%	10.9%	10.3%	23.1%	18.1%	15.4%	14.4%
0.71	17.5%	13.4%	10.8%	9.8%	36.7%	28.7%	23.4%	21.5%
0.81	14.9%	11.9%	8.6%	7.5%	43.7%	37.0%	29.3%	26.4%
0.91	6.2%	8.0%	4.6%	3.5%	37.3%	42.2%	32.1%	28.4%
0.93	3.8%	6.8%	3.6%	2.5%	32.8%	42.6%	32.1%	28.2%
0.95	1.3%	5.6%	2.5%	1.6%	26.3%	42.7%	31.8%	27.7%
0.97	0.0%	4.3%	1.5%	0.7%	16.6%	42.5%	31.0%	26.7%
0.99	10.4%	3.5%	0.7%	0.2%	2.2%	41.6%	29.5%	25.0%

Table 10: Optimality gap versus ρ ($\rho > 0.5$) when $\frac{\sqrt{v}}{\mu} = 2$.

δ	0.3			1			2		
ρ	ADR (0.66)	ADR (0.75)	DR	ADR (0.66)	ADR (0.75)	DR	ADR (0.66)	ADR (0.75)	DR
0.51	0.4%	1.7%	5.0%	0.9%	3.8%	11.4%	1.3%	5.3%	15.7%
0.53	0.4%	1.8%	5.2%	1.0%	4.0%	11.7%	1.4%	5.4%	16.0%
0.55	0.5%	1.8%	5.3%	1.1%	4.1%	12.0%	1.5%	5.6%	16.4%
0.57	0.5%	1.9%	5.5%	1.3%	4.4%	12.4%	1.7%	5.9%	16.8%
0.59	0.4%	1.9%	5.6%	1.5%	4.7%	12.9%	2.0%	6.3%	17.4%
0.61	0.4%	1.9%	5.8%	1.7%	5.0%	13.4%	2.3%	6.7%	18.0%
0.63	0.5%	2.0%	6.0%	1.9%	5.2%	13.9%	2.8%	7.2%	18.7%
0.65	0.5%	2.1%	6.3%	2.0%	5.4%	14.2%	3.3%	7.8%	19.5%
0.67	0.6%	2.3%	6.6%	2.0%	5.5%	14.5%	3.9%	8.5%	20.5%
0.69	0.7%	2.5%	6.9%	1.9%	5.4%	14.6%	4.6%	9.3%	21.5%
0.71	0.8%	2.6%	7.2%	1.6%	5.2%	14.6%	5.2%	10.0%	22.4%
0.73	1.0%	2.8%	7.6%	1.3%	5.0%	14.5%	5.6%	10.5%	23.2%
0.75	1.2%	3.1%	8.0%	1.0%	4.8%	14.5%	5.9%	10.9%	23.8%
0.77	1.4%	3.4%	8.6%	0.8%	4.7%	14.6%	6.0%	11.1%	24.1%
0.79	1.8%	3.9%	9.3%	0.7%	4.6%	14.8%	5.9%	11.0%	24.2%
0.81	2.3%	4.4%	10.1%	0.8%	4.8%	15.2%	5.4%	10.6%	23.9%
0.83	2.8%	5.1%	11.0%	1.0%	5.1%	15.8%	4.5%	9.7%	23.1%
0.85	3.2%	5.6%	11.8%	1.4%	5.7%	16.7%	3.2%	8.4%	21.9%
0.87	3.5%	6.0%	12.6%	2.2%	6.6%	18.0%	2.0%	7.1%	20.7%
0.89	3.6%	6.3%	13.2%	3.4%	8.0%	19.9%	1.0%	6.2%	19.9%
0.91	3.8%	6.7%	14.2%	5.2%	10.0%	22.5%	0.5%	5.8%	19.7%
0.93	4.4%	7.5%	15.6%	6.8%	11.9%	25.0%	0.8%	6.2%	20.4%
0.95	5.8%	9.3%	18.2%	9.2%	14.6%	28.6%	2.6%	8.2%	23.0%
0.97	9.2%	13.3%	23.7%	14.7%	20.7%	36.2%	8.2%	14.3%	30.3%
0.99	22.7%	28.3%	42.6%	33.5%	41.0%	60.6%	18.8%	25.7%	44.0%

Table 11: Optimality Gap for DRNP.