Piecewise linear lower and upper bounds for the standard normal first order loss function

CrossMark

Roberto Rossi [a,*], S. Armagan Tarim [b], Steven Prestwich [c], Brahim Hnich [d]

[a] Business School, University of Edinburgh, 29 Buccleuch place, EH8 9JS, Edinburgh, United Kingdom
[b] Institute of Population Studies, Hacettepe University, Merkez Kampüsü D Blok Kat:5 Sıhhiye 06100 Ankara, Turkey
[c] Insight Centre for Data Analytics, University College Cork, Western Gateway Building, Western Road, Cork, Ireland
[d] Dept. of Computer Engineering, Izmir University of Economics, Sakarya Caddesi No:156, 35330 balcova, İzmir, Turkey

ARTICLE INFO

ABSTRACT

The first order loss function and its complementary function are extensively used in practical settings. When the random variable of interest is normally distributed, the first order loss function can be easily expressed in terms of the standard normal cumulative distribution and probability density function. However, the standard normal cumulative distribution does not admit a closed form solution and cannot be easily linearised. Several works in the literature discuss approximations for either the standard normal cumulative distribution or the first order loss function and their inverse. However, a comprehensive study on piecewise linear upper and lower bounds for the first order loss function is still missing. In this work, we initially summarise a number of distribution independent results for the first order loss function and its complementary function. We then extend this discussion by focusing first on random variables featuring a symmetric distribution, and then on normally distributed random variables. For the latter, we develop effective piecewise linear upper and lower bounds that can be immediately embedded in MILP models. These linearisations rely on constant parameters that are independent of the mean and standard deviation of the normal distribution of interest. We finally discuss how to compute optimal linearisation parameters that minimise the maximum approximation error.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Consider a random variable ω and a scalar variable x. The first order loss function is defined as

$$\mathcal{L}(x, \omega) = \mathrm{E}[\max(\omega - x, 0)], \tag{1}$$

where E denotes the expected value. The complementary first order loss function is defined as

$$\widehat{\mathcal{L}}(x, \omega) = \mathrm{E}[\max(x - \omega, 0)]. \tag{2}$$

The first order loss function and its complementary function play a key role in several application domains. In inventory control [13] it is often used to express expected inventory holding or shortage costs, as well as service level measures such as the widely adopted "fill rate", also known as β service level [1, p. 94]. In finance the first order loss function may be employed to

---

* Corresponding author. Address: University of Edinburgh, Business School, EH8 9JS Edinburgh, United Kingdom.
 E-mail addresses: roberto.rossi@ed.ac.uk (R. Rossi), armagan.tarim@hacettepe.edu.tr (S.A. Tarim), s.prestwich@cs.ucc.ie (S. Prestwich), hnich.brahim@gmail.com (B. Hnich).

capture risk measures such as the so-called "conditional value at risk" (see e.g. [11]). These examples illustrate possible applications of this function. Of course, the applicability of this function goes beyond inventory theory and finance.

In Section 2, we first summarise a number of distribution independent results for the first order loss function and its complementary function. We then focus on symmetric and normal distributions, for which we discuss ad hoc results in Section 3.

According to one of these results, the first order loss function can be expressed in terms of the cumulative distribution function of the random variable under scrutiny. Depending on the probability distribution adopted, integrating this function may constitute a challenging task. For instance, if the random variable is normally distributed, no closed formulation exists for its cumulative distribution function. Several approximations have been proposed in the literature (see e.g. [3–5,10,12,16,17]), which can be employed to approximate the first order loss function. However, these approximations are generally nonlinear and cannot be easily embedded in mixed integer linear programming (MILP) models.

In Sections 4 and 5, we introduce piecewise linear lower and upper bounds for the first order loss function and its complementary function for the case of normally distributed random variables. These bounds are based on standard bounding techniques from stochastic programming, i.e. Jensen's lower bound and Edmundson–Madansky upper bound [9, pp. 167–168]. The bounds can be readily used in MILP models and do not require instance dependent tabulations. Our linearisation strategy is based on standard optimal linearisation coefficients computed in such a way as to minimise the maximum approximation error, i.e. according to a minimax approach – see [7,8,15] for a similar approach. Optimal coefficients for approximations comprising from two to eleven segments will be presented in Tables 1 and 2; these can be reused to approximate the loss function associated with any normally distributed random variable.

2. The first order loss function and its complementary function

Consider a continuous random variable ω with support over $\mathbb{R}$, probability density function $g_\omega(x) : \mathbb{R} \to (0, 1)$ and cumulative distribution function $G_\omega(x) : \mathbb{R} \to (0, 1)$. The first order loss function can be rewritten as

$$\mathcal{L}(x, \omega) = \int_{-\infty}^{\infty} \max(t - x, 0) g_\omega(t) \, dt = \int_x^{\infty} (t - x) g_\omega(t) \, dt. \tag{3}$$

The complementary first order loss function can be rewritten as

$$\widehat{\mathcal{L}}(x, \omega) = \int_{-\infty}^{\infty} \max(x - t, 0) g_\omega(t) \, dt = \int_{-\infty}^{x} (x - t) g_\omega(t) \, dt. \tag{4}$$

We introduce the following two well-known lemmas.

Lemma 1 [14, p. 338, C.3]. *The first order loss function $\mathcal{L}(x, \omega)$ can also be expressed as*

$$\mathcal{L}(x, \omega) = \int_x^{\infty} (1 - G_\omega(t)) \, dt. \tag{5}$$

Lemma 2 [14, p. 338, C.4]. *The complementary first order loss function $\widehat{\mathcal{L}}(x, \omega)$ can also be expressed as*

$$\widehat{\mathcal{L}}(x, \omega) = \int_{-\infty}^{x} G_\omega(t) \, dt. \tag{6}$$

There is a close relationship between the first order loss function and the complementary first order loss function.

Lemma 3 [14, p. 338, C.5]. *The first order loss function $\mathcal{L}(x, \omega)$ can also be expressed as*

$$\mathcal{L}(x, \omega) = \widehat{\mathcal{L}}(x, \omega) - (x - \tilde{\omega}), \tag{7}$$

where $\tilde{\omega} = \mathrm{E}[\omega]$.

Because of the relation discussed in Lemma 3, in what follows without loss of generality most of the results will be presented for the complementary first order loss function.

Another known result for the first order loss function and its complementary function is their convexity, which we present next.

Lemma 4. *$\mathcal{L}(x, \omega)$ and $\widehat{\mathcal{L}}(x, \omega)$ are convex in x.*

Proof. It is sufficient to show that $\frac{d^2}{dx^2} \widehat{\mathcal{L}}(x, \omega)$ is positive; furthermore, the proof for $\mathcal{L}(x, \omega)$ follows from Lemma 3 and from the fact that $-x$ is convex. $\square$

For a random variable ω with symmetric probability density function, we introduce the following results.

Lemma 5. *If the probability density function of ω is symmetric about a mean value $\tilde{\omega}$, then*

$$\mathcal{L}(x, \omega) = \widehat{\mathcal{L}}(2\tilde{\omega} - x, \omega).$$

Proof.

$$\mathcal{L}(x, \omega) = \int_x^\infty (1 - G_\omega(t)) \, dt = \int_{-\infty}^{\tilde{\omega} - (x - \tilde{\omega})} G_\omega(t) \, dt = \widehat{\mathcal{L}}(2\tilde{\omega} - x, \omega) \qquad \square$$

Lemma 6. *If the probability density function of ω is symmetric about a mean value $\tilde{\omega}$, then*

$$\widehat{\mathcal{L}}(x, \omega) = \widehat{\mathcal{L}}(2\tilde{\omega} - x, \omega) + (x - \tilde{\omega})$$

and

$$\mathcal{L}(x, \omega) = \mathcal{L}(2\tilde{\omega} - x, \omega) - (x - \tilde{\omega}).$$

Proof. Follows immediately from Lemmas 3 and 5. $\square$

The results presented so far are easily extended to the case in which the random variable is discrete. In the following section we present results for the case in which the random variable is normally distributed.

3. The first order loss function for a normally distributed random variable

Let ζ be a normally distributed random variable with mean μ and standard deviation σ. Recall that the normal probability density function is defined as

$$g_\zeta(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{8}$$

No closed form expression exists for the cumulative distribution function

$$G_\zeta(x) = \int_{-\infty}^x g_\zeta(x) \, dx.$$

Let $\phi(x)$ be the standard normal probability density function and $\Phi(x)$ the respective cumulative distribution function.

We next present three known results for a normally distributed random variable: a standardisation result in Lemma 7, and two closed form expressions for the computation of the loss function and of its complementary function in Lemmas 8 and 9.

Lemma 7 [14, p. 338, C.13]. *The complementary first order loss function of ζ can be expressed in terms of the standard Normal cumulative distribution function as*

$$\widehat{\mathcal{L}}(x, \zeta) = \sigma \int_{-\infty}^{\frac{x-\mu}{\sigma}} \Phi(t) \, dt = \sigma \widehat{\mathcal{L}}\left(\frac{x-\mu}{\sigma}, Z\right), \tag{9}$$

where Z is a standard Normal random variable.

Lemma 8 [14, p. 338, C.9]. *The complementary first order loss function $\widehat{\mathcal{L}}(x, \zeta)$ can be rewritten in closed form as*

$$\widehat{\mathcal{L}}(x, \zeta) = \sigma\left(\phi\left(\frac{x-\mu}{\sigma}\right) + \Phi\left(\frac{x-\mu}{\sigma}\right)\frac{x-\mu}{\sigma}\right).$$

From Lemma 3 the first order loss function of ζ can be expressed as

$$\mathcal{L}(x, \zeta) = -(x - \mu) + \sigma \int_{-\infty}^{\frac{x-\mu}{\sigma}} \Phi(t) \, dt. \tag{10}$$

Recall that an alternative expression is obtained via Lemma 1,

$$\mathcal{L}(x, \zeta) = \sigma \int_{\frac{x-\mu}{\sigma}}^\infty (1 - \Phi(t)) \, dt. \tag{11}$$

From Lemmas 3 and 8 we obtain the closed form expression

$$\mathcal{L}(x, \zeta) = -(x - \mu) + \sigma\left(\phi\left(\frac{x-\mu}{\sigma}\right) + \Phi\left(\frac{x-\mu}{\sigma}\right)\frac{x-\mu}{\sigma}\right). \tag{12}$$

Lemma 9 [14, p. 338, C.8]. *The first order loss function $\mathcal{L}(x, \zeta)$ can be rewritten in closed form as*

$$\mathcal{L}(x, \zeta) = \sigma\left(\phi\left(\frac{x-\mu}{\sigma}\right) - \left(1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right)\frac{x-\mu}{\sigma}\right). \qquad \text{x} \qquad\qquad \text{xi}$$

Furthermore, since the Normal distribution is symmetric, both Lemmas 5 and 6 hold.

4. Jensen's lower bound for the standard normal first order loss function

We introduce a well-known inequality from stochastic programming.

4.1. Jensen's lower bound

Theorem 1 (*Jensen's inequality, [2, p. 140]*). *Consider a random variable ω with support Ω and a function $f(x, s)$, which for a fixed x is convex for all $s \in \Omega$, then*

$$\mathrm{E}[f(x, \omega)] \geqslant f(x, \mathrm{E}[\omega]).$$

Common discrete lower bounding approximations in stochastic programming are extensions of Jensen's inequality. The usual strategy is to find a low cardinality discrete set of realisations representing a good approximation of the true underling distribution. Birge and Louveaux [2, p. 288], discuss one of these discrete lower bounding approximations which consists in partitioning the support Ω into a number of disjoint regions, Jensen's bound is then applied in each of these regions.

More formally, let $g_\omega(\cdot)$ denote the probability density function of ω and consider a partition of the support Ω of ω into N disjoint compact subregions $\Omega_1, \ldots, \Omega_N$. We define, for all $i = 1, \ldots, N$

$$p_i = \mathrm{Pr}\{\omega \in \Omega_i\} = \int_{\Omega_i} g_\omega(t)\, dt$$

and

$$\mathrm{E}[\omega|\Omega_i] = \frac{1}{p_i} \int_{\Omega_i} t g_\omega(t)\, dt.$$

Theorem 2 [2, p. 289].

$$\mathrm{E}[f(x, \omega)] \geqslant \sum_{i=1}^{N} p_i f(x, \mathrm{E}[\omega|\Omega_i]).$$

Theorem 3. *Given a random variable ω Jensen's bound (Theorem 1) is applicable to the first order loss function $\mathcal{L}(x, \omega)$ and its complementary function $\widehat{\mathcal{L}}(x, \omega)$.*

Proof. Follows immediately from Lemma 4. □

Having established this result, we must then decide how to partition the support ω in order to obtain a good bound. In fact, to generate good lower bounds, it is necessary to carefully select the partition of the support ω. The optimal partitioning strategy will depend, of course, on the probability distribution of the random variable ω.

4.2. Minimax discrete lower bounding approximation

We discuss a minimax strategy for generating discrete lower bounding approximations of the (complementary) first order loss function. In this strategy, we partition the support ω into a predefined number of regions N in order to minimise the maximum approximation error.

Consider a random variable ω and the associated complementary first order loss function

$$\widehat{\mathcal{L}}(x, \omega) = \mathrm{E}[\max(x - \omega, 0)];$$

assume that the support Ω of ω is partitioned into N disjoint subregions $\Omega_1, \ldots, \Omega_N$.

Lemma 10. *For the (complementary) first order loss function the lower bound presented in Theorem 2 is a piecewise linear function with $N + 1$ segments.*

Proof. Consider the bound presented in [Theorem 2](#) and let $f(x, \omega) = \max(x - \omega, 0)$,

$$\widehat{\mathcal{L}}_{lb}(x, \omega) = \sum_{i=1}^{N} p_i \max(x - E[\omega|\Omega_i], 0) \qquad \text{Jensen}$$

this function is equivalent to

$$\widehat{\mathcal{L}}_{lb}(x, \omega) = \begin{cases} 0 & -\infty \leqslant x \leqslant E[\omega|\Omega_1], \\ p_1 x - p_1 E[\omega|\Omega_1] & E[\omega|\Omega_1] \leqslant x \leqslant E[\omega|\Omega_2], \\ (p_1 + p_2)x - (p_1 E[\omega|\Omega_1] + p_2 E[\omega|\Omega_2]) & E[\omega|\Omega_2] \leqslant x \leqslant E[\omega|\Omega_3], \\ \vdots & \vdots \\ (p_1 + p_2 + \ldots + p_N)x - (p_1 E[\omega|\Omega_1] + p_2 E[\omega|\Omega_2] + \ldots + p_N E[\omega|\Omega_N]) & E[\omega|\Omega_{N-1}] \leqslant x \leqslant E[\omega|\Omega_N], \end{cases}$$

which is piecewise linear in x with breakpoints at $E[\omega|\Omega_1], E[\omega|\Omega_2], \ldots, E[\omega|\Omega_N]$. The proof for the first order loss function follows a similar reasoning. □

Lemma 11. *Consider the ith linear segment of* $\widehat{\mathcal{L}}_{lb}(x, \omega)$

$$\widehat{\mathcal{L}}_{lb}^i(x, \omega) = x \sum_{k=1}^{i} p_k - \sum_{k=1}^{i} p_k E[\omega|\Omega_k] \quad E[\omega|\Omega_i] \leqslant x \leqslant E[\omega|\Omega_{i+1}],$$

where $i = 1, \ldots, N$. *Let* $\Omega_i = [a, b]$, *then* $\widehat{\mathcal{L}}_{lb}^i(x, \omega)$ *is tangent to* $\widehat{\mathcal{L}}(x, \omega)$ *at* $x = b$. *Furthermore, the 0th segment* $x = 0$ *is tangent to* $\widehat{\mathcal{L}}(x, \omega)$ *at* $x = -\infty$.

Proof. Note that

$$\widehat{\mathcal{L}}_{lb}^i(x, \omega) = x \sum_{k=1}^{i} \int_{\Omega_k} g_\omega(t)\, dt - \sum_{k=1}^{i} \int_{\Omega_k} t g_\omega(t)\, dt$$

and that

$$\Omega_1 \cup \Omega_2 \cup \ldots \cup \Omega_i = (-\infty, b]$$

it follows

$$\widehat{\mathcal{L}}_{lb}^i(x, \omega) = x \int_{-\infty}^{b} g_\omega(t)\, dt - \int_{-\infty}^{b} t g_\omega(t)\, dt$$

and

$$\widehat{\mathcal{L}}_{lb}^i(x, \omega) = G_\omega(b)(x - b) + \int_{-\infty}^{b} G_\omega(t)\, dt,$$

which is the equation of the tangent line to $\widehat{\mathcal{L}}(x, \omega)$ at a given point b, that is

$$y = \widehat{\mathcal{L}}(b, \omega)'(x - b) + \widehat{\mathcal{L}}(b, \omega).$$

$x = 0$ is tangent to $\widehat{\mathcal{L}}(x, \omega)$ at $x = -\infty$ since $\widehat{\mathcal{L}}(x, \omega)$ is convex, positive and

$$\lim_{x \to -\infty} \widehat{\mathcal{L}}(x, \omega) = 0.$$

The very same reasoning can be easily applied to the first order loss function. □

Lemma 12. *The maximum approximation error between* $\widehat{\mathcal{L}}_{lb}(x, \omega)$ *and* $\widehat{\mathcal{L}}(x, \omega)$ *will be attained at a breakpoint.*

Proof. By recalling that $\widehat{\mathcal{L}}(x, \omega)$ is convex ([Lemma 4](#)), since $\widehat{\mathcal{L}}_{lb}(x, \omega)$ is piecewise linear ([Lemma 10](#)) and each segment of $\widehat{\mathcal{L}}_{lb}(x, \omega)$ is tangent to $\widehat{\mathcal{L}}(x, \omega)$ ([Lemma 11](#)), it follows that the maximum error will be attained at a breakpoint. □

Theorem 4. *Given the number of regions* N, $\Omega_1, \ldots, \Omega_N$ *is an optimal partition of the support* Ω *of* ω *under a minimax strategy, if and only if approximation errors at breakpoints are all equal.*

Proof. The proof of this theorem follows a reasoning similar to the one presented in [8]. □

By using this last result it is possible to derive a set of equations that can be solved for computing an optimal partitioning. Let us consider the error e_i at breakpoint i, this can be expressed as

$$e_i = \widehat{\mathcal{L}}(E[\omega|\Omega_i], \omega) - \widehat{\mathcal{L}}_{lb}^i(E[\omega|\Omega_i], \omega),$$

where $\Omega_i = [a_i, b_i]$. Since we have N breakpoints to check, we must solve a system comprising the following $N - 1$ equations

$$e_1 = e_i \quad \text{for } i = 2, \ldots, N$$

under the following restrictions

$$
\begin{aligned}
&a_1 = -\infty, \\
&b_N = \infty, \\
&a_i \leqslant b_i \quad \text{for } i = 1, \ldots, N, \\
&b_i = a_{i+1} \quad \text{for } i = 1, \ldots, N-1,
\end{aligned}
$$

The system therefore involves $N - 1$ variables, each of which identifies the boundary between two disjoint regions Ω_i and Ω_{i+1}.

Theorem 5. *Assume that the probability density function of ω is symmetric about a mean value $\bar{\omega}$. Then, under a minimax strategy, if $\Omega_1, \ldots, \Omega_N$ is an optimal partition of the support Ω of ω, breakpoints will be symmetric about $\bar{\omega}$.*

Proof. This follows from Lemma 6 and Theorem 4. □

In this case, by exploiting the symmetry of the piecewise linear approximation, an optimal partitioning can be derived by solving a smaller system comprising $\lceil N/2 \rceil$ equations, where N is the number of regions Ω_i and $\lceil x \rceil$ rounds x to the next integer value.

Unfortunately, equations in the above system are nonlinear and do not admit a closed form solution in the general case.

4.2.1. Normal distribution

We will next discuss the system of equations that leads to an optimal partitioning for the case of a standard Normal random variable Z. This partitioning leads to a piecewise linear approximation that is, in fact, easily extended to the general case of a normally distributed variable ζ with mean μ and standard deviation σ via Lemma 7. This equation suggests that the error of this approximation is independent of μ and proportional to σ.

Consider a partitioning for the support Ω of Z into N adjacent regions $\Omega_i = [a_i, b_i]$, where $i = 1, \ldots, N$. From Theorem 5, if N is odd, then $b_{\lceil N/2 \rceil} = 0$ and $b_i = -b_{N+1-i}$, if N is even, then $b_i = -b_{N+1-i}$. We shall use Lemma 8 for expressing $\widehat{\mathcal{L}}(x, Z)$. Then, by observing that

$$\int_{a_i}^{b_i} t\phi(t)\, dt = \phi(a_i) - \phi(b_i)$$

and that $p_1 + p_2 + \cdots + p_i = \Phi(b_i)$, we rewrite

$$\widehat{\mathcal{L}}_{lb}^i(E[Z|\Omega_i], Z) = \Phi(b_i)E[Z|\Omega_i] - \sum_{k=1}^{i}(\phi(a_i) - \phi(b_i)) \tag{13}$$

$$= \Phi(b_i)E[Z|\Omega_i] - (\phi(a_1) - \phi(b_i)) \tag{14}$$

$$= \Phi(b_i)E[Z|\Omega_i] + \phi(b_i). \tag{15}$$

To express the conditional expectation $E[Z|\Omega_i]$ we proceed as follows: let $p_i = \Phi(b_i) - \Phi(a_i)$, it follows

$$E[Z|\Omega_i] = \frac{\phi(a_i) - \phi(b_i)}{\Phi(b_i) - \Phi(a_i)}.$$

To solve the above system of non-linear equations we will exploit the close connections between finding a local minimum and solving a set of nonlinear equations. In particular, we will use the Gauss–Newton method to find a partition $\Omega_1, \ldots, \Omega_N$ of the support of Z that minimises the following sum of squares

$$\sum_{k=2}^{N}(e_1 - e_k)^2$$

This minimisation problem can be solved by software packages such as Mathematica (see `NMinimize`).

4.2.2. Numerical examples

The classical Jensen's bound for the complementary first order loss function of a standard Normal random variable Z is show in Fig. 1. This is obtained by considering a degenerate partition of the support of Z comprising only a single region $\Omega_1 = [-\infty, \infty]$. In practice, we simply replace Z by its expected value, i.e. zero. Therefore we simply have $\widehat{\mathcal{L}}_{lb} = \max(x, 0)$. The maximum error of this piecewise linear approximation occurs for $x = 0$ and it is equal to $1/\sqrt{2\pi}$.

If we split the support of Z into four regions (Fig. 2), the solution to the system of nonlinear equations prescribes to split Ω at $b_1 = -0.886942$, $b_2 = 0$, $b_3 = 0.886942$. The maximum error is 0.0339052 and it is observed at $x \in \{\pm 1.43535, \pm 0.415223\}$.

In Table 1 we report parameters of $\widehat{\mathcal{L}}_{lb}(x, Z)$ with up to eleven segments. In Fig. 3 we present the approximation error of $\widehat{\mathcal{L}}_{lb}(x, Z)$ with up to eleven segments.
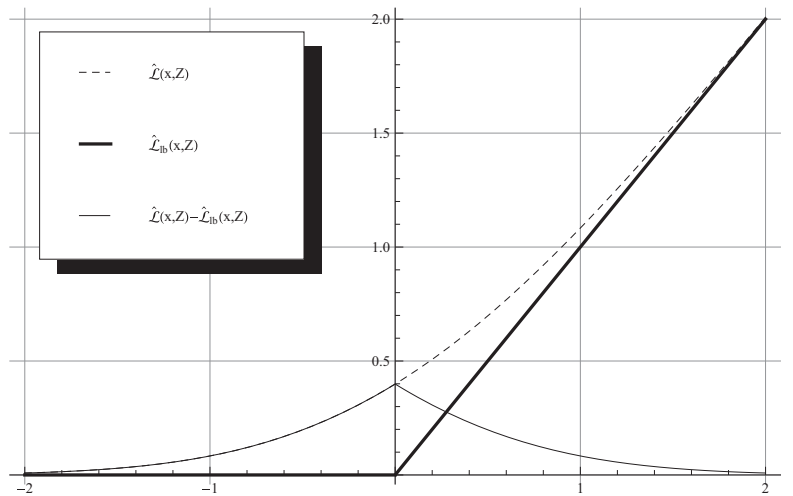
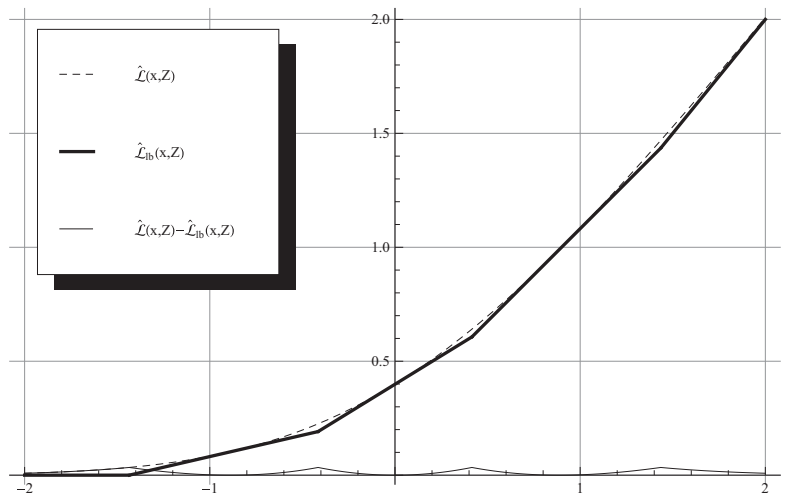**Fig. 1.** Classical Jensen's bound for $\widehat{\mathcal{L}}(x, Z)$.

Fig. 2. Five-segment piecewise Jensen's bound for $\widehat{\mathcal{L}}(x, Z)$.

In Fig. 4 we exploited Lemma 7 to obtain the five-segment piecewise Jensen's bound for $\widehat{\mathcal{L}}(x, \zeta)$, where ζ is a normally distributed random variable with mean $\mu = 20$ and standard deviation $\sigma = 5$. The maximum error is $\sigma 0.0339052$ and it is observed at $x \in \{\sigma(\pm 1.43535) + \mu, \sigma(\pm 0.415223) + \mu\}$.

5. A piecewise linear upper bound for the standard normal first order loss function

In this section we introduce a simple bounding technique that exploits convexity of the (complementary) first order loss function to derive a piecewise linear upper bound.

5.1. A piecewise linear upper bound

Without loss of generality we shall introduce the bound for the complementary first order loss function. Consider a random variable ω with support Ω. From Lemma 4, $\widehat{\mathcal{L}}(x, \omega)$ is convex in x regardless of the distribution of ω. Given an interval $[a, b] \in \mathbb{R}$, it is possible to construct an upper bound by exploiting the very same definition of convexity, that is by constructing a straight line $\widehat{\mathcal{L}}_{ub}(x, \omega)$ between the two points $(a, \widehat{\mathcal{L}}(a, \omega))$ and $(b, \widehat{\mathcal{L}}(b, \omega))$. The slope ($\alpha$) and the intercept ($\beta$) of this line can be easily computed

Table 1
Parameters of $\widehat{\mathcal{L}}_{lb}(x, Z)$ with up to eleven segments.

| Segments	Error	i	1	2	3	4	5	6	7	8	9	10
Piecewise linear approximation parameters												
2	0.398942	b_i	∞									
		p_i	1									
		$E[\omega\|\Omega_i]$	0									
3	0.120656	b_i	0	∞								
		p_i	0.5	0.5								
		$E[\omega\|\Omega_i]$	−0.797885	0.797885								
4	0.0578441	b_i	−0.559725	0.559725	∞							
		p_i	0.287833	0.424333	0.287833							
		$E[\omega\|\Omega_i]$	−1.18505	0	1.18505							
5	0.0339052	b_i	−0.886942	0	0.886942	∞						
		p_i	0.187555	0.312445	0.312445	0.187555						
		$E[\omega\|\Omega_i]$	−1.43535	−0.415223	0.415223	1.43535						
6	0.0222709	b_i	−1.11507	−0.33895	0.33895	1.11507	∞					
		p_i	0.132411	0.234913	0.265353	0.234913	0.132411					
		$E[\omega\|\Omega_i]$	−1.61805	−0.691424	0	0.691424	1.61805					
7	0.0157461	b_i	−1.28855	−0.579834	0	0.579834	1.28855	∞				
		p_i	0.0987769	0.182236	0.218987	0.218987	0.182236	0.0987769				
		$E[\omega\|\Omega_i]$	−1.7608	−0.896011	−0.281889	0.281889	0.896011	1.7608				
8	0.0117218	b_i	−1.42763	−0.765185	−0.244223	0.244223	0.765185	1.42763	∞			
		p_i	0.0766989	0.145382	0.181448	0.192942	0.181448	0.145382	0.0766989			
		$E[\omega\|\Omega_i]$	−1.87735	−1.05723	−0.493405	0	0.493405	1.05723	1.87735			
9	0.00906529	b_i	−1.54317	−0.914924	−0.433939	0	0.433939	0.914924	1.54317	∞		
		p_i	0.0613946	0.118721	0.152051	0.167834	0.167834	0.152051	0.118721	0.0613946		
		$E[\omega\|\Omega_i]$	−1.97547	−1.18953	−0.661552	−0.213587	0.213587	0.661552	1.18953	1.97547		
10	0.00721992	b_i	−1.64166	−1.03998	−0.58826	−0.19112	0.19112	0.58826	1.03998	1.64166	∞	
		p_i	0.0503306	0.0988444	0.129004	0.146037	0.151568	0.146037	0.129004	0.0988444	0.0503306	
		$E[\omega\|\Omega_i]$	−2.05996	−1.30127	−0.8004	−0.384597	0.	0.384597	0.8004	1.30127	2.05996	
11	0.00588597	b_i	−1.72725	−1.14697	−0.717801	−0.347462	0.	0.347462	0.717801	1.14697	1.72725	∞
		p_i	0.0420611	0.0836356	0.110743	0.127682	0.135878	0.135878	0.127682	0.110743	0.0836356	0.0420611
		$E[\omega\|\Omega_i]$	−2.13399	−1.39768	−0.9182	−0.526575	−0.17199	0.17199	0.526575	0.9182	1.39768	2.13399
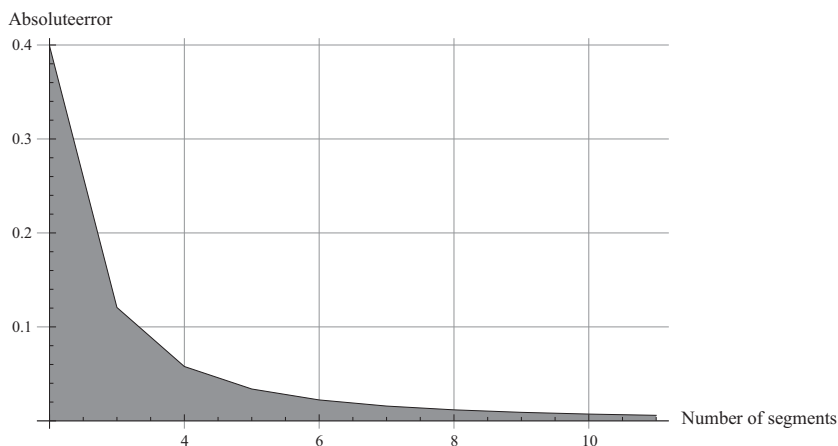
Fig. 3. Approximation error of $\widehat{\mathcal{L}}_{lb}(x, Z)$ with up to eleven segments.
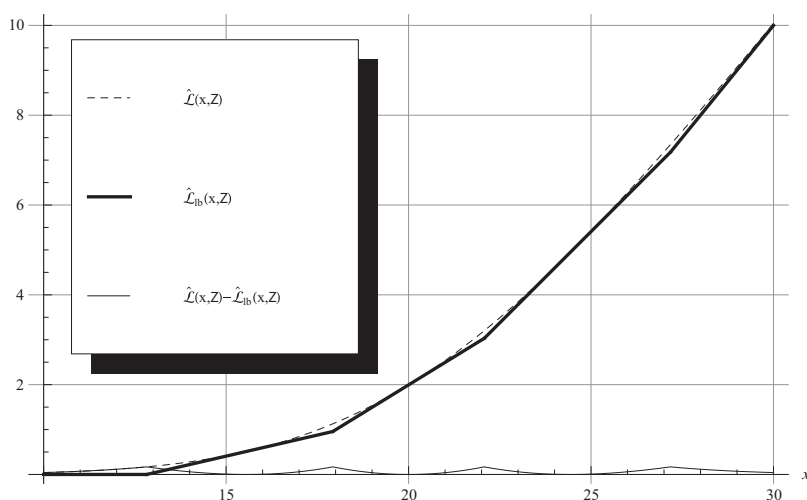

Fig. 4. Five-segment piecewise Jensen's bound for $\widehat{\mathcal{L}}(x, \zeta)$, where $\mu = 20$ and $\sigma = 5$.

$$\alpha = \frac{\widehat{\mathcal{L}}(b, \omega) - \widehat{\mathcal{L}}(a, \omega)}{b - a}, \quad \beta = \frac{b\widehat{\mathcal{L}}(a, \omega) - a\widehat{\mathcal{L}}(b, \omega)}{b - a}.$$

The upper bound is then

$$\widehat{\mathcal{L}}_{ub}(x, \omega) = \alpha x + \beta \quad a \leqslant x \leqslant b \tag{16}$$

$$= \widehat{\mathcal{L}}(a, \omega)\frac{b - x}{b - a} + \widehat{\mathcal{L}}(b, \omega)\frac{x - a}{b - a} \quad a \leqslant x \leqslant b \tag{17}$$

We can improve the quality of this bound by partitioning the domain $\mathbb{R}$ of $\widehat{\mathcal{L}}(x, \omega)$ into N disjoint regions $\mathcal{D}_i = [a_i, b_i]$, $i = 1, \ldots, N$. The selected regions must be all compact and adjacent. Because of the convexity of $\widehat{\mathcal{L}}(x, \omega)$ the bound can be then applied to each of these regions considered separately.

However, since $\widehat{\mathcal{L}}(x, \omega)$ is defined over $\mathbb{R}$, it is not possible to guarantee a complete covering of the domain by using compact regions. We must therefore add two extreme regions $\mathcal{D}_0 = [-\infty, b_0 = a_1]$ and $\mathcal{D}_{N+1} = [a_{N+1} = b_N, \infty]$ to ensure the one obtained is indeed an upper bound for each $x \in \mathbb{R}$. By noting that

$$\lim_{x \to -\infty} \widehat{\mathcal{L}}(x, \omega) = 0 \quad \text{and} \quad \lim_{x \to \infty} \widehat{\mathcal{L}}(x, \omega) = x$$

it is easy to derive equations for the lines associated with these two extra regions. In particular, we associate with $\mathcal{D}_0$ a horizontal line with slope $\alpha = 0$ and intercept $\beta = \widehat{\mathcal{L}}(a_1, \omega)$, and with $\mathcal{D}_{N+1}$ a line with slope $\alpha = 1$ and intercept $\beta = \widehat{\mathcal{L}}(b_N, \omega) - b_N$.

Also in this case, we must then decide how to partition the domain $\mathbb{R}$ into $N + 2$ intervals $\mathcal{D}_0, \ldots, \mathcal{D}_{N+1}$ to obtain a tight bound. Once more, the optimal partitioning strategy will depend on the probability distribution of the random variable ω.

5.2. Minimax piecewise linear upper bound

We discuss a minimax strategy for generating a piecewise linear upper bound of the (complementary) first order loss function $\widehat{\mathcal{L}}(x, \omega)$. In this strategy, we partition the domain $\mathbb{R}$ of x into a predefined number of regions $N + 2$ in order to minimise the maximum approximation error. Note that, since this domain is not compact, one needs at least two regions to derive a piecewise linear upper bound.

Consider a random variable ω and the associated complementary first order loss function

$$\widehat{\mathcal{L}}(x, \omega) = \mathrm{E}[\max(x - \omega, 0)];$$

assume that the domain $\mathbb{R}$ of x is partitioned into $N + 2$ disjoint adjacent subregions $\mathcal{D}_0, \ldots, \mathcal{D}_{N+1}$, where $\mathcal{D}_0 = [-\infty, a_1]$, $\mathcal{D}_i = [a_i, b_i]$, for $i = 1, \ldots, N$, and $\mathcal{D}_{N+1} = [b_N, \infty]$, and consider the following piecewise linear upper bound

$$\widehat{\mathcal{L}}_{ub}(x, \omega) = \begin{cases} \widehat{\mathcal{L}}(a_1, \omega) & x \in \mathcal{D}_0, \\ \vdots \\ \widehat{\mathcal{L}}(a_i, \omega) \frac{b_i - x}{b_i - a_i} + \widehat{\mathcal{L}}(b_i, \omega) \frac{x - a_i}{b_i - a_i} & x \in \mathcal{D}_i, \\ \vdots \\ x + \widehat{\mathcal{L}}(b_N, \omega) - b_N & x \in \mathcal{D}_{N+1}. \end{cases}$$

Let $\widehat{\mathcal{L}}_{ub}^i(x, \omega)$ be the linear segment of $\widehat{\mathcal{L}}_{ub}(x, \omega)$ over $\mathcal{D}_i$, for $i = 0, \ldots, N + 1$.

Lemma 13. *Consider $\widehat{\mathcal{L}}_{ub}^i(x, \omega)$, where $i = 1, \ldots, N$; the maximum approximation error between $\widehat{\mathcal{L}}_{ub}^i(x, \omega)$ and $\widehat{\mathcal{L}}(x, \omega)$ will be attained for*

$$\bar{x}_i = G_\omega^{-1}\left(\frac{\widehat{\mathcal{L}}(b_i, \omega) - \widehat{\mathcal{L}}(a_i, \omega)}{b_i - a_i}\right).$$

Proof. The idea here is to derive a line that is tangent to $\widehat{\mathcal{L}}(x, \omega)$ and that has a slope equal to that of the ith linear segment of $\widehat{\mathcal{L}}_{ub}(x, \omega)$. We have already discussed in Lemma 11 that the equation of the tangent to $\widehat{\mathcal{L}}(x, \omega)$ at a given point $\bar{x}_i$ is

$$y = \widehat{\mathcal{L}}(\bar{x}_i, \omega)'(x - p) + \widehat{\mathcal{L}}(\bar{x}_i, \omega)$$

that is

$$\widehat{\mathcal{L}}_{lb}^i(x, \omega) = G_\omega(\bar{x}_i)(x - p) + \int_{-\infty}^{\bar{x}_i} G_\omega(t)\, dt.$$

The slope $G_\omega(\bar{x}_i)$ only depends on $\bar{x}_i$. To find a tangent with a slope equal to that of the ith linear segment of $\widehat{\mathcal{L}}_{ub}(x, \omega)$, we simply let

$$G_\omega(\bar{x}_i) = \frac{\widehat{\mathcal{L}}(b_i, \omega) - \widehat{\mathcal{L}}(a_i, \omega)}{b_i - a_i}$$

and invert the cumulative distribution function. $\square$

Note that the maximum approximation error for the linear segment over $\mathcal{D}_0$ is $\widehat{\mathcal{L}}(a_1, \omega)$ and the maximum approximation error for the linear segment over $\mathcal{D}_{N+1}$ is $\widehat{\mathcal{L}}(b_N, \omega) - b_N$. This can be inferred from the fact that $\widehat{\mathcal{L}}(x, \omega)$ monotonically approaches 0 for $x \to -\infty$ and x for $x \to \infty$.

Theorem 6. *$\mathcal{D}_0, \ldots, \mathcal{D}_{N+1}$ is an optimal partition under a minimax strategy, if and only if the maximum approximation error between $\widehat{\mathcal{L}}(x, \omega)$ and each linear segment of $\widehat{\mathcal{L}}_{ub}(x, \omega)$ is the same.*

Proof. The proof follows from Lemma 13 and Theorem 4. $\square$

The key insight needed to understand this result is the following. In Theorem 4 we showed that the approximation errors at breakpoints for the piecewise linear lower bound presented are all equal to each other and also equal to the maximum approximation error; in Lemma 11 we showed that the ith piecewise linear segment of this lower bound

agrees with the original function at point b_i, where $\Omega_i = [a_i, b_i]$ is the ith partition of the support of Z, for $i = 1, \ldots, N$. Since the first order loss function is convex and we know the maximum approximation error, by shifting up the piecewise linear lower bound by a value equal to the maximum approximation error we immediately obtain a piecewise linear upper bound comprising $N + 1$ segments. This upper bound agrees with the original function at $E[Z|\Omega_i]$, for $i = 1, \ldots, N$. The maximum approximation error will be attained at those points in which the lower bound was tangent to the original function, that is $a_1, b_1, b_2, \ldots, b_N$.

By using this result it is possible to derive a set of equations that can be solved for computing an optimal partitioning. Let us consider the maximum approximation error e_i associated with the ith linear segment of $\widehat{\mathcal{L}}_{ub}(x, \omega)$, this can be expressed as

$$e_i = \widehat{\mathcal{L}}_{ub}^i(\bar{x}_i, \omega) - \widehat{\mathcal{L}}(\bar{x}_i, \omega),$$

where $i = 1, \ldots, N$; furthermore $e_0 = \widehat{\mathcal{L}}(a_1, \omega)$ and $e_{N+1} = \widehat{\mathcal{L}}(b_N, \omega) - b_N$. Since we have $N + 2$ segments to check, we must solve a system comprising the following $N + 1$ equations

$$e_0 = e_i \quad \text{for } i = 1, \ldots, N + 1$$

under the following restrictions

$$a_i \leqslant b_i \quad \text{for } i = 1, \ldots, N,$$
$$b_i = a_{i+1} \quad \text{for } i = 1, \ldots, N - 1.$$

The system involves $N + 1$ variables, each of which identifies the boundary between two disjoint adjacent regions $\mathcal{D}_i$ and $\mathcal{D}_{i+1}$.

Theorem 7. *Assume that the probability density function of ω is symmetric about a mean value $\bar{\omega}$. Then, under a minimax strategy, if $\mathcal{D}_0, \ldots, \mathcal{D}_{N+1}$ is an optimal partition of the domain, breakpoints will be symmetric about the mean value $\bar{\omega}$.*

Proof. This follows from Lemma 6 and Theorem 6.   □

In this case, by exploiting the symmetry of the piecewise linear approximation, an optimal partitioning can be derived by solving a smaller system comprising $\lceil (N + 1)/2 \rceil$ equations, where $N + 2$ is the number of regions $\mathcal{D}_i$ and $\lceil x \rceil$ rounds x to the next integer value.

As in the case of Jensen's bound, equations in the above system are nonlinear and do not admit a closed form solution in the general case. We will briefly discuss next how to derive the system of nonlinear equations for the case of a standard normal random variable Z. The upper bound presented is closely related to a well-known inequality from stochastic programming, see e.g. [9, p. 168], [6, p. 316], and [2, pp. 291–293]. As pointed out in [9, p. 168], Edmundson–Madanski's upper bound can be seen as a bound where the original distribution is replaced by a two point distribution and the problem itself is unchanged, or it can be viewed as a bound where the distribution is left unchanged and the original function is replaced by a linear affine function represented by a straight line. The above discussion clearly demonstrates the dual nature of this upper bound.

5.3. Normal distribution

We will next discuss the system of equations that leads to an optimal partitioning for the case of a standard Normal random variable Z. This partitioning leads to a piecewise linear approximation that is, in fact, easily extended to the general case of a normally distributed variable ζ with mean μ and standard deviation σ via Lemma 7. Also for this second approximation this equation suggests that the error is independent of μ and proportional to σ.

Consider a partitioning for the domain of x in $\widehat{\mathcal{L}}(x, \omega)$ into $N + 2$ adjacent regions $\mathcal{D}_i = [a_i, b_i]$, where $i = 0, \ldots, N + 1$. From Theorem 7, if N is odd, then $b_{\lceil N/2 \rceil} = 0$ and $b_i = -b_{N+1-i}$, if N is even, then $b_i = -b_{N+1-i}$. Also in this case, we shall use Lemma 8 for expressing $\widehat{\mathcal{L}}(x, Z)$, and we will exploit the close connections between finding a local minimum and solving a set of nonlinear equations. We will therefore use the Gauss–Newton method to minimise the following sum of squares

$$\sum_{k=1}^{N+1} (e_0 - e_k)^2.$$

This minimisation problem can be solved by software packages such as Mathematica (see `NMinimize`).

5.3.1. Numerical examples

A two-segment piecewise linear upper bound for the complementary first order loss function of a standard Normal random variable Z is shown in Fig. 5. This bound has been obtained, under the minimax criterion previously described, by considering a single breakpoint in the domain, i.e. $x = 0$. Of course, the maximum error of this piecewise linear approximation occurs for $x = \pm\infty$ and it is equal to $1/\sqrt{2\pi}$. It is easy to observe that this upper bound can be obtained by adding to the classical Jensen's lower bound presented in Fig. 1 a constant value equal to its maximum approximation error, i.e. $1/\sqrt{2\pi}$.
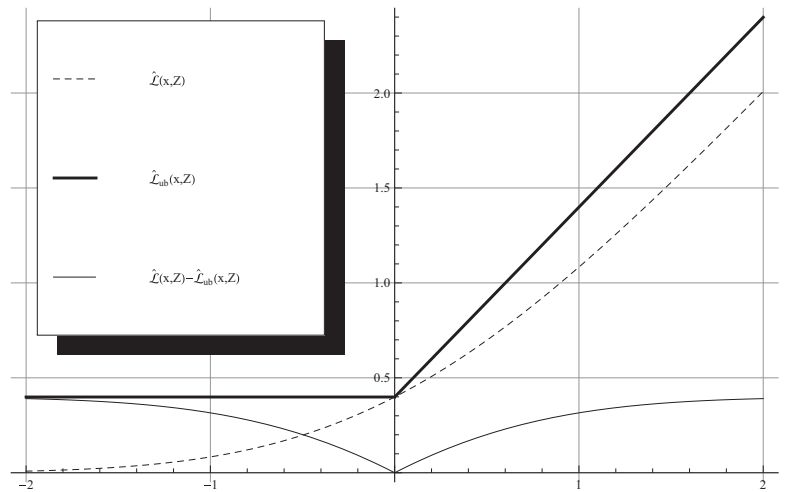
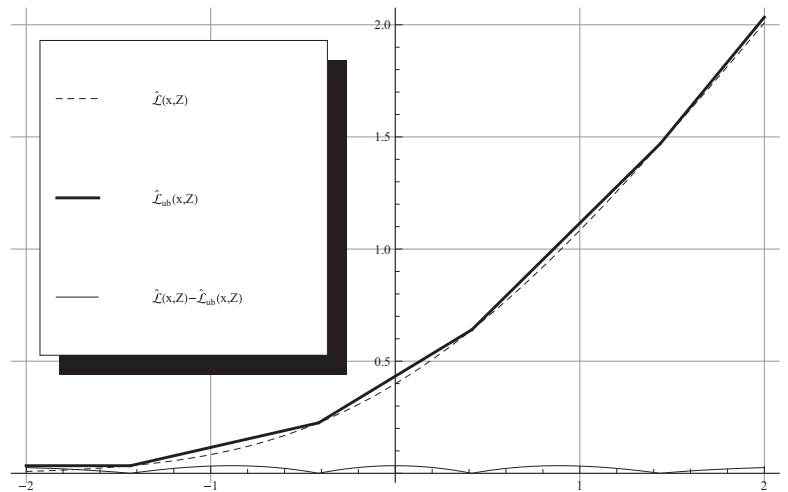**Fig. 5.** Two-segment piecewise linear upper bound for $\widehat{\mathcal{L}}(x, Z)$.

Fig. 6. Five-segment piecewise linear upper bound for $\widehat{\mathcal{L}}(x, Z)$.

We next present a more interesting case, in which the domain has been split into five regions (Fig. 6). Breakpoints are positioned at $x \in \{\pm 1.43535, \pm 0.415223\}$. These were the locations at which the maximum error, i.e. 0.0339052, was observed in Fig. 2. Also in this case, the five-segment piecewise linear upper bound can be obtained by adding to the five-segment piecewise Jensen's lower bound a value equal to its maximum approximation error.

Finally, in Fig. 7, we show an example in which we exploited Lemma 7 to obtain, from the approximation presented in Fig. 6, the five-segment piecewise linear upper bound for $\widehat{\mathcal{L}}(x, \zeta)$, where ζ is a normally distributed random variable with mean $\mu = 20$ and standard deviation $\sigma = 5$. The maximum error is $\sigma 0.0339052$ and it is observed at $x \in \{\pm \infty, \sigma(\pm 0.886942) + \mu, \mu\}$.

In Table 2 we report parameters of $\widehat{\mathcal{L}}_{ub}(x, Z)$ with up to eleven segments.

6. Conclusions

We summarised a number of distribution independent results for the first order loss function and its complementary function. We then focused on symmetric distributions and on normal distributions; for these we discussed ad hoc results. Based on the results discussed, we developed effective piecewise linear approximation strategies based on a minimax framework. More specifically, we developed piecewise linear upper and lower bounds for the first order loss function and its

Table 2
Parameters of $\widehat{\mathcal{L}}_{ub}(x,Z)$ with up to eleven segments.

| Segments	Error	i	0	1	2	3	4	5	6	7	8	9
Piecewise linear approximation parameters												
2	0.398942	b_i	0									
		$\widehat{\mathcal{L}}(x,Z)$	$1/\sqrt{2\pi}$									
3	0.120656	b_i	−0.797885	0.797885								
		$\widehat{\mathcal{L}}(x,Z)$	0.120656	0.918541								
4	0.0578441	b_i	−1.18505	0	1.18505							
		$\widehat{\mathcal{L}}(x,Z)$	0.0578441	0.398942	1.2429							
5	0.0339052	b_i	−1.43535	−0.415223	0.415223	1.43535						
		$\widehat{\mathcal{L}}(x,Z)$	0.0339052	0.225236	0.640459	1.46926						
6	0.0222709	b_i	−1.61805	−0.691424	0	0.691424	1.61805					
		$\widehat{\mathcal{L}}(x,Z)$	0.0222709	0.144966	0.398942	0.83639	1.64032					
7	0.0157461	b_i	−1.7608	−0.896007	−0.281886	0.281886	0.896007	1.7608				
		$\widehat{\mathcal{L}}(x,Z)$	0.0157462	0.101168	0.273745	0.555631	0.997175	1.77654				
8	0.0117218	b_i	−1.87735	−1.05723	−0.493404	0	0.493404	1.05723	1.87735			
		$\widehat{\mathcal{L}}(x,Z)$	0.0117218	0.0746248	0.199839	0.398942	0.693243	1.13185	1.88907			
9	0.00906529	b_i	−1.97547	−1.18953	−0.661552	−0.213587	0.213587	0.661552	1.18953	1.97547		
		$\widehat{\mathcal{L}}(x,Z)$	0.00906529	0.0573174	0.152415	0.301214	0.514801	0.813967	1.24685	1.98453		
10	0.00721992	b_i	−2.05996	−1.30127	−0.8004	−0.384597	0	0.384597	0.8004	1.30127	2.05996	
		$\widehat{\mathcal{L}}(x,Z)$	0.00721992	0.0454051	0.120123	0.23579	0.398942	0.620387	0.920523	1.34668	2.06718	
11	0.00588597	b_i	−2.13399	−1.39768	−0.9182	−0.526575	−0.17199	0.17199	0.526575	0.9182	1.39768	2.13399
		$\widehat{\mathcal{L}}(x,Z)$	0.00588598	0.0368557	0.0971251	0.189721	0.318833	0.490823	0.716296	1.01533	1.43454	2.13987
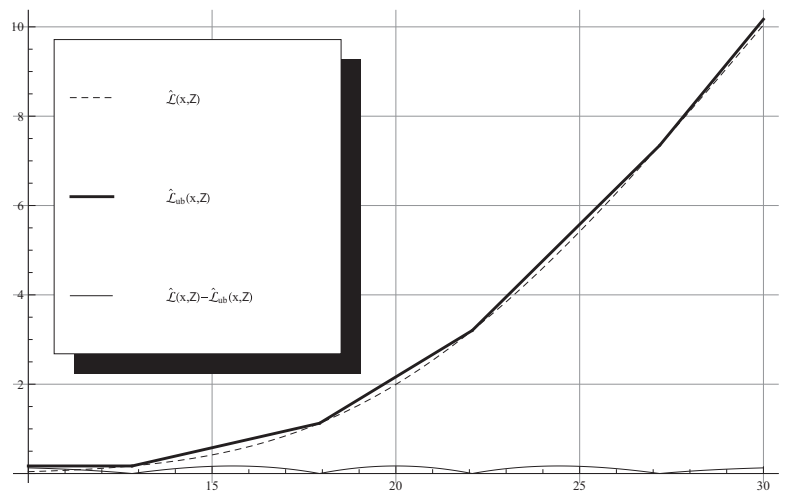
Fig. 7. Five-segment piecewise linear upper bound for $\widehat{\mathcal{L}}(x, \zeta)$, where $\mu = 20$ and $\sigma = 5$.

complementary function. These bounds rely on constant parameters that are independent of the means and standard deviation of the normal distribution considered. We discussed how to compute optimal parameters that minimise the maximum approximation error and we also provided a table with pre-computed optimal parameters for piecewise bound with up to eleven segments. These bounds can be easily embedded in existing MILP models.

Acknowledgements

References

[1] Sven Axsater, Inventory control, Springer Verlag, 2006.
[2] John R. Birge, François Louveaux, Introduction to stochastic programming, corrected ed., Springer Series in Operations Research and Financial Engineering, Springer, 1997.
[3] Wlodzimierz Bryc, A uniform approximation to the right normal tail integral, Appl. Math. Comput. 127 (2–3) (2002) 365–374.
[4] Steven K. De Schrijver, El-Houssaine Aghezzaf, Hendrik Vanmaele, Double precision rational approximation algorithm for the inverse standard normal first order loss function, Appl. Math. Comput. 219 (3) (2012) 1375–1382.
[5] Steven K. De Schrijver, El-Houssaine Aghezzaf, Hendrik Vanmaele, Double precision rational approximation algorithms for the standard normal first and second order loss functions, Appl. Math. Comput. 219 (4) (2012) 2320–2330.
[6] Karl Frauendorfer, Michael Schürle, Stochastic linear programs with recourse and arbitrary multivariate distributions, in: Christodoulos A. Floudas, Panos M. Pardalos (Eds.), Encyclopedia of Optimization, Springer, US, 2001, pp. 2488–2493.
[7] Momčilo M. Gavrilović, Optimal approximation of convex curves by functions which are piecewise linear, J. Math. Anal. Appl. 52 (2) (1975) 260–282.
[8] Alyson Imamoto, Ben Tang, Optimal piecewise linear approximation of convex functions, in: S.I. Ao, Craig Douglas, W.S. Grundfest, Lee Schruben, Jon Burgstone (Eds.), Proceedings of the World Congress on Engineering and Computer Science 2008, WCECS 2008, October 22–24, 2008, IAENG, San Francisco, USA, 2008, pp. 1191–1194.
[9] Peter Kall, Stein W. Wallace, Stochastic programming, Wiley Interscience Series in Systems and Optimization, John Wiley & Sons, 1994.
[10] Jean M. Linhart, Algorithm 885: computing the logarithm of the normal distribution, ACM Trans. Math. Software 35 (3) (2008).
[11] Renata Mansini, Włodzimierz Ogryczak, Maria Grazia Speranza, Conditional value at risk and related linear programming models for portfolio optimization, Ann. Oper. Res. 152 (1) (2007) 227–256.
[12] Haim Shore, Simple approximations for the inverse cumulative function, the density function and the loss integral of the normal distribution, J. R. Stat. Soc. Ser. C (Appl. Stat.) 31 (2) (1982) 108–114.
[13] Edward A. Silver, David F. Pyke, Rein Peterson, Inventory Management and Production Planning and Scheduling, John Wiley & Sons, New York, 1998.
[14] Lawrence V. Snyder, Zuo-Jun M. Shen, Fundamentals of Supply Chain Theory, first ed., Wiley, 2011.
[15] S. Armagan Tarim, Brian G. Kingsman, Modelling and computing (Rn,Sn) policies for inventory systems with non-stationary stochastic demand, Eur. J. Oper. Res. 174 (1) (October 2006) 581–599.
[16] Gary R. Waissi, Donald F. Rossin, A sigmoid approximation of the standard normal integral, Appl. Math. Comput. (1996) 91–95.
[17] Marvin Zelen, Norman C. Severo, Probability functions, in: Milton Abramowitz, Irene A. Stegun (Eds.), Handbook of Mathematical Functions, Applied Mathematics Series, vol. 5, GPO, 1964, pp. 925–995.