Piecewise linearisation of the first order loss function for families of arbitrarily distributed random variables

Citation for published version:

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Preprint (usually an early version)

Publisher Rights Statement:

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Piecewise linearisation of the first order loss function for families of arbitrarily distributed random variables

Roberto Rossi¹ and Eligius M. T. Hendrix²

¹ Business School, University of Edinburgh, Edinburgh, United Kingdom, roberto.rossi@ed.ac.uk
² Department of Computer Architecture, Málaga University, Málaga, Spain eligius.hendrix@wur.nl

Abstract We discuss the problem of computing optimal linearisation parameters for the first order loss function of a family of arbitrarily distributed random variable. We demonstrate that, in contrast to the problem in which parameters must be determined for the loss function of a single random variable, this problem is nonlinear and features several local optima and plateaus. We introduce a simple and yet effective heuristic for determining these parameters and we demonstrate its effectiveness via a numerical analysis carried out on a well known stochastic lot sizing problem.

Keywords: First order loss function, piecewise linear bounds, Jensen, Edmundson-Madanski, lot sizing

1. Introduction

Consider a random variable $\omega$ with expected value $\mu$ and density function $g_\omega$, and a scalar variable $x$. The first order loss function is defined as $L_\omega(x) = E[\max(\omega - x, 0)]$, where $E$ denotes the expected value. The complementary first order loss function is defined as $\bar{L}_\omega(x) = E[\max(x - \omega, 0)]$. Note that $L_\omega(x) = \mu - x + \bar{L}_\omega(x)$. The first order loss function $L_\omega(x)$ and its complementary function $\bar{L}_\omega(x)$ are extensively used in several application domains, such as inventory control [8] and finance (see e.g. [5]).

In general, $L_\omega(x)$ does not admit a closed form and cannot be evaluated without resorting to numerical approximations (see e.g. [1]). $L_\omega(x)$ and its numerical approximations are nonlinear in $x$ and cannot directly be embedded in mixed integer linear programming (MILP) models.

In [7] the authors discuss piecewise linear upper and lower bounds for the first order loss function, which can be immediately embedded in MILP models. These bounds are particularly convenient for a number of reasons: they rely on constant parameters that are independent of the mean and standard deviation of the normal distribution of interest; it is easy to obtain bounds for generic, i.e. non standard, normally distributed random variables via a simple linear transformation. Optimal linearisation parameters are derived following an approach similar to the one discussed in [2, 3], which minimise the maximum approximation error.

In this work, we extend the approach in [7] to the case of generic, i.e. non normal, distributions and we discuss how to embed into a MILP model piecewise linear upper and lower bounds of the first order loss function for a predefined family of generic random variables $\omega_1, \omega_2, \ldots, \omega_N$. These bounds are computed in such a way as to minimise the maximum approximation error over the given family of random variables. We demonstrate the effectiveness of this technique to address a well known stochastic lot sizing problem. Because of the relation that exists between the function $L_\omega(x)$ and its complement $\bar{L}_\omega(x)$, the following discussion will be limited to the complementary first order loss function $\bar{L}_\omega(x)$. 

2. Piecewise linear upper and lower bounds

The complementary first order loss function function \( \bar{L}_\omega(x) \) is convex in \( x \) regardless of the distribution of \( \omega \). For this reason, both Jensen’s lower bound and Edmundson-Madanski upper bound are applicable [4]. Consider a partition of the support \( \Omega \) of \( \omega \) into \( W \) disjoint compact subregions \( \Omega_1, \ldots, \Omega_W \). We define, for all \( i = 1, \ldots, W \)

\[
p_i = \Pr\{\omega \in \Omega_i\} = \int_{\Omega_i} g_\omega(t) \, dt \quad \text{and} \quad \mathbb{E}[\omega|\Omega_i] = \frac{1}{p_i} \int_{\Omega_i} t g_\omega(t) \, dt.
\]

Lemma 1. Let \( \Omega_1, \ldots, \Omega_W \) be a partition of the support \( \Omega \) of random variable \( \omega \), \( p_i \) and \( \mathbb{E}[\omega|\Omega_i] \) defined by (1) and lower bounding function \( \Lambda^i_\omega(x) = x \sum_{k=1}^i p_k - \sum_{k=1}^i p_k \mathbb{E}[\omega|\Omega_k] \). Then lower bounding function

\[
\Lambda_\omega(x) = \max_{i} \left( \max_{i} \Lambda^i_\omega(x), 0 \right) \leq \bar{L}_\omega(x)
\]
is a piecewise linear function with \( W + 1 \) segments.

This lower bound is a direct application of Jensen’s inequality. Let us then consider the maximum approximation error \( e_W = \max_x (\bar{L}_\omega(x) - \Lambda_\omega(x)) \) for the lower bound in Lemma 1 associated with a given partition. A piecewise linear upper bound, i.e. the Edmundson-Madanski bound, is \( \Lambda_\omega(x) + e_W \), which is obtained by shifting up the lower bound in Lemma 1 by a value \( e_W \). These lower and upper bounds for any random variable \( \omega \) can directly be used in an MILP model.

Having established this results, the question is how to partition the support \( \Omega \) in order to obtain good bounds. A number of works discussed how to obtain an optimal partitioning of the support under a framework that minimises the maximum approximation error [2, 3]. In short, these works demonstrate that, in order to minimise the maximum approximation error, one must find parameters ensuring approximation errors at piecewise function breakpoints are all equal. This result unfortunately does not hold when optimal linearisation parameters must be found for complementary first order loss functions of a family of generic random variables.

Consider the complementary first order loss functions for a family of generic random variables \( \omega_1, \ldots, \omega_n, \ldots, \omega_N \). From (1) it is clear that \( \mathbb{E}[\omega|\Omega_i] \) is uniquely determined by the choice of \( p_i \). From this choice of the coefficients \( p_i \) follows for each \( \omega_n \) the function \( \Lambda_\omega(x) \). Within an MILP model, one can select the desired bounding function via a binary selector variable \( y_n \):

\[
\Lambda(x) = \sum_{n=1}^N \max \left( \max_{i} \left( x \sum_{k=1}^i p_k - y_n \sum_{k=1}^i p_k \mathbb{E}[\omega_n|\Omega_k] \right), 0 \right)
\]

adding \( \sum_{n=1}^N y_n = 1 \). These expressions generalise those discussed in [7], which only hold for normally distributed random variables.
The challenge is, of course, to compute an optimal partition of random variable supports into $W$ disjoint compact subregions with probability masses $p_1, \ldots, p_i, \ldots, p_W$. We shall first demonstrate that this is a nonconvex optimisation problem in contrast to computing optimal linearisation parameters for a single loss function. To do so, we consider a simple instance involving two random variables $\omega_1$ and $\omega_2$ with probability density function as shown in Fig. 1a and 1b.

We split the support of $\omega_1$ and $\omega_2$ into five regions with probability mass $p_1, \ldots, p_5$, respectively. In Fig. 2 we plot the maximum approximation error

$$e_W = \max_x \left( \max (\bar{L}_{\omega_1}(x) - \Lambda_{\omega_1}(x)), \max (\bar{L}_{\omega_2}(x) - \Lambda_{\omega_2}(x)) \right),$$

when $p_1$ and $p_4$ are free to vary, $p_2 = 0.3$, $p_3 = 0.1$ and $p_5 = 1 - p_1 - p_2 - p_3 - p_4$; note that this is a standard 2-simplex in $\mathbb{R}^3$ projected in $\mathbb{R}^2$ and coloured to reflect the value of $e_W$. It is clear that this function has a number of local minima. In fact, the function is also constant in some regions, i.e. it has wide plateaus. It is intuitive to observe this, if one considers the fact that the slope of the $i$-th linear segment is given by $\sum_{i=1}^k p_i$, i.e. the slope of segment 1, the $y$ axis represents $p_1 + p_2 + p_3 + p_4$, i.e. the slope of segment 4; darker regions mean lower error.

---

**Figure 2.** Maximum approximation error of our piecewise linear lower bound when $p_1 + p_4 + p_5 \leq 1 - p_2 - p_3$; the $x$ axis represents $p_1$, i.e. the slope of segment 1, the $y$ axis represents $p_1 + p_2 + p_3 + p_4$, i.e. the slope of segment 4; darker regions mean lower error.

**Figure 3.** Optimality gap trend as a function of the number of segments used in the linearisation. The optimality gap is presented as a percentage of the optimal solution obtained with the model embedding a piecewise linear upper bound.

Therefore we developed a metaheuristic to find good, but not necessarily optimal parameter values. The heuristic is a combination of simple random sampling (SRS) and coordinate descent (CD) from the best solution produced by the SRS. For the above example, this strategy produced the following partitioning: $p_1 = 0.24$, $p_2 = 0.18$, $p_3 = 0.215$, $p_4 = 0.175$, $p_5 = 0.19$, with associated maximum approximation error of 0.639. This approximation error amounts to 1.5% of the expected value of $\omega_1$ and to 1.3% of the expected value of $\omega_2$. As we will demonstrate, this strategy produced fairly good outcomes in practical applications.
3. An application to stochastic lot-sizing

We applied the metaheuristic to compute near optimal linearisation parameters for the stochastic lot sizing problem discussed in [6]. We tested the approach over a test bed discussed in [6] comprising 270 instances. In our experiments, in contrast to the original test bed in which demand is normally distributed, demand follows different distributions in different periods: normal, Poisson, exponential and uniform. All instances took just a few seconds to be solved. Note that each of these instances comprises \( N = 15 \) periods in which demand is observed. Consequently, there are \( N(N + 1)/2 \) loss functions in the family for which optimal linearisation parameters must be computed; these parameters must be computed for each instance separately, this is why a lightweight heuristic is desirable. The average optimality gap trend — the difference between the optimal solution of the model embedding our piecewise linear upper bound and that of the model embedding our piecewise linear lower bound — as a function of the number of segments used in the linearisation is shown in Fig. 3. As we see, the gap drops initially with the number of segments and it is well below 1\% of the optimal cost even when just four segments are employed. For higher number of segments the gap fluctuates. This is due to the fact that the simple heuristic here proposed gets stuck in local minima for high dimensional spaces. Future research should therefore investigate more effective approaches to compute optimal parameters for linearisations involving a large number of segments.

4. Conclusions

We have shown that finding optimal linearisation parameters to approximate loss functions when several non-normal random variables are considered, is a challenging global optimisation problem. We discuss how to handle this in a practical setting to generate reasonable results that can be used in MILP models for inventory control.

References


