Proceedings of MAGO 2014, pp. 1 – 4.

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

Roberto Rossi[1] and Eligius M. T. Hendrix[2]

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

In general, $\mathcal{L}_\omega(x)$ does not admit a closed form and cannot be evaluated without resorting to numerical approximations (see e.g. [1]). $\mathcal{L}_\omega(x)$ 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 $\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_N$. 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 $\mathcal{L}_\omega(x)$ and its complement $\bar{\mathcal{L}}_\omega(x)$, the following discussion will be limited to the complementary first order loss function $\bar{\mathcal{L}}_\omega(x)$.

2. Piecewise linear upper and lower bounds

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

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

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

$$\Lambda_\omega(x) = \max\left(\max_i \Lambda^i_\omega(x), 0\right) \leq \bar{\mathcal{L}}_\omega(x)$$

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 $e_W = \max_x(\bar{\mathcal{L}}_\omega(x) - \Lambda_\omega(x))$ 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 $\boldsymbol{\omega}$ can directly be used in an MILP model.

Having established this results, the question is how to partition the support Ω 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.

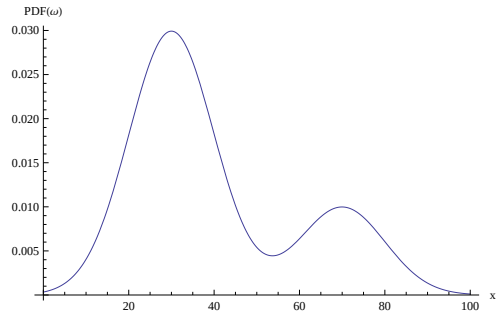

Figure 1a. Probability density function (PDF) of ω_1.

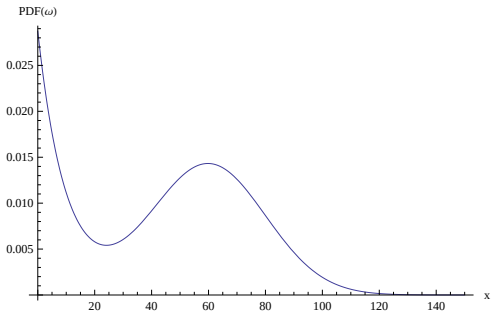

Figure 1b. Probability density function (PDF) of ω_2.

Consider the complementary first order loss functions for a family of generic random variables $\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_n, \ldots, \boldsymbol{\omega}_N$. From (1) it is clear that $\mathrm{E}[\boldsymbol{\omega}|\Omega_i]$ is uniquely determined by the choice of p_i. From this choice of the coefficients p_i follows for each $\boldsymbol{\omega}_n$ the function $\Lambda_\omega(x)$. Within an MILP model, one can select the desired bounding function via a binary selector variable y_n:

$$\Lambda(x) = \sum_{n=1}^{N} \max\left(\max_i \left(x \sum_{k=1}^{i} p_k - y_n \sum_{k=1}^{i} p_k \mathrm{E}[\boldsymbol{\omega}_n|\Omega_k]\right), 0\right)$$

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 $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ with probability density function as shown in Fig. 1a and 1b.

We split the support of $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ into five regions with probability mass $p_1, \ldots, p_5$, respectively. In Fig. 2 we plot the maximum approximation error

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

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

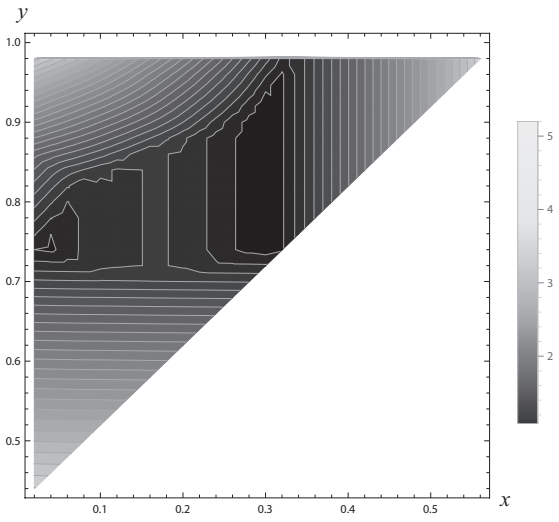

Figure 2. Maximum approximation error of our piecewise linear lower bound when $p_1 + p_4 + p_5 \leq 1 - p_2 - p_3$; the x axis represents p_1, i.e. the slope of segment 1, the y axis represents $p_1 + p_2 + p_3 + p_4$, i.e. the slope of segment 4; darker regions mean lower error.

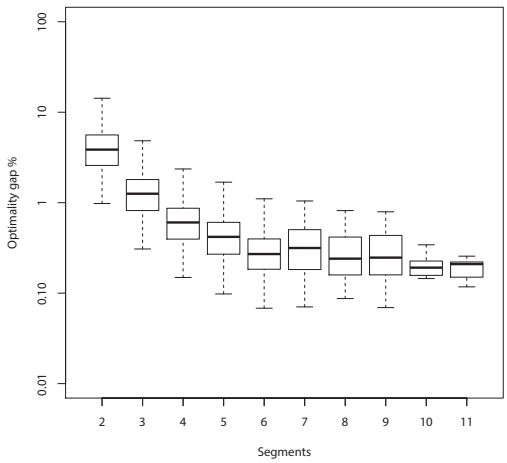

Figure 3. Optimality gap trend as a function of the number of segments used in the linearisation. The optimality gap is presented as a percentage of the optimal solution obtained with the model embedding a piecewise linear upper bound.

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 $\sum_{i=1}^{k} p_i$ and that by varying the slope of one or more segments the maximum approximation error — attained at one or more breakpoints — may easily remain the same. Finding a global optimum of this function is a challenge.

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 $\boldsymbol{\omega}_1$ and to 1.3% of the expected value of $\boldsymbol{\omega}_2$. 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 comput 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

[1] Steven K. De Schrijver, El-Houssaine Aghezzaf, and Hendrik Vanmaele. Double precision rational approximation algorithms for the standard normal first and second order loss functions. Applied Mathematics and Computation, 219(4):2320–2330, November 2012.

[2] Momčilo M. Gavrilović. Optimal approximation of convex curves by functions which are piecewise linear. Journal of Mathematical Analysis and Applications, 52(2):260–282, November 1975.

[3] Alyson Imamoto and Ben Tang. Optimal piecewise linear approximation of convex functions. In S. I. Ao, Craig Douglas, W. S. Grundfest, Lee Schruben, and Jon Burgstone, editors, Proceedings of the World Congress on Engineering and Computer Science 2008 WCECS 2008, October 22 - 24, 2008, San Francisco, USA, pages 1191–1194. IAENG, 2008.

[4] Peter Kall and Stein W. Wallace. Stochastic Programming (Wiley Interscience Series in Systems and Optimization). John Wiley & Sons, August 1994.

[5] Renata Mansini, Włodzimierz Ogryczak, and Maria Grazia Speranza. Conditional value at risk and related linear programming models for portfolio optimization. Annals of Operations Research, 152(1):227–256, July 2007.

[6] Roberto Rossi, Onur A. Kilic, and S. Armagan Tarim. Piecewise linear approximations for the static-dynamic uncertainty strategy in stochastic lot-sizing, December 2013.

[7] Roberto Rossi, S. Armagan Tarim, Steven Prestwich, and Brahim Hnich. Piecewise linear lower and upper bounds for the standard normal first order loss function. Applied Mathematics and Computation, 231:489–502, March 2014.

[8] Edward A. Silver, David F. Pyke, and Rein Peterson. Inventory Management and Production Planning and Scheduling, 3rd Edition. Wiley, 3 edition, January 1998.