

Notes on Distributionally Robust Optimization for the calculation of Ruin Probabilities

Alex Infanger

March 19, 2018

All of the following is based on or directly from [1].

The following assumptions will be made throughout the work.

Assumption 1. (A1) The cost function $c : S \times S \rightarrow \mathbb{R}^+$ is a non-negative lower semicontinuous function satisfying $c(x, y) = 0$ if and only if $x = y$.

Assumption 2. (A2) The function $f : S \rightarrow \mathbb{R}$ whose expectation we are interested in lives in $L^1(d\mu)$ and is upper semicontinuous.

We will take for granted the “DRO duality” proved in class for the compact S case (with discussion over the extension). We restate it here.

Theorem 1. (*Strong Duality for DRO*) Under the assumptions (A1) and (A2), we have that the primal problem

$$\begin{aligned} & \sup \int f d\nu \\ \text{s.t. } & d_c(\mu, \nu) \leq \delta \end{aligned}$$

is equal to the dual problem,

$$\begin{aligned} \inf \quad & \lambda\delta + \int \phi d\mu \\ \text{s.t. } \quad & \lambda \geq 0 \\ & \phi(x) + \lambda c(x, y) \geq f(y) \text{ for all } x, y \end{aligned}$$

(which upon noting the form of the optimizer can be written as a one dimensional problem)

$$\inf_{\lambda \geq 0} \lambda\delta + \mathbf{E}_\mu \left[\sup_{y \in S} (f(y) - \lambda c(X, y)) \right].$$

Along with this strong duality, we have the complementary slackness conditions,

$$\begin{aligned} \lambda (d_c(\mu, \nu) - \delta) &= 0 \\ \phi(x) &= \sup_y (f(y) - \lambda c(x, y)) \quad \pi^* \text{ a.s. } . \end{aligned}$$

The Build-Up of Theorem 3

Our mission now is to simplify the above when we our function $f = \mathbf{1}_A$ is an indicator, so that we can calculate ruin probabilities. Let me extend our definition slightly from class,

$$c(x, A) = \begin{cases} \inf\{c(x, y), y \in A\} & x \notin A \\ 0 & \text{otherwise} \end{cases}.$$

We begin by recognizing that

$$\sup_{y \in S} \{\mathbf{1}_A(y) - \lambda c(x, y)\} = (1 - \lambda c(x, A))^+,$$

which is clear casewise. Notice the left hand side is bounded above by one because $\lambda, c \geq 0$. Then if $x \in A$ choose $y = x$ to achieve that bound. On the other hand if $x \notin A$ the choice $y = x$ gets you zero, and the question is whether we can choose $y \in A$ that is c -close enough to x to have $\mathbf{1}_A(y) - \lambda c(x, y) \geq 0$. hence we have,

$$\sup_{y \in S} \{\mathbf{1}_A(y) - \lambda c(x, y)\} = \max\{1 - \lambda c(x, A), 0\} = (1 - \lambda c(x, A))^+.$$

Potentially, for the sake of time, skip the build up of theorem 3.

Theorem 2. (*Simplified Theorem 3*) Suppose that Assumption (A1) is in force. If A is a nonempty closed subset of the Polish space S , and $\lambda \in [0, \infty)$ attains the infimum in Theorem 1, and the function,

$$h(u) \triangleq \int_{\{x: c(x, A) \leq u\}} c(x, A) d\mu(x)$$

is continuous, we can guarantee

$$\sup \{P(A) : d_c(\mu, P) \leq \delta\} = \mu \left\{ x : c(x, A) \leq \frac{1}{\lambda^*} \right\}.$$

An Example

We work here with the Cramer-Lundberg model for insurance claims. This is a continuous time stochastic process, where the amount of money in the bank $R(t)$ satisfies

$$R(t) = u + ct - \sum_{i=1}^{N_t} X_i.$$

Here u is the initial money in the bank, c is the premium rate, and the X_i are claim sizes. They come in at a rate that is a Poisson process, N_t . You have to assume the X_i are distributed in some way. Let this distribution have first and second moments m_1 and m_2 respectively. We can assume $m_1 < \infty$ else the insurance company would have no point insuring such a risk. Just for the insurance company to make even, the amount of money coming in has to be the amount of money being paid out on average, which corresponds to the choice

$$c = \nu m_1$$

where ν is the rate of the Poisson process. To make money, the firm adds a safety loading – the so-called risk premium – so that,

$$c = (1 + \eta)\nu m_1.$$

Hence our final model is of the form,

$$R(t) = u + (1 + \eta)\nu m_1 t - \sum_{i=1}^{N_t} X_i.$$

We are interested in the probability of ruin

$$\psi(u, T) \triangleq \mathbf{Pr} \left[\inf_{t \in [0, T]} R(t) \leq 0 \right].$$

It turns out that it is computationally intractable to solve the above so we instead work with the following Brownian motion approximation,

$$\begin{aligned} R_B(t) &\triangleq u + (1 + \eta)\nu m_1 t - (\nu m_1 t + \sqrt{\nu m_2} B(t)) \\ &= u + \eta \nu m_1 t - \sqrt{\nu m_2} B(t). \end{aligned}$$

(There may be a typo here, double check if it's ν^2). Our mission is to robustify this estimate by allowing for some non-trivial movement away from the Brownian motion. For this purpose, we identify the Polish space where the stochastic processes of interest live as the *Skorokhod space*,

$$S = D([0, T], \mathbb{R})$$

the space of real valued right-continuous functions with left limits equipped with the J_1 metric.

The J_1 Metric

The J_1 metric is supposed to be like a sup metric that is robust against time shifts that go to zero. Let Λ be the set of strictly increasing and continuous functions $\lambda : [0, T] \rightarrow [0, T]$ such that $\lambda(0) = 0$ and $\lambda(1) = 1$, and let e be the identity map on $[0, T]$. Then we define,

$$d_{J_1}(x_1, x_2) \triangleq \inf_{\lambda \in \Lambda} \{ \|x_1 \circ \lambda - x_2\|_\infty \vee \|\lambda - e\|_\infty \} \quad (1)$$

where \vee is the max function.

Intuitively, this is suppose to capture the sup norm except you're allowed to perturb the function input a litte bit. So it's a bit weaker than the sup norm. In fact, plugging in $\lambda = e$ in Equation 1 tells us that everything is closer together with the J_1 metric compared to the sup metric. Take for example when $T = 1$,

$$x_n = (1 + n^{-1})\mathbf{1}_{\{[0, 2^{-1} + n^{-1})\}}, \quad x = \mathbf{1}_{\{[0, 2^{-1})\}}.$$

Then $\|x_n - x\|_\infty \geq 1$ for all n but $d_{J_1}(x_n, x) \rightarrow 0$ as $n \rightarrow \infty$. To show this notice that,

$$\begin{aligned} d_{J_1}(x_n, x) &= \inf_{\lambda \in \Lambda} \{ \|x_n \circ \lambda - x\|_\infty \vee \|\lambda - e\|_\infty \} \\ &\leq \max \{ \|x_n \circ \lambda_n - x\|_\infty, \|\lambda_n - e\|_\infty \}, \end{aligned}$$

for some feasible choices λ_n . Let us choose

$$\lambda_n = \begin{cases} (1 + 2n^{-1})t & t \in [0, 2^{-1}] \\ (1 - 2n^{-1})t + 2n^{-1} & t \in [2^{-1}, 1]. \end{cases}$$

Notice that the difference between $x_n \circ \lambda_n$ and x sees contribution from the part of the domain $t \in [0, \frac{1}{2})$, so that,

$$\|x_n \circ \lambda_n - x\|_\infty = \frac{1}{n}.$$

On the other hand we have that the identity map and λ differ the most at the point $x = \frac{1}{2}$ where they differ by $\frac{1}{n}$, so that we conclude,

$$\begin{aligned} d_{J_1}(x_1, x_2) &= \inf_{\lambda \in \Lambda} \{ \|x_n \circ \lambda\|_\infty \vee \|\lambda - e\|_\infty \} \\ &\leq \max \{ \|x_n \circ \lambda_n\|_\infty, \|\lambda_n - e\|_\infty \} \\ &\leq \frac{1}{n} \rightarrow 0. \end{aligned}$$

For readability we follow the convention of Billingsley $\lambda t \triangleq \lambda(t)$.

Theorem 3. Noting the above, a good characterization of convergence in J_1 is that, $d_{J_1}(x_n, x) \rightarrow 0$ if and only if you can find some sequence λ_n such that $\lim_n (x_n(\lambda_n(t))) \rightarrow x(t)$ uniformly in t and $\lambda_n t$ converges uniformly to t .

On uniformly continuous functions, Skorokhod convergence is equivalent to uniform convergence. From the triangle inequality,

$$|x_n(t) - x(t)| \leq |x_n(t) - x(\lambda_n t)| + |x(\lambda_n t) - x(t)|,$$

which means if $d_{J_1}(x_n, x) \rightarrow 0$ then the left hand term goes to zero uniformly in t . The right hand term goes to zero uniformly in t if we assume x is uniformly continuous.

[If there is time put in the $j(x)$ example].

Back to the Process of Robustification

We intend to invoke Theorem 2 so that we can show that our robust ruin calculation simplifies to a standard Brownian motion ruin calculation with a different initial money-in-the-bank u . The set of interest is the ruin set,

$$A_u \triangleq \left\{ x \in S : \sup_{t \in [0, T]} (\sqrt{\nu m_2} x(t) - \nu m_1 t) \geq u \right\}.$$

In order to invoke Theorem 2, we must show that the above is closed. Before that we will want the following simple proposition,

Proposition 1. The following inequality holds

$$\sup_{t \in [0, T]} y(t) \leq \sup_{t \in [0, T]} x(t) + d_{J_1}(x, y).$$

Proof. The point is that we want to switch the well known inequality,

$$\left| \sup_t y(t) - \sup_t x(t) \right| \leq \|x - y\|_\infty$$

with the slightly stronger statement,

$$\left| \sup_t y(t) - \sup_t x(t) \right| \leq \|x - \lambda y\|_\infty \text{ for all } \lambda \in \Lambda.$$

The desired conclusion will follow from taking an infimum over λ . Hence we follow a similar path to that of the former inequality. Notice that,

$$\begin{aligned} f(t) &\leq \sup_t f(t) \text{ for all } t \\ g\lambda(t) &\leq \sup_t g(t) \text{ for all } t, \lambda, \end{aligned}$$

from which we conclude,

$$\sup_t (f(t) + g\lambda(t)) \leq \sup_t f(t) + \sup_t g(t).$$

For the choice $f = f - g\lambda$ we achieve,

$$\sup_t f(t) - \sup_t g(t) \leq \sup_t (f - g\lambda), \text{ for all } \lambda.$$

Finally we have that,

$$\sup_t f(t) - \sup_t g(t) \leq \inf_\lambda \|f - g\lambda\|_\infty \leq d_{J_1}(f, g).$$

□

In the Appendix, we consider the more general set

$$A_u = \left\{ x \in D([0, T], \mathbb{R}) : \sup_{t \in [0, 1]} x(t) \geq u \right\},$$

and prove A_u is closed.

Lemma 1. The set $A_u \subseteq S$ is closed in the J_1 metric.

Proof. We will show the compliment is open. Pick any x in the compliment, and let

$$\epsilon \triangleq \frac{(u - \sup_{t \in [0, T]} x(t))}{2} > 0.$$

Now consider some $y \in S$ which is within ϵ of x . We have using proposition 1

$$\sup_{t \in [0, T]} y(t) \leq \sup_{t \in [0, T]} x(t) + d_{J_1}(y, x) < \sup_{t \in [0, T]} x(t) + \epsilon < u.$$

Hence we have found an open ball around x that is contained in A_u . □

Lemma 2. The minimum cost to the set A_u satisfies,

$$c(x, A_u) \triangleq \inf \{c(x, y), y \in A_u\} = \left(u - \sup_{t \in [0, T]} x(t) \right)^p$$

Proof. We will show this for $p = 1$. The result follows. Let $\epsilon > 0$ be smaller than $u - \sup_t x$. Define the shifted version of $x(t)$,

$$z(t) \triangleq x(t) + \inf_t (u - x(t)) + \epsilon,$$

and notice that $z(t) \in A_u$. Then we have that,

$$c(x, A_u) \leq d_{J_1}(x, z) \leq \sup_t |z(t) - x(t)| \stackrel{\text{constant}}{=} \inf_t u - x(t) + \epsilon = u - \sup_t x + \epsilon.$$

Next we show the other side of the inequality. Notice that,

$$c(x, A_u) = \inf_{y \in A_u} \inf_{\lambda \in \Lambda} \{ \|x - y\lambda\|_\infty \vee \|\lambda - e\|_\infty \}.$$

The good news is that $y \in A_u$ implies $y\lambda \in A$ so we can set $\lambda = e$ and find,

$$\begin{aligned} c(x, A_u) &= \inf_{y \in A_u} \sup_t (y - x) \\ &\geq \inf_{y \in A_u} \left(\inf_{t: y > u - \epsilon} y - \sup_t x \right) \\ &\geq u - \sup_t x - \epsilon. \end{aligned}$$

For both bounds ϵ was arbitrary so indeed $c(x, A_u) = u - \sup_t x$.

Now our mission is to extend this for $p \geq 0$. That is, we want to show that,

$$\inf_{y \in A_u} c^p(x, y) = (u - \sup_t x)^p = \left(\inf_{y \in A_u} c(x, y) \right)^p.$$

To this end, let y_i attain the $\inf_{y \in A_u} c(x, y)$ and recognize that,

$$0 \leq \lim_i c(x, y_i) \leq c(x, y) \text{ for all } y \in A_u$$

which immediately implies,

$$0 \leq \left(\lim_i c(x, y_i) \right)^p \leq c^p(x, y) \text{ for all } y \in A_u.$$

Recognizing that $(\lim_i c(x, y_i))^p = \lim_i c^p(x, y_i)$ for $p \geq 1$ we have that

$$0 \leq \lim_i c^p(x, y_i) \leq c^p(x, y) \text{ for all } y \in A_u.$$

Hence we have found a sequence y_i which attains a lower bound for $c^p(x, y)$ for everyone in A_u . We conclude that this $\lim_i c(x, y_i)^p = (\lim_i c(x, y_i))^p = (u - \sup_t x)^p$ is the infimum. \square

Lemma 3. The function,

$$h(s) = \mathbf{E}_\mu [c(X, A); c(x, A) \leq s]$$

is continuous.

Proof. Supposedly it is clear by inspection. \square

With the above statements you can identify $\lambda^* = h^{-1}(\delta) = \inf \{s \geq 0 : h(s) \geq \delta\}$. We can use Theorem 2 to show,

$$\psi_{\text{rob}}(u, T) = \mathbf{Pr} \left[\sup_{t \in [0, T]} (\sqrt{\nu m_2} x(t) - \eta \nu m_1 t) > u - \left(\frac{1}{\lambda^*} \right)^{1/p} \right] = \psi_B(\tilde{u}, T), \quad (2)$$

and so finding the robust ruin probability just settles down to finding a different Brownian motion probability.

Numerical Example

In the numerical example from Jose Blanchet and Karthyek Murthy, they use the Pareto distribution for the claim sizes, with $\alpha = 2.2$. One thing we need to do is to pick δ . Ideally, we pick δ just big enough to capture the distance from the Brownian motion to our process. We approximate the optimal coupling by finding a good coupling between the Brownian motion and the Poisson process. The coupling offered by Khoshnevisan simulates a coupled Brownian motion and Poisson process so that they are very correlated. Hence we seem to assume the cost function is supermodular. With this embedding we can estimate the cost δ with Monte Carlo, and then use this to find λ^* . At this point we generate a new Brownian motion according to Equation 2.

Questions and things still to solve

1. I am worried that I might have implemented that Khosnevisan embedding incorrectly for the following reason: I have to simulate more Brownian motion time steps than Poisson timesteps. This stems from the fact that

$$\sigma(t) \triangleq \inf\{s : A(s) = m_1 t\}.$$

But $A(t) = N(t) + S(t)$ is effectively the sup of a Brownian motion. Because of this it goes like $\sqrt{t} = o(t)$ such that the above equality (or in the computer it is a \geq) exists only for t less than that simulated.

2. How would I implement the large deviations simulation?
3. The uncountable inf of lower semi-continuous is not necessarily lower semicontinuous. How do I show that for our case,

$$h(u) = \int_{\{x: c(x, A) \leq u\}} c(x, A) d\mu(x)$$

is continuous? This is a stochastic integral in our case.

4. I seemed to have disproven the fact that A_u is closed; where is my mistake or is this a typo?
5. It seems to me that the Khoshnevisan embedding is trying hard to correlate the two processes: does this mean we are assuming the cost function is supermodular?

References

- [1] Jose Blanchet, Karthyek R.A. Murthy. “Quantifying Distributional Model Risk via Optimal Transport.” <https://arxiv.org/abs/1604.01446>.