Optimization under Uncertainty and Risk

Michael Orlitzky

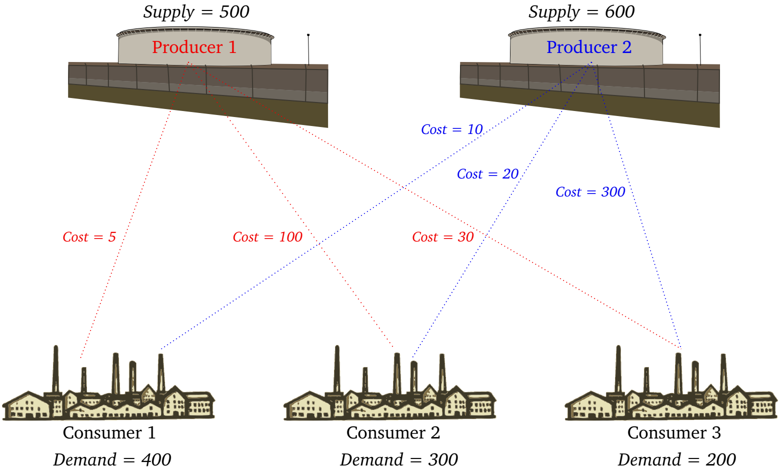
Project overview

The Applied Math Lab is a long-running program of the Towson University Mathematics Department. Under the sponsorship of CSAC, our current project involves the following:

- Describe the transport of chlorine throughout the United States.
- Create models for the risk associated with TIH transportation.
- Develop software systems implementing these models.

This talk presents some of the progress made towards a transportation model.

Transportation problem example



Problem constraints

Our goal is to minimize the total cost of transportation, that is, we would like to minimize the function,

$$z = 5x_{11} + 100x_{12} + 30x_{13} + 10x_{21} + 20x_{22} + 300x_{23}$$

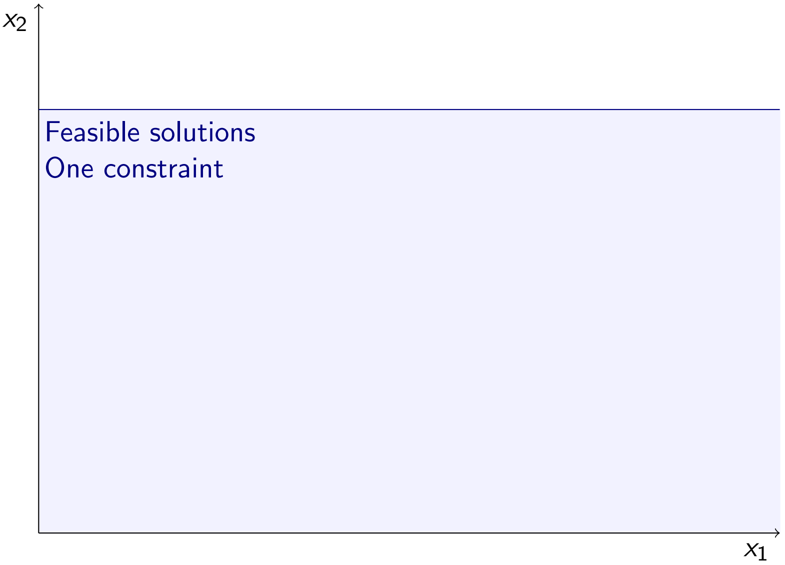
But, we can't over-sell,

$$\begin{aligned} x_{11} + x_{12} + x_{13} &\leq Supply_1 = 500 \\ x_{21} + x_{22} + x_{23} &\leq Supply_2 = 600 \end{aligned}$$

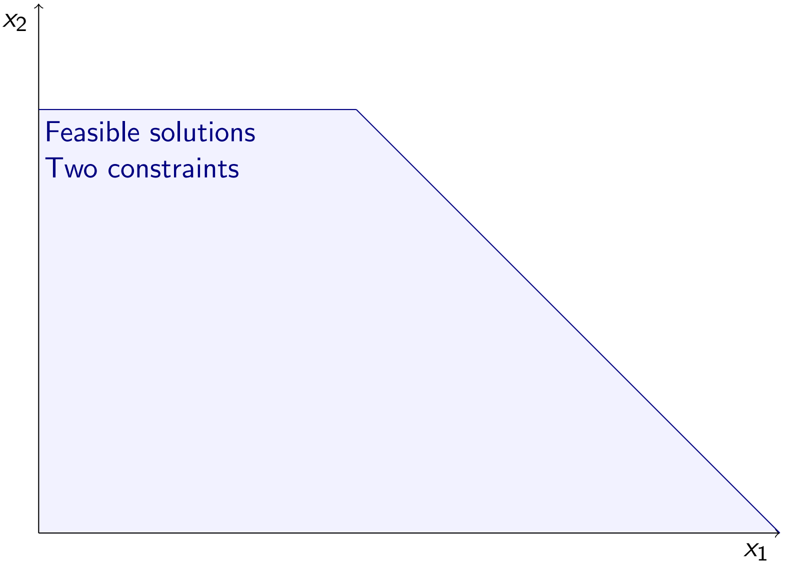
Or over-consume,

$$\begin{aligned} x_{11} + x_{21} &= Demand_1 &= 400 \\ x_{12} + x_{22} &= Demand_2 &= 300 \\ x_{13} + x_{23} &= Demand_3 &= 200 \end{aligned}$$

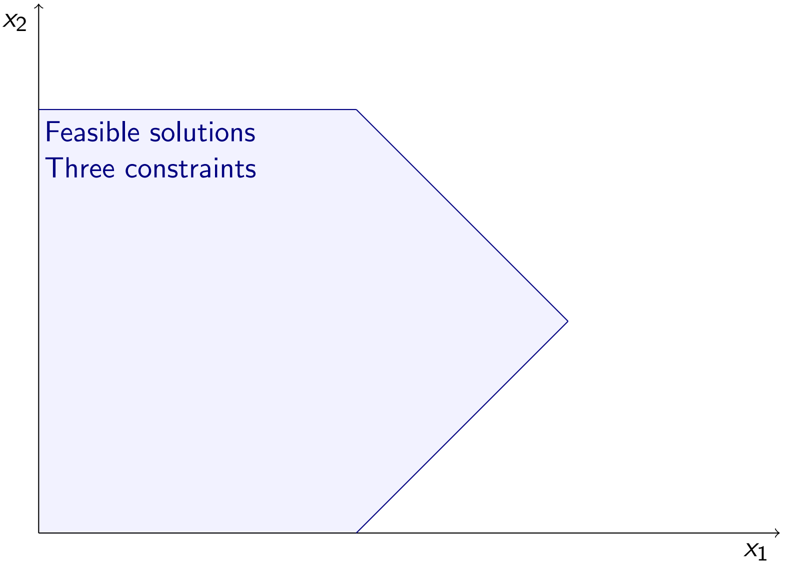
Graphic representation: non-negativity


x_2
Feasible solutions
One constraint
x_1

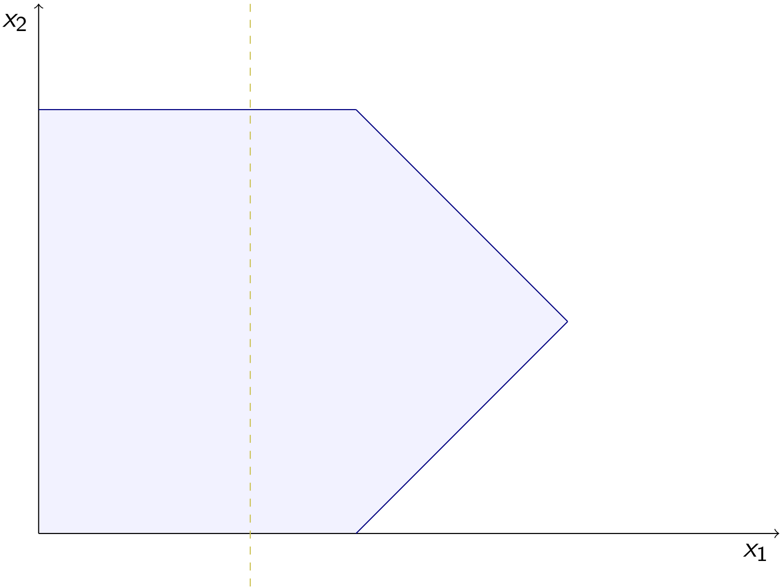
Graphic representation: two constraints



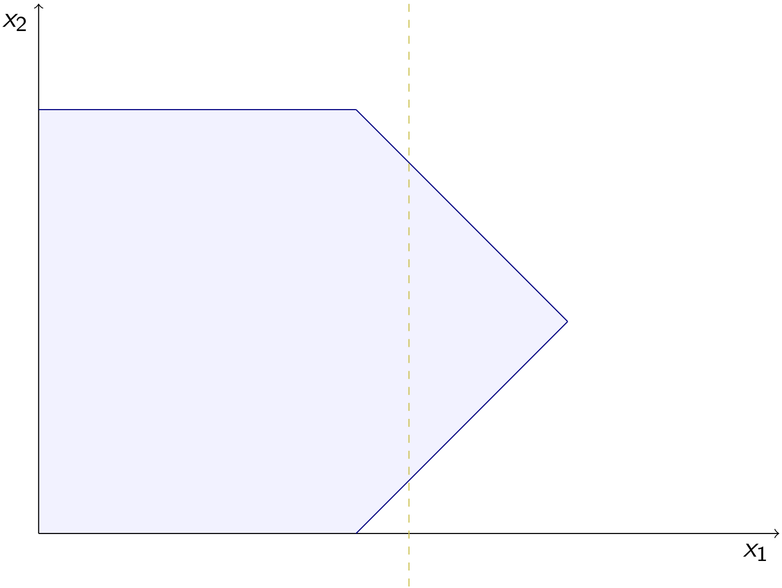
Graphic representation: three constraints



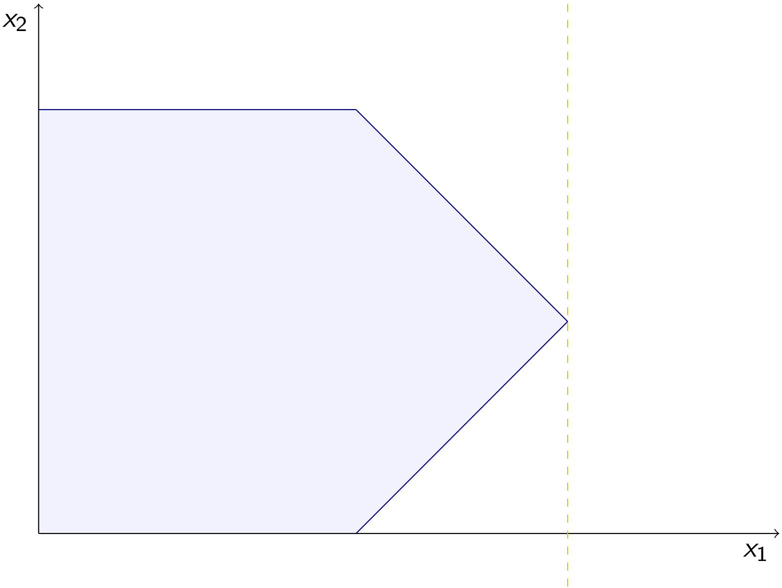
Objective function: good



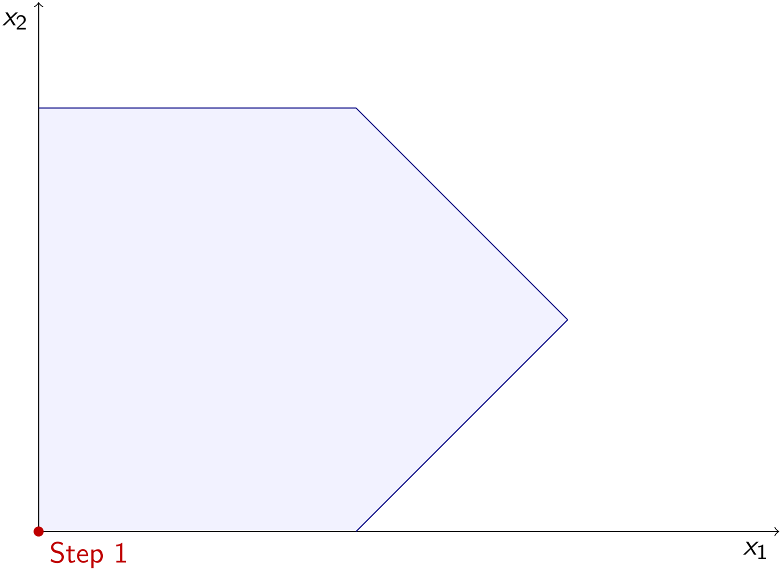
Objective function: better



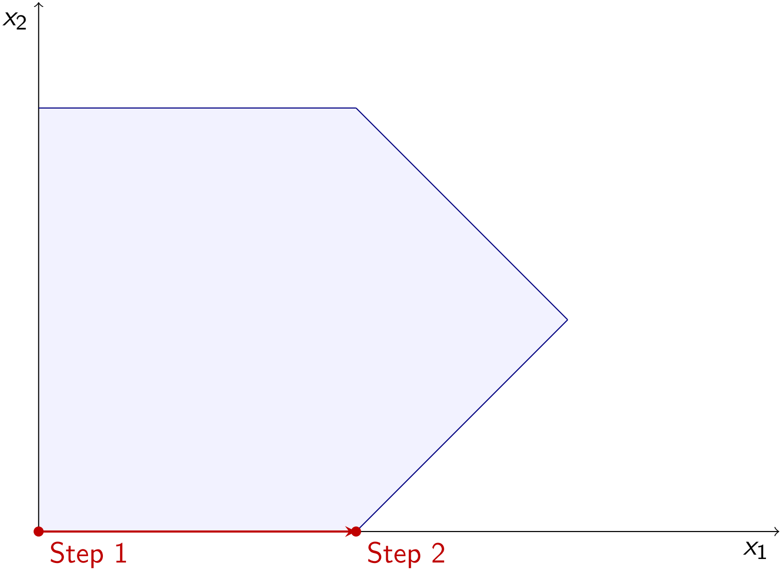
Objective function: best



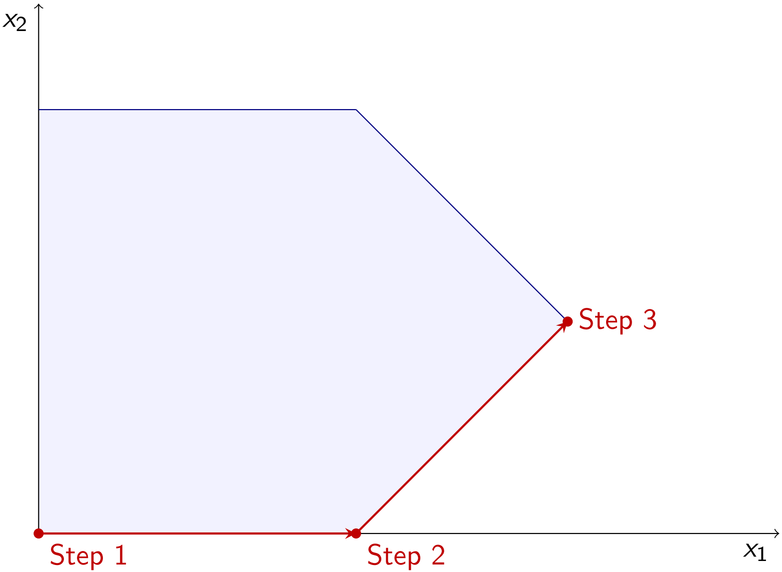
Simplex method: step 1



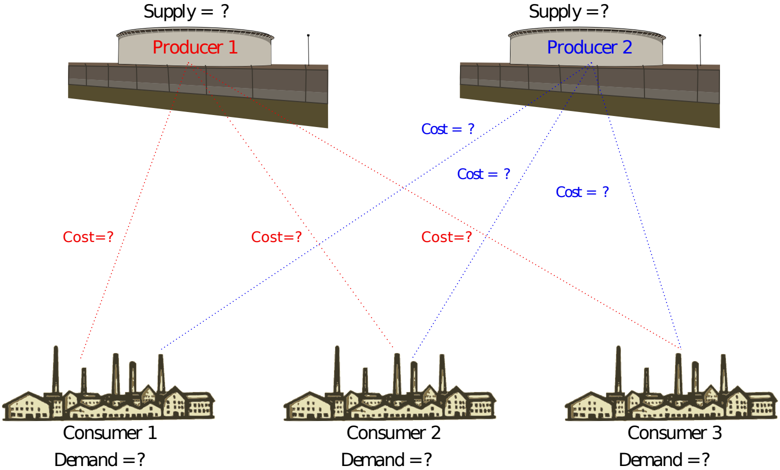
Simplex method: step 2



Simplex method: step 3



What if we don't know our parameters?



Stochastic programming

- Allows uncertainty in both objective function and constraints
- Focuses on problems with two or more distinct phases
- Maximizes the expected value of some objective function

In our problem, there is only one phase. The expected value of the objective function is easy to compute, but we are more interested in "how wrong" we could possibly be.

References

- http://stoprog.org/
- http://tucson.sie.arizona.edu/SPX/tutorial_slides/Philpott.pdf
- http://tucson.sie.arizona.edu/SPX/tutorial_slides/Henrion.pdf

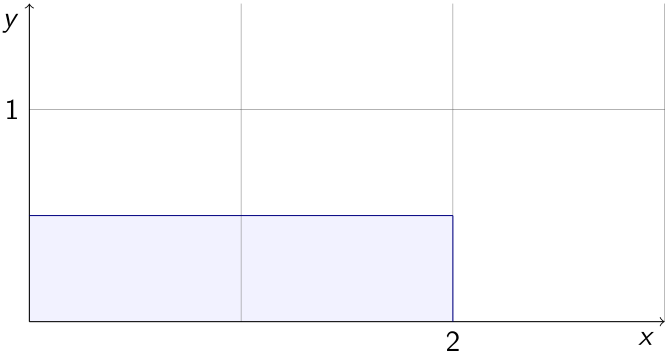
Robust optimization

- Tends to optimize the worst case
- Assumes only a fixed number of coefficients are uncertain
- Uses penalty functions to encourage feasibility
- Can still produce infeasible solutions

References

- http://www.bschool.nus.edu/staff/dscsimm/docs/ROThesis.pdf
- http://robust.moshe-online.com/
- http://hostdb.ece.utexas.edu/~cmcaram/pubs/RobustOptimizationPaper.pdf

We're ready to get rid of the known costs. We can replace the costs c_{ij} by random uniform variables $\mathcal{C}_{ij}$.



With uniform random variables, all possible values are equally likely.

New objective function

Our new objective is to minimize the function,

$$z = \mathcal{C}_{11}x_{11} + \mathcal{C}_{12}x_{12} + \mathcal{C}_{13}x_{13} + \mathcal{C}_{21}x_{21} + \mathcal{C}_{22}x_{22} + \mathcal{C}_{23}x_{23}$$

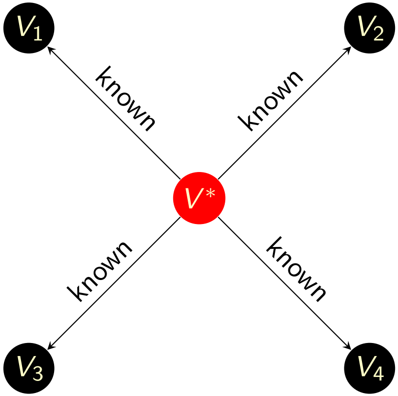
where the $\mathcal{C}_{ij}$ are the uniform variables mentioned previously. Let $\tilde{c}_{ij}$ denote the mean of the random variable distributed by $\mathcal{C}_{ij}$. Note that $\tilde{c}_{ij}$ is just a constant, so we can solve the following problem by the usual process.

$$\tilde{z} = \tilde{c}_{11}x_{11} + \tilde{c}_{12}x_{12} + \tilde{c}_{13}x_{13} + \tilde{c}_{21}x_{21} + \tilde{c}_{22}x_{22} + \tilde{c}_{23}x_{23}$$

We'll call the optimal solution to this problem V^*

Starting vertex V^*

We have a set of cost coefficients for which V^* is optimal, so we can compute the probability distributions for the adjacent vertices.



Let's take for an example the following modification of our original two-supplier three-consumer model.

$$
\begin{aligned}
c_{11} &\rightarrow U_{11} \sim \mathcal{U}_{11}(0,10) \\
c_{12} &\rightarrow U_{12} \sim \mathcal{U}_{12}(75,125) \\
c_{13} &\rightarrow U_{13} \sim \mathcal{U}_{13}(20,40) \\
c_{21} &\rightarrow U_{21} \sim \mathcal{U}_{21}(0,20) \\
c_{22} &\rightarrow U_{22} \sim \mathcal{U}_{22}(10,30) \\
c_{23} &\rightarrow U_{23} \sim \mathcal{U}_{23}(200,400)
\end{aligned}
$$

Note that the means of these distributions equal the cost coefficients from the original problem (this is just for convenience).

Optimal at the mean

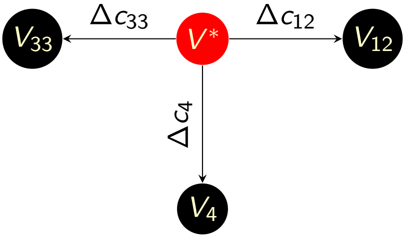
The optimal solution for the mean costs is just the optimal for our original problem:

$$V^* = (300, 0, 200, 100, 300, 0)$$

There are three directions we can move from V^*:

$$\begin{aligned}
\Delta V_{12} &= (-1, 1, 0, 1, -1, 0, 0, 0) \\
\Delta V_{33} &= (1, 0, -1, -1, 0, 1, 0, 0) \\
\Delta V_{4} &= (-1, 0, 0, 1, 0, 0, 1, -1)
\end{aligned}$$

A (non-negative!) cost is associated with each direction. We can express these costs in terms of the coefficients which are random variables.



$$\begin{aligned}
\Delta c_{12} &= U_{21}(0,20) - U_{11}(0,10) + U_{12}(75,125) - U_{22}(10,30) \\
\Delta c_{33} &= U_{11}(0,10) - U_{13}(20,40) + U_{23}(200,400) - U_{21}(0,20) \\
\Delta c_4 &= U_{21}(0,20) - U_{11}(0,10)
\end{aligned}$$

Movement probability

The probability that we move in a direction is simply the probability that the cost in that direction is negative. We'll do the easy direction (actually, its complement).

$$
\begin{aligned}
P\left(U_{11} + \tilde{U}_{21} \leq 0\right) &= \int_{-\infty}^{\infty} F_{U_{11}}(-y)\, \tilde{u}_{21}(y)\ dy \\
&= \int_{-\infty}^{\infty} F_{U_{11}}(y)\, u_{21}(y)\ dy \\
&= \int_{0}^{10} F_{U_{11}}(y)\, u_{21}(y)\ dy + \int_{10}^{20} F_{U_{11}}(y)\, u_{21}(y)\ dy \\
&= \int_{0}^{10} \frac{x}{10} \frac{1}{20}\ dy + \int_{10}^{20} 1 \cdot \frac{1}{20}\ dy \\
&= \frac{1}{200} \frac{100}{2} + \frac{1}{20}\left[20 - 10\right] \\
&= \frac{100}{400} + \frac{10}{20} \\
&= \frac{3}{4}
\end{aligned}
$$

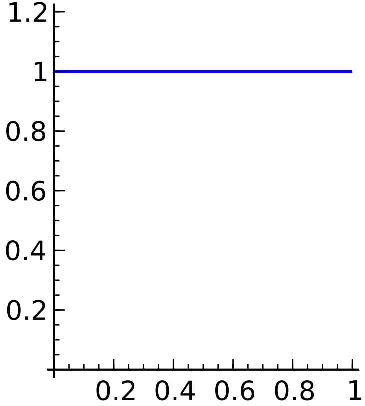
We can test the result by running 100,000 trials (samples) using our random distributions. The resulting vertex frequencies are shown below.

```
linear_programs $ ./lp_solve trials-uniform-small
Solution Vector : Count
----------------------
[100.0, 0.0, 200.0, 300.0, 300.0, 0.0] : 24974
[300.0, 0.0, 200.0, 100.0, 300.0, 0.0] : 75026
```

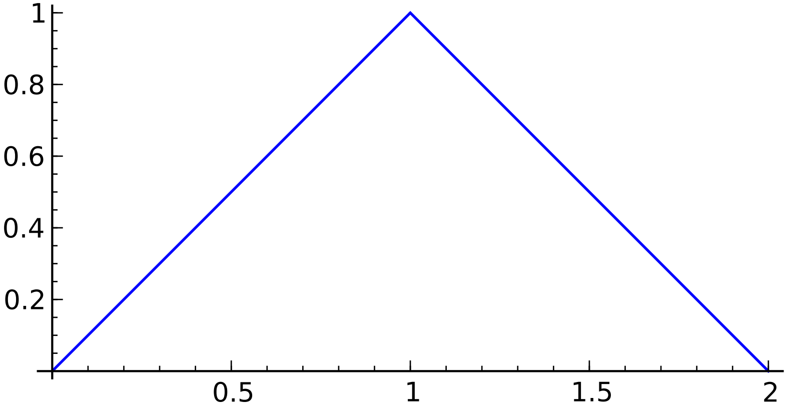
As expected, 25% of the time, we move to V_4. The other vertices are missing because we move to them with probability zero.

Probability computation: one variable

Each of the probability calculations involve determining whether or not the sum of random variables is negative. Here's one variable (easy).

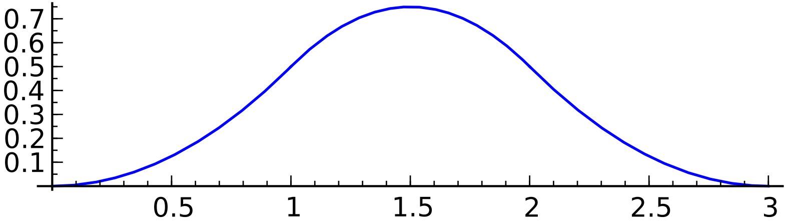


In two variables, we're still piecewise linear so the probabilities can be calculated quickly.



Probability computation: three variables

With three or more variables, we're in trouble. The graph below changes for every combination of costs.



Risk incorporation

Given (uncertain) information about the costs along a transportation network, we study the distribution of material along that network. However, monetary costs are not the only costs we consider.

$$\textbf{cost} = \textbf{money}\ (\textit{uncertain}) + \textbf{risk}\ (\textit{uncertain})$$

The *risk* along each route is also a “cost” that may be known with some uncertainty. If we incorporate risk into the model, we can analyze the resulting distribution to find the risk’s effect on the network.

Future work

- Are these computations fast enough to do on a large scale?
  - Caching
  - Numerical integration
  - Monte Carlo methods
- If so, how can we optimize and extend the model for non-uniform distributions?
- If not, do we really need to do that many of them? How far does a typical problem deviate from V^*?
- Is this the best method to obtain this information? Is it worth the effort?

Thanks and acknowledgements

Chemical Security Analysis Center

- Dr. George Famini
- Dr. Shannon Fox
- Rachel Gooding

Towson University

- Fisher College of Science and Mathematics
- Drs. Kolesnikov, Borodachov, and Kumchev
- Jonathan Howe
- Dennis Howell
- Sara Miller
- Derek Morgan
- Patrick O'Neill
- Solomon Stein
- Samantha Stevenson
- Alana Thompson