

Modeling Uncertainty in Optimization Problems

Andy Philpott*

June 7, 2010

1 Introduction

This expository article discusses approaches for modeling optimization problems that involve uncertainty. The emphasis of the paper is on motivation and intuition rather than technical completeness. By virtue of its length, the paper is inevitably incomplete, and the reader is advised to look at other papers in this encyclopedia to explore the richness of modeling approaches that this paper can only hint at.

How to model real decision problems using optimization models depends on the exact form of the problem being faced, and so it is difficult to give a general description of modeling techniques without some context in which to view the techniques. For this reason, we have chosen to illustrate the modeling approaches we discuss by placing them in the context of optimization of some models of electricity generation.

Optimization problems that involve uncertainty fall into several broad classes that admit different modeling approaches. A good starting point is to consider any system with random effects that can be simulated. As an example consider a simple unit commitment problem in which the owner of a thermal generation unit and a wind farm must decide ahead of time whether to start up their thermal unit to meet a known future demand. If the wind blows enough then there will be sufficient power to meet the load, and so the cost incurred by the startup will be unnecessary. On the other hand, if the wind does not blow, and the unit is not started then the owner will incur some penalty of not supplying the demand. The decision in this case is a simple binary choice, and our model should enable the decision maker to make the correct choice.

The natural question to ask at this point is “how do we determine the correct choice?” Since there are only two alternatives, one can investigate all the implications of making each choice and compare them. We suppose here, and throughout the rest of the paper that the best choice would be clear if there were no uncertainty in the model. In other words, if we

*Department of Engineering Science, University of Auckland, Auckland, New Zealand, e-mail: a.philpott@auckland.ac.nz.

had a perfect forecast of wind then our choice would be to start the thermal plant only if the cost of doing so were smaller than the cost of shortage that the startup avoids. (Of course there are settings in which we might consider other criteria such as environmental impact, in which case we seek the solution to a multi-criteria optimization problem, but we shall not be concerned with this aspect in this paper, and the reader should consult e.g. [9].)

If a perfect forecast of wind is not available, then one is faced with the problem of determining the best choice under uncertainty. A good question to ask at this point is “can we construct or estimate a probability distribution for uncertain wind events?” for example from historical data and/or meteorological models. If so, then we say that the uncertainty is *probabilistic*. (It is well-known that economists like to distinguish between *risk* and *uncertainty*, where risk involves events having probability distributions, and uncertainty a situation for which we have no previous data or experience so that it is impossible to model statistically, but since we shall use the word “risk” in a different sense, we call these probabilistic and non-probabilistic uncertainty.)

Let us first consider the case of probabilistic uncertainty. The distribution of this might take the form of some well-known distribution from a parametric family (such as a lognormal distribution for example) or it might be empirically determined (for example from a Monte Carlo simulation). The existence of such a distribution means in principle that the probability distribution of total cost of either switching on the unit, or keeping it off can be computed. Armed with these distributions the decision maker can express a preference for one to determine the best decision. The message here is that optimization under uncertainty with known probability distributions is about choosing between a (possibly infinite) set of probability distributions of total cost. Furthermore, as we shall discuss in this article, the existence of probability distributions allows us to use probability theory to explore mathematically different ways of making this choice. We shall discuss this more fully in section 2.

How does one handle non-probabilistic uncertainty? On one hand the decision-maker might be faced with a collection of data points with insufficient a priori knowledge to confidently determine a probability distribution from which these data are drawn. In our unit commitment example, we may have only 10 observations of previous wind outcomes. In this context the modeling choices become less obvious. On one hand, the collected data could be viewed as an empirical discrete distribution. In our example this would give rise to 10 wind outcomes (or *scenarios*) each with probability 0.1. Even if they were not evenly weighted, at least we would want to investigate the candidate decisions in each of the 10 scenarios to see what the cost would have been in these scenarios. The decision maker can then choose the vector of 10 outcomes that is preferred. How to do this soundly is less obvious when there is no probability distribution, but there is a rich literature on making these choices according to different criteria. We shall return to this in section 3 of this paper.

The final form of uncertainty the decision maker might be confronting is that for which there is no probability distribution, or reliable data, or even a coherent understanding. Uncertain events might occur that are not even foreseen - the so-called black swans (the probability of a swan being white was 1 until black swans were discovered in Western Aus-

tralia) that have been popularized in the recent book by Nicolas Taleb [17] that warns of the dangers of quantitative modeling in finance. However faced with such comprehensive uncertainty, there is little one can do from an Operations Research perspective. This does not mean that one should be entirely solipsistic - stress testing candidate decisions under some outlandish scenarios can yield valuable insights even those these might be supposed to occur with miniscule probability. Nevertheless the black swan will by definition always elude the analyst.

The remainder of this paper is laid out as follows. In the next section we will give an overview of probabilistic modeling in optimization problems. Section 3 will then deal with the treatment of uncertainty when probability distributions are not available. Finally in section 4 we will provide some brief bibliographical references.

2 Optimization modeling with probability

Probabilistic optimization models take advantage of the fact that probability distributions governing the data are known or can be estimated. In this section we will discuss how to model these with reference to problem examples drawn from electricity generation.

To begin the discussion we recall the unit commitment problem discussed in the Introduction, and formulate this mathematically. For simplicity, assume that there is a single hour of demand d MW to be met. Suppose that the thermal unit has a capacity g MW, a startup cost of $\$h$, a running cost of $\$c/\text{MWh}$, and the supply of wind power is the random variable $S(\omega)$, where $\omega \in \Omega$ denotes a random wind event for the hour in question. Finally suppose that any shortages of electricity cost $\$s/\text{MWh}$. Let

$$z = \begin{cases} 1, & \text{if the unit is started} \\ 0, & \text{if the unit is not started} \end{cases}$$

and suppose $X(\omega)$ is the (random) power generated by the unit. Observe that in any outcome ω the optimal choice of $X(\omega)$ is

$$X(\omega) = z \min\{g, \max(0, d - S(\omega))\}.$$

Then the cost of decision z in outcome ω is

$$C(z, \omega) = hz + cz \min\{g, \max(0, d - S(\omega))\} + s \max(0, d - S(\omega) - gz).$$

As discussed in the Introduction, a decision maker can compare $C(0, \omega)$ and $C(1, \omega)$ for every realization of ω and choose which decision they prefer. As an illustration these functions are plotted in Figure 1 for $h = 20$, $c = 1$, $g = 5$, $d = 6$, and $s = 10$ over the range $S(\omega)$ from 0 to 10.

They show that not committing the unit is a costly decision when wind generation is low, but is inexpensive when wind becomes plentiful. Conversely committing the unit incurs an expense but saves on shortage cost, except when $S(\omega)$ is less than 1. From these plots alone it is not clear which is the better decision.

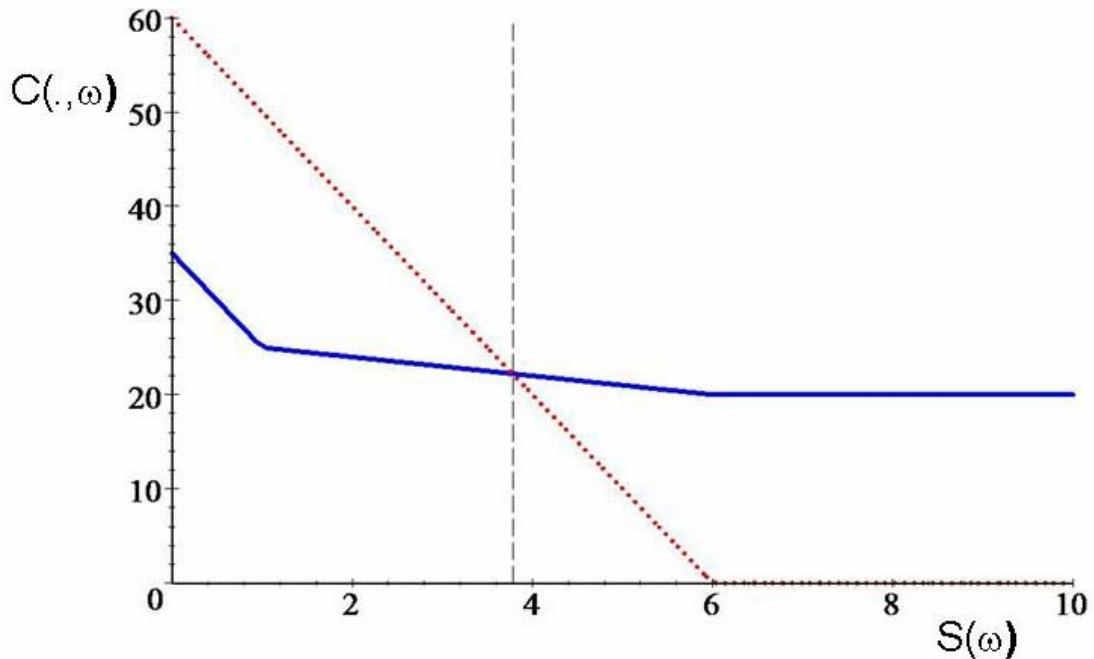


Figure 1: Plot of $C(0, \omega)$ (dotted) and $C(1, \omega)$ (solid) for unit commitment example.

If a probability distribution is known for wind outcomes ω , then it is possible to analyse the choices in more detail. For example if there is zero probability of observing wind generation less than $\frac{34}{9}$ MWh, then it is less costly to leave the unit off, as this policy is less costly with probability 1. This approach can be explored further by attempting to order the probability distributions that emerge from different decisions. The theory of *stochastic dominance* gives a methodology for doing this. A random variable C_1 is said to *stochastically dominate* random variable C_2 to *first order* if for every x , $\Pr(C_1 \leq x) \leq \Pr(C_2 \leq x)$. When minimizing cost we might seek a policy that gives rise to a (random) objective function that is dominated by those of every alternative policy.

In the unit commitment problem suppose that $S(\omega)$ has a uniform distribution on $[0, 10]$. Then $C(0, \omega)$ has distribution function

$$\Pr(C(0, \omega) \leq x) = \begin{cases} 0, & x \leq 0 \\ 0.4 + 0.01x, & 0 < x \leq 60 \\ 1, & 60 < x \end{cases}$$

and $C(1, \omega)$ has distribution function

$$\Pr(C(1, \omega) \leq x) = \begin{cases} 0, & x \leq 20 \\ -1.6 + 0.1x, & 20 < x \leq 25 \\ 0.65 + 0.01x, & 25 < x \leq 35 \\ 1, & 35 < x \end{cases}$$

as shown in Figure 2. It is easy to see that neither $C(0, \omega)$ nor $C(1, \omega)$ is first-order stochas-

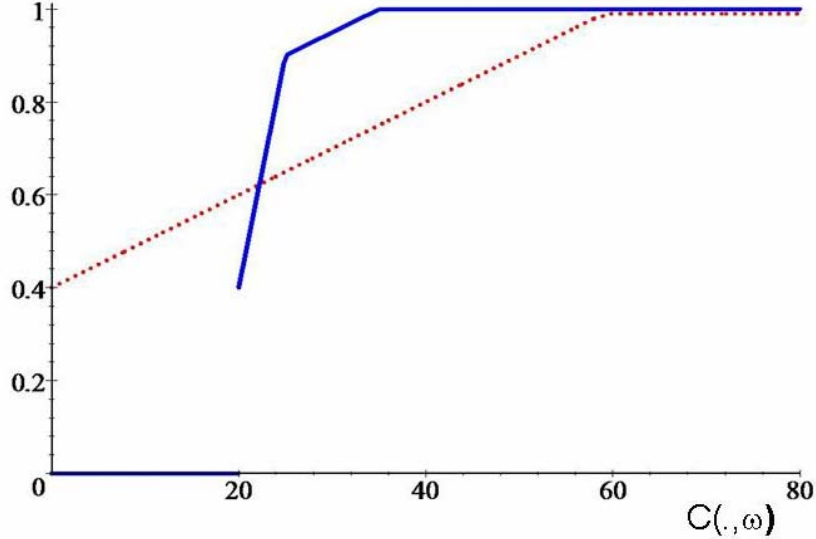


Figure 2: Cumulative distribution functions of $C(0, \omega)$ (dotted) and $C(1, \omega)$ (solid).

tically dominant. This example illustrates that this is a very strong form of optimality, and often precludes finding a suitable solution. (Second- and higher-order stochastic dominance relations can be defined that admit more possible solutions; we shall not discuss these here but refer the reader to [15].)

As an alternative to stochastic dominance, we might constrain the policies that we are prepared to accept to those that have a high cost only with low probability. This is an example of a *chance constraint*, which takes the form

$$\Pr(C(z, \omega) \leq \tau) \geq 1 - \alpha$$

For example, if the probability distribution of wind outcomes $S(\omega)$ is uniform on $[0, 10]$, then we could constrain z to satisfy

$$\Pr(C(z, \omega) \leq 30) \geq 0.9.$$

Since $\Pr(C(0, \omega) > 30) = 0.3$, and $\Pr(C(1, \omega) > 30) = 0.05$, this constraint is satisfied by $z = 1$, but not $z = 0$.

The choice of a policy becomes more obvious in settings in which decisions are made repeatedly in essentially the same circumstances, and the objective is to come up with a decision that will perform well on average. In our unit commitment example this would be the case if the commitment decision was being repeated every day or every few hours. In this setting the goal is to find some policy that minimizes the expectation $\mathbb{E}[C(z, \omega)]$ of $C(z, \omega)$. This objective is justified by the law of large numbers for infinitely repeated trials which implies that a policy that minimizes expected costs will result in the minimum long-run average cost. In the example with a uniform distribution of wind outcomes, $\mathbb{E}[C(1, \omega)] = 22.25$ and $\mathbb{E}[C(0, \omega)] = 18$, and so the minimum expected cost policy is to keep the unit off.

The mathematical expectation operator is an example of a *risk measure*, a function $\rho(Z)$ from a space of random variables to \mathbb{R} . It assigns a score to any random variable Z that can then be ranked against the scores of others. In the example $Z_0 = C(0, \omega)$, and $Z_1 = C(1, \omega)$, and $\rho(Z_0) = 18$ is preferable to $\rho(Z_1) = 22.25$. When modeling optimization problems it is often crucial that the risk measure be convex and monotone, so that efficient numerical procedures can be applied to solve the problem. More generally a risk measure is called *coherent* if it has the following properties.

- (i) It is convex, i.e., if Z_1 and Z_2 are two random variables and $t \in [0, 1]$, then

$$\rho(tZ_1 + (1-t)Z_2) \leq t\rho(Z_1) + (1-t)\rho(Z_2).$$

- (ii) It is monotone, i.e., if Z_1 and Z_2 are two random variables such that with probability one $Z_2 \geq Z_1$, then $\rho(Z_2) \geq \rho(Z_1)$.
- (iii) It is positively homogeneous, i.e., if Z is a random variable and $t > 0$, then $\rho(tZ) = t\rho(Z)$.
- (iv) It has the (translation equivariance) property: $\rho(Z + a) = \rho(Z) + a$ for any $a \in \mathbb{R}$.

Properties (i) and (ii) ensure that if $C(\cdot, \omega)$ is convex for almost every ω , then the function $f(z) := \rho(C(z, \omega))$ is also convex.

As mentioned above, the expected cost is an appropriate risk measure for an optimization problem that is solved repeatedly as variations of cost around the expected cost will balance out in the long run. For a solution that is to be used only once, we might seek to control this random variation, especially if it produces high costs. One way of doing this is by adding a chance constraint of the form

$$\Pr(C(z, \omega) \leq \tau) \geq 1 - \alpha. \tag{2.1}$$

Unfortunately this is not convex in general, and so does not correspond to a coherent risk measure.

The constraint (2.1) can be approximated by

$$\text{CVaR}_{1-\alpha}(C(z, \omega)) \leq \tau$$

where CVaR is called the *conditional value at risk*, defined to be

$$\text{CVaR}_{1-\alpha}(Z) := \inf_{t \in \mathbb{R}} \{t + \alpha^{-1} \mathbb{E}[\max\{0, Z - t\}]\}$$

It can be shown that the coherent measure CVaR is the tightest safe convex approximation to (2.1). In the example, the CVaR of each unit commitment decision at level 0.95 is respectively

$$\begin{aligned} \text{CVaR}_{0.95}(C(0, \omega)) &= \inf_{t \in \mathbb{R}} \{t + 20 \mathbb{E}[\max\{0, C(0, \omega) - t\}]\} = \inf_{t \geq 0} \{-11t + 360 + \frac{1}{10}t^2\} = \frac{115}{2}, \\ \text{CVaR}_{0.95}(C(1, \omega)) &= \inf_{t \in \mathbb{R}} \{t + 20 \mathbb{E}[\max\{0, C(1, \omega) - t\}]\} = \inf_{t \geq 25} \{-6t + \frac{245}{2} + \frac{1}{10}t^2\} = \frac{65}{2}, \end{aligned}$$

so with this risk measure there is less risk in running the unit. Observe in this example that the underlying (discrete) optimization problem is not convex, and so the use of a coherent risk measure here does not retain convexity as it would if applied to a linear program.

Finally we mention the classical approach to modeling risk that can be traced back to Bernoulli [4] that uses utility and disutility functions as risk measures. (Here we speak of disutility when minimizing and utility when maximizing). In a minimization problem like our unit commitment problem, the risk measure $\rho(Z)$ is defined to be $\mathbb{E}[u(Z)]$ for some disutility function u . Typically utility functions are chosen to be concave and disutility functions are convex. This represents aversion to risk on the part of decision makers. To illustrate this, consider the unit commitment example with the disutility function $u(z) = z^2$. Then

$$\mathbb{E}[u(C(0, \omega))] = 720, \quad \mathbb{E}[u(C(1, \omega))] = 505,$$

so with this risk measure the decision maker would prefer to run the unit.

The unit commitment example we have been studying involves choosing from two possible decisions, to run the thermal unit or not to run it. Implicit in each of these choices is an operating level $X(\omega)$ for the plant under each wind realization. Because of its cost structure, our model can assume this to be the minimum level required to meet demand minus wind generation if this does not exceed the capacity of the unit.

In many circumstances this implicit optimization is not possible, and for each unit commitment decision we must solve a dispatch problem to determine the optimal level at which to run each plant. This leads to the widely studied *two-stage stochastic program with recourse*, where the first stage of the problem determines the unit commitment and the second stage, or recourse problem, determines the optimal generation level in each scenario given this unit commitment decision.

In general the classical two-stage linear stochastic programming problem can be formulated as

$$\min_{x \in X} \{c^\top x + \mathbb{E}[Q(x, \xi)]\}, \tag{2.2}$$

where $Q(x, \xi)$ is the optimal value of the second-stage problem

$$\min_y q^\top y \quad \text{subject to } Tx + Wy \leq h. \tag{2.3}$$

Here x is the first-stage decision vector that must lie in some set X , $y \in \mathbb{R}^m$ is the second-stage decision vector and $\xi = (q, T, W, h)$ contains the data of the second-stage problem. In this formulation, at the first stage we have to make a “here-and-now” decision x before the realization of the uncertain data ξ , viewed as a random vector, is known. At the second stage, after a realization of ξ becomes available, we optimize our behavior by solving an appropriate optimization problem.

An important practical question is whether the formulated problem can be solved numerically. In that respect the standard approach is to assume that random vector ξ has a finite number of possible realizations or scenarios, say ξ_1, \dots, ξ_K , with respective (positive) probabilities p_1, \dots, p_K . Then the expectation can be written as the summation

$$\mathbb{E}[Q(x, \xi)] = \sum_{k=1}^K p_k Q(x, \xi_k), \quad (2.4)$$

and, moreover, the two-stage problem (2.2)–(2.3) can be formulated as one optimization problem:

$$\begin{aligned} \min_{x, y_1, \dots, y_K} \quad & c^\top x + \sum_{k=1}^K q_k^\top y_k \\ \text{s.t.} \quad & x \in X, T_k x + W_k y_k \leq h_k, k = 1, \dots, K. \end{aligned} \quad (2.5)$$

In the above formulation (2.5) we make one copy y_k , of the second stage decision vector, for every scenario $\xi_k = (q_k, T_k, W_k, h_k)$. By solving (2.5) we obtain an optimal solution \bar{x} of the first-stage problem and optimal solutions \bar{y}_k of the second-stage problem for each scenario ξ_k , $k = 1, \dots, K$. Given \bar{x} , each \bar{y}_k gives an optimal second-stage decision corresponding to a realization $\xi = \xi_k$ of the respective scenario.

In practice this recourse structure exists at many stages, and the two-stage model is an approximation of a more complicated process in which observations of uncertain parameters occur over time, and decisions are made that adapt to these observations. In mathematical terms, the random variables form a *stochastic process*, and the decisions to be made must be measurable with respect to the filtration defined by this process.

As an example of a problem with such a structure consider the problem of controlling a reservoir with maximum volume a and initial storage r , that is used for hydro-electricity generation in conjunction with a thermal unit to minimize the expected cost of meeting demand over some planning horizon $t = 1, 2, \dots, T$. For simplicity, we assume that the thermal unit has zero startup cost but capacity g MW, and a running cost of $\$c$ /MWh. We also assume that (known) electricity demand is $d(t)$, $t = 1, 2, \dots, T$, and the supply of wind power is zero. As before suppose that any shortages of electricity cost $\$s$ /MWh.

The new feature of our model is a random inflow sequence $I(t)$, $t = 1, 2, \dots, T$, that may be stored for electricity generation through a turbine with capacity b MW. This gives the

following model:

$$\begin{aligned}
\text{MSP: } \min & \quad \mathbb{E}[\sum_{t=1}^T cX(t) + sY(t)] \\
\text{s.t.} & \quad X(t) + U(t) + Y(t) = d, & t = 1, 2, \dots, T, \\
& \quad R(t) = R(t-1) - \beta U(t) - V(t) + I(t), & t = 1, 2, \dots, T, \\
& \quad 0 \leq R(t) \leq a, \quad R(0) = r, & t = 1, 2, \dots, T, \\
& \quad 0 \leq X(t) \leq g, & t = 1, 2, \dots, T, \\
& \quad 0 \leq U(t) \leq b, & t = 1, 2, \dots, T, \\
& \quad V(t) \geq 0, & t = 1, 2, \dots, T.
\end{aligned}$$

Here $X(t)$ is thermal generation, $U(t)$ is hydro generation, $Y(t)$ is unserved demand, $V(t)$ is spill, and $R(t)$ is reservoir storage at the end of period t , all of which are random variables. The parameter $\beta \text{m}^3/\text{MWh}$ converts the electric power generated in a period into the corresponding volume of water released.

This general formulation MSP can be interpreted in many ways depending on how the random inflows $I(t)$ are modeled. A possible model could consider all T inflows to be revealed to the decision maker at once, who plans a (possibly different) least cost electricity generation schedule for each realization of inflows and then computes its cost. This *wait-and-see* version of the problem is of course unrealistically optimistic, and will give policies that will not be able to be implemented unless perfect foresight is available. In contrast, a *here-and-now* version of the problem might seek a fixed plan of thermal generation decisions $x(t)$ that must be feasible for all realizations of $I(t)$ and minimizes $\sum_{t=1}^T cx(t) + \mathbb{E}[\sum_{t=1}^T sY(t)]$. Because of its inflexibility, this model is likely to result in expensive policies.

Between these extremes there are many models that incorporate flexibility in the policies. The modeling choice then becomes a tradeoff between realism and computational tractability. If $I(t)$ follows a general stochastic process then MSP becomes a multi-stage stochastic linear program. Traditional stochastic programming methodologies for solving these models use *scenario trees* which approximate the stochastic process by one defined on a finite set of outcomes in each stage leading to a large-scale mathematical programming problem. If either the number of stages or dimension of the random variable become large then scenario-tree approximations can be quite poor, and so this approach has serious limitations for many realistically-sized problems.

On the other hand, when $I(t)$ is stage-wise independent or is a Markov process, MSP can be treated as a Markov decision problem that can be attacked using methods of dynamic programming. For the reservoir problem we define $Q_{t-1}(r)$ to be the minimum expected thermal and shortage cost incurred over the period $t, t+1, \dots, T$, if the reservoir storage at the end of period $t-1$ is r , and an optimal policy is followed. In the case where $I(t)$ is

stage-wise independent this gives the recursion:

$$\begin{aligned}
Q_{t-1}(r) = \text{DP}(t): \quad & \min \quad cx(t) + sy(t) + \mathbb{E}[Q_t(R(t))] \\
& \text{s.t.} \quad x(t) + u(t) + y(t) = d, \\
& \quad \quad R(t) = r - \beta u(t) - V(t) + I(t), \\
& \quad \quad 0 \leq R(t) \leq a, \\
& \quad \quad 0 \leq x(t) \leq g, \\
& \quad \quad 0 \leq u(t) \leq b, \\
& \quad \quad V(t) \geq 0.
\end{aligned}$$

Solving such a recursion in general is possible only if the state variable r has low dimension.

We conclude here by discussing the advantages of modelling a multi-stage decision problem as a dynamic program rather than a multi-stage stochastic program. As mentioned above scenario-tree models become computationally intractable as the number of stages and random outcomes grows. In contrast, dynamic programming can admit many stages, but suffers from a curse of dimensionality in the state variables. So there are computational advantages when the state space is of low dimension.

A second advantage of dynamic programming is that it gives a state-feedback policy, rather than an open-loop set of actions adapted to the filtration of the stochastic process. This makes simulation much easier. To simulate a multi-stage stochastic programming solution in a rolling horizon implementation requires a model like MSP to be generated and solved as each stage unfolds, a considerable computational effort considering that $T - 1$ such solves are required for a single realization of random outcomes over T stages. In contrast, dynamic programming given $Q_t(\cdot)$ would solve $\text{DP}(t)$ at each stage in such a simulation, a much less intensive effort.

3 Optimization modeling without probability

In this section we give a brief discussion of optimization under uncertainty without using probability distributions. The principle in this setting is to make decisions that are somehow *robust* in the face of variations in data. Different modeling approaches are available here, and we give a brief overview, returning to the unit commitment problem of section 2 as an illustration. Recall the plot of the cost outcomes of the alternative decisions, shown again in Figure 3.

The classical *minimax* criterion used in modeling without probability seeks to minimize the worst case. In the unit commitment example the worst-case cost is 60 for $z = 0$ and is 35 for $z = 1$, so $z = 1$ is the preferred solution under a minimax objective. Observe that this would still be the minimax solution even if $C(0, \omega)$ was 36 when $S(\omega) = 0$, and zero for $S(\omega) > 0$, which would make the choice of $z = 1$ an overly cautious solution.

An alternative criterion is to minimize maximum regret. Here *regret* is defined to be the difference in cost between the decision taken and the optimal decision with hindsight. The maximum regret for the decision $z = 1$ in the example is 20, which occurs when $S(\omega) \geq 6$.

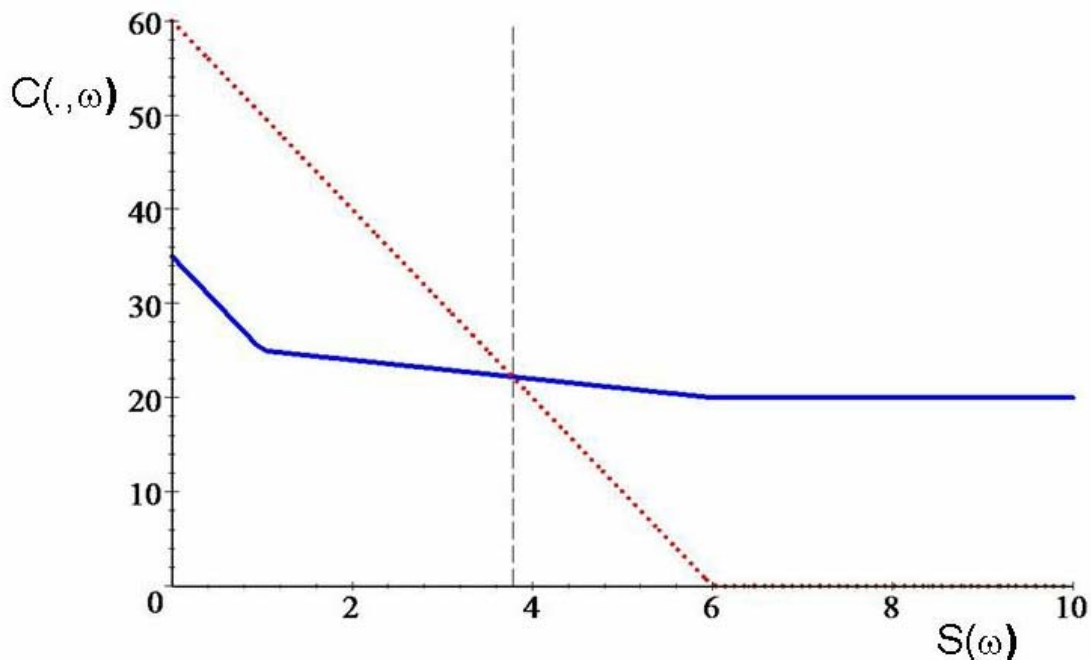


Figure 3: Plot of $C(0, \omega)$ (dotted) and $C(1, \omega)$ (solid) for unit commitment example.

The maximum regret for the decision $z = 0$ in the example is 25, which occurs when $S(\omega) \leq 1$. Thus we should choose $z = 1$. Observe that if $C(0, \omega)$ was 36 when $S(\omega) = 0$, and zero for $S(\omega) > 0$ then the maximum regret for the decision $z = 0$ in the example is 1, which makes it preferred to $z = 1$. This instance does however illustrate a paradox with this approach. If $C(0, \omega)$ was 56 when $S(\omega) = 0$, and zero for $S(\omega) > 0$ then the maximum regret for the decision $z = 0$ is 21, which makes $z = 1$ preferred again, even though it is a worse decision in every outcome except when $S(\omega) = 0$. This highlights the advantages of a probabilistic model that can quantify the tradeoff between different outcomes.

Since the minimax criterion is widely recognized as being too conservative, there have been a number of efforts to make it less so. *Robust optimization* applies the minimax criterion to a subset of possible outcomes called an *uncertainty set*. Formally robust optimization seeks the solution to the semi-infinite program

$$\begin{aligned} \text{RP: } \min_x \quad & f(x) \\ \text{s.t.} \quad & g(x, u) \leq 0, \quad u \in U, \end{aligned}$$

where the functional dependence of g on u and uncertainty set U is specially constructed so that this problem can be solved using an efficient algorithm.

In the unit-commitment example, one might choose

$$U_1 = \{\omega \mid 2 \leq S(\omega) \leq 8\},$$

and solve for example

$$\begin{aligned} \text{RP: } \min_{x,z} \quad & x \\ \text{s.t.} \quad & C(z, \omega) \leq x, \omega \in U_1. \end{aligned}$$

This has solution $z = 1$. On the other hand choosing

$$U_2 = \{\omega \mid 4 \leq S(\omega) \leq 10\},$$

gives optimal solution $z = 0$.

There has been considerable discussion in the literature on the relative merits of robust optimization and chance-constrained programming. The latter approach as part of the optimization seeks the best set U of outcomes with probability $1 - \alpha$ for which RP gives the least cost, and so from a modeling approach it is richer. Thus if we assume a uniform distribution in the above example, the chance-constrained problem

$$\begin{aligned} \text{CP: } \min_{x,z} \quad & x \\ \text{s.t.} \quad & \Pr(C(z, \omega) \leq x) \geq 0.6, \end{aligned}$$

has optimal solution $z = 0$, corresponding to the optimal choice of uncertainty set U_2 .

On the other hand, the optimal set U may be poorly behaved (making even RP for this U a difficult problem to solve) and so a chance-constrained model might prove impossible to solve for instances with lots of variables. In this case, robust optimization provides an alternative modeling approach.

4 Bibliographical notes

The concept of two-stage linear stochastic programming with recourse was introduced in Beale [2] and Dantzig [8]. For the theory and applications of two-stage and multi-stage stochastic programming see the monographs [6, 12, 14, 15, 18]. Chance constraints problems were introduced in Charnes, Cooper and Symonds [7]. For a thorough discussion and development of that concept we may refer to Prékopa [12]. An introductory treatment is available in the on-line tutorial by Henrion [10].

The first mathematical discussion of risk aversion (using utility functions) is often attributed to Daniel Bernoulli [4]. The concept of utility was formally defined and expounded by von Neumann and Morgenstern [11]. Coherent risk measures were first introduced by Artzner *et al* [1]. The approach of using CVaR for approximating chance constraints is due to Rockafellar and Uryasev [13].

Robust optimization models using ellipsoidal uncertainty sets were introduced by Ben Tal and Nemirovski [3] who showed how to formulate these as second-order cone problems. An alternative approach is due to Bertsimas and Sim [5].

Acknowledgements

Much of this paper is drawn from the material in the on-line tutorial paper [16]. The author would like to acknowledge the contributions of Alexander Shapiro to this work.

References

- [1] Artzner, P. Delbaen, F., Eber, J.-M. and Heath, D., Coherent measures of risk, *Mathematical Finance*, **9**, 203–228 (1999).
- [2] Beale, E.M.L., On minimizing a convex function subject to linear inequalities, *Journal of the Royal Statistical Society, Series B*, **17**, 173–184 (1955).
- [3] Ben Tal, A. and Nemirovski, A., Robust Convex Optimization, *Mathematics of Operations Research* **23**, 4, 769–805 (1998).
- [4] Bernoulli, D., Exposition of a new theory on the measurement of risk, 1738. (Translated by L. Sommer in *Econometrica* **22** (1): 22–36, 1954)
- [5] Bertsimas, D. and Sim, M., The Price of Robustness, *Operations Research* **52**, 1, 35–53 (2004).
- [6] Birge, J.R. and Louveaux, F., *Introduction to Stochastic Programming*, Springer, 1997.
- [7] Charnes, A., Cooper, W.W. and G.H. Symonds, Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil, *Management Science*, **4**, 235–263 (1958).
- [8] Dantzig, G.B., Linear programming under uncertainty, *Management Science*, **1**, 197–206 (1955).
- [9] Figueira, J.R., Greco, S. and Ehrgott, M. *Multiple Criteria Decision Analysis: State of the Art Surveys* International Series in Operations Research & Management Science, Vol. 78, 2005.
- [10] Henrion, R., Introduction to chance-constrained programming, <http://www.stoprog.org>.
- [11] von Neumann, J. and Morgenstern, O., *Theory of Games and Economic Behavior*, 1944.
- [12] Prékopa, A., *Stochastic Programming*, Kluwer, Dordrecht, Boston, 1995.
- [13] Rockafellar, R.T. and Uryasev, S.P., Optimization of conditional value-at-risk, *The Journal of Risk*, **2**, 21–41, 2000.
- [14] Ruszczyński, A. and Shapiro, A., (Eds.), *Stochastic Programming*, Handbook in OR & MS, Vol. 10, North-Holland Publishing Company, Amsterdam, 2003.
- [15] Shapiro, A., Dentcheva, D. and Ruszczyński, A., *Lectures on Stochastic Programming: Modeling and Theory*, SIAM, 2009.
- [16] Shapiro, A. and Philpott, A.B., A tutorial on stochastic programming, <http://www.stoprog.org>.

- [17] Taleb, N. *The Black Swan: the Impact of the Highly Improbable*, Random House, 2007.
- [18] Ziemba, W.T. and Wallace, S.W., *Applications of Stochastic Programming*, SIAM, 2006.