Notes on Stochastic Optimization Problems Formulation, Warren Powell

compiled by D.Gueorguiev, 2/25/25

# Formulating a Stochastic Optimization Problem

We will consider the ubiquitous inventory problem as an example and will study deterministic and stochastic formulations of the problem.

This requires the introduction of the notion of *policy* (or *control law*) which is a method for making decisions.

## Deterministic inventory problem

We want to solve a simple inventory in which we have to decide how much to order, , at time . We are going to assume that when we order , the items cannot be used until time . Let be the cost of items ordered in period (which can vary from one time period to the next), and assume we are paid a price when we satisfy the demand given by . Let be the sales at time , which is limited by the demand , and the available product which is our inventory plus our orders , so we can write

(1)

(2)

(3)

(4)

We assume that unsatisfied demand is lost. The left-over inventory is:

(5)

We set .

Now we formulate our deterministic optimization problem as

(6)

subject to the constraints (1) – (6). The solution is in the form of the vector of production and sales decisions . It becomes clear that we need to solve the inventory problem over the entire horizon to make the best decision now. For example, if we specified in advance , then this could easily change what we do now, , at time .

# The transition to a stochastic formulation

Let us consider random demand and random inventory . This means that the order quantity is random, as is the sales . Given this, the optimization problem (6) simply does not make sense as performing optimization over a set of random variables is not meaningful concept.

We fix this deficiency by replacing the decisions with a function, known as *policy* (or *control law*)

where is our state variable, capturing what we need to know to make a decision. For the inventory problem the state is represented by just the inventory . This allows us to rewrite our objective function as

(7)

This equation is still solved given the constraints (1)-(5) but they are applied differently.

The policy has to obey (1)-(4) which can be done because at time both and are known.

Then, after computing we can compute from (5). The expectation in (7) is obtained by sampling over all different sequences . In practice we cannot do that, so we create a sample where represents a particular sequence of possible values of . Assuming that each sequence of demands is equally likely, we approximate (7) with

(8)

where is the number of sample paths in .

Question: how to find those policies which will be advantageous for our problem?

# Choosing Inventory Policies

Policy is a function that returns a feasible decision using the information contained in the state variable .

Policies come in different forms. For the inventory problem we could implement a simple rule as a policy. For example:

Let . We write the policy as

(9)

Then we set our sales quantity as , which means we satisfy as much demand as we can.

This policy is parametrized by , so now we just have to find the values that produce the highest profit.

From (7) the objective function becomes:

(10)

To solve (10) we could do grid search over all possible values of using the sampled approximation (8).

Using the sampled approximation (9) we can rewrite (10) as

(11)

The “order-up-to” policy would not work well in settings with highly variable (yet predictable) demands. In such a setting, it makes more sense to use a look-ahead policy, where we optimize a deterministic approximation over some appropriately chosen horizon H. For this, let

- the forecast of the demand made at time .

Now, we formulate a deterministic approximation of the problem at time using

(12)

– the prediction at time that we will sell this number of units at a future moment

- the price of a unit quantity at a moment

- the cost of a unit quantity at a moment

– the prediction at time that we will order this number of units at a future moment

– forecasted demand , the best estimate of given what we know at time .

Then the lookahead version of the constraints (1)-(5) is given with

# References

[1] [Reinforcement Learning and Stochastic Optimization: A Unified Framework for Sequential Decisions, Warren Powel, 2019](https://github.com/dimitarpg13/reinforcement_learning_and_game_theory/blob/main/books/Powell-Reinforcement-Learning-and-Stochastic-Optimization.pdf)