ISyE/CS 719: Stochastic Programming
Fall 2016
Assignment #4 Example Solutions

1. A hospital needs to determine how many nurses to staff in a 12 hour shift for its four departments: Emergency ($i = 1$), Intensive Care ($i = 2$), Surgery ($i = 3$), and General ($i = 4$) (let the department set be $D = \{1, 2, 3, 4\}$). There are four classes of nurses, corresponding to the four departments, representing the type of care they are trained for (so nurse types are also given by the set $D = \{1, 2, 3, 4\}$). The cost to the hospital for staffing a nurse of each type for the shift are listed in the table below.

	Emergency	Surgery	Intensive Care	General
Cost/nurse (c_i)	2400	3000	2600	2000
Time/patient (a_i)	0.5	1.0	0.75	0.3

The number of patients that will need to be cared for during the shift at each department is a random variable, ξ_i, $i \in D$. Nurses associated with department i spend an average of a_i hours per patient in that department. However, during the course of the shift, some types of nurses can also help provide service in departments outside their main department. When doing so, the time spent per patient is about 20% higher than the time that would be spent by a nurse within that department. Specifically, nurses from any department can care for patients in the General department, and in addition, Surgery nurses can also care for patients in the Intensive Care and Emergency departments. Thus, for example, if a nurse from the Emergency department cares for a patient in the General department, the average time spent would be about $1.2(0.3) = 0.36$ hours. The number of nurses to have working in each department must be decided before the random number of patients is observed. However, the decision to use nurses from one department within another can be done in real-time, and hence is performed after observing the random variables. Assume furthermore that nurses can cover any fraction of a patient's required service time (so assignment of patients in a department to nurse types does not have to be integer). It is difficult to estimate the cost of having an insufficient number of nurses staffed, so the hospital wishes to impose a restriction that with probability at least 0.95 the staffing levels are sufficient to care for all patients.

(a) Formulate a chance-constrained stochastic program to minimize the staffing costs while meeting the service requirement. Here, the formulation should be "generic" - i.e., without assuming a discrete distribution on the random vector ξ.

Answer:

Following are the first-stage decision variables:

- x_i: number of nurses to assign to the shift in department i, for $i \in D$

These are integer decision variables. The objective is to minize cost:

$$\min \sum_{i \in D} c_i x_i$$

With the first-stage decision variables fixed at x, and given the observed random demand vector ξ, the system is feasible if there exists an assignment of nurses to departments (home or different) that meets all customer demands. We model this assignment with the continuous decision variables:

- y_{ij}: number of patients in department j that are cared for by a nurse in department i, $i, j \in D$

By assumption the care of a patient can be split between nurses, so these are continuous variables. An assignment is feasible if it meets all customer demands, and does not exceed the available time of the scheduled nurses:

$$\sum_{i \in D} y_{ij} \geq \xi_j, \quad j \in D \tag{1}$$

$$a_i y_{ii} + \sum_{j \in D: j \neq i} (1.2a_i) y_{ij} \leq 12 x_i, \quad i \in D \tag{2}$$

$$y_{ij} \geq 0, \quad i \in D, j \in D \tag{3}$$

Let $P(\xi) = \{x \in \mathbb{R}^{|D|} : \exists y \text{ s.t. } (1)--(3)\}$. Then $P(\xi)$ is the set of values of x for which a feasible recourse action exists when the demands are ξ. Thus, the chance constraint can be written as:

$$\mathbb{P}(x \in P(\xi)) \geq 0.95.$$

$\Diamond$

(b) Now assume the random vector ξ has finitely many equally likely scenarios, $\xi^1, \ldots, \xi^N$. Formulate the chance-constrained problem as an explicit mixed-integer linear program. Whena "big-M" value is needed, describe how you can obtain a large enough big-M value as a function of the data.

Answer:

We now introduce the following explicit second-stage variables:

- y_{ij}^k: number of patients in department j that are cared for by a nurse in department i in scenario k, $i, j \in D$ and $k = 1, \ldots, N$.
- z_k: binary variable where $z_k = 1$ implies all patients are cared for in scenario k

A complete extensive form can then be written down as follows:

$$\begin{aligned}
\min \quad & \sum_{i \in D} c_i x_i & & \\
\text{s.t.} \quad & \sum_{i \in D} y_{ij}^k \geq \xi_j^k z_k, & & j \in D, k = 1, \ldots, N & (4)\\
& a_i y_{ii}^k + \sum_{j \in D: j \neq i} (1.2a_i) y_{ij} \leq 12 x_i, & & i \in D, k = 1, \ldots, N & (5)\\
& \sum_{k=1}^{N} z_k \geq (0.95)N & & & (6)\\
& y_{ij}^k \geq 0, & & i \in D, j \in D, k = 1, \ldots, K & (7)\\
& z_k \in \{0, 1\}, & & k = 1, \ldots, K & (8)
\end{aligned}$$

Note that a direct application of the "big-M" procedure would also include big-M coefficients multipled with the term $(1 - z_k)$ on the right-hand-side of the constraints (5). However, the above formulation can directly be argued to be valid. In particular, when $z_k = 1$, it precisely enforces the required constraints for scenario k. On the other hand, when $z_k = 0$, the constraints (4) reduce to $\sum_{i \in D} y_{ij}^k \geq 0$, which can be satisfied by setting $y_{ij}^k = 0$, and hence constraints (5) are also satisfied for any $x_i \geq 0$. (Full credit will be given for a solution that uses big-M coefficients in both sets of constraints, provided valid values for the big-M coefficients are provided.) $\diamondsuit$

(c) Assume the random vector of customer service amounts is given by $\xi_1 = \xi_1^0$, $\xi_2 = \xi_2^0 + 0.2\xi_1^0$, $\xi_3 = \xi_3^0 + 0.2\xi_1^0 + 0.1\xi_2^0$, $\xi_4 = \xi_4^0 + 0.3\xi_1^0 + 0.6\xi_2^0 + 0.2\xi_3^0$, where the random variables ξ_i^0, $i = 1, 2, 3, 4$ are Poisson random variables with means $(50, 35, 30, 10)$. (This model captures dependence between the patient volumes. Here the variables ξ_i^0 represent patients that arrive directly to that department, whereas the other terms represent patients that go to that department from another department. E.g., among patients who arrive directly at the emergency room, 20% also go to surgery, 20% to intensive care, and 30% to the general department.) Take a sample of $N = 100$ from this distribution and formulate and solve the corresponding sample average approximation problem, using $\epsilon = 0.05$ in the chance constraint.

(d) Replicate the sample problem solved in the previous part 10 times to obtain a statistical lower bound on the optimal value. Also, choose one solution from these 10 problems and use a large sample $M = 100,000$ to obtain a confidence interval on the probability that this solution is feasible. (Note: this will likely be larger than the target $\epsilon = 0.05$.)

Answer:
See the example implementation solution posted on the course web site. In my sample of 10 the minimum value was 31,600, which is a lower bound on the true optimal value with confidence at least 99%. In addition, I arbitrarily selected the last solution of the 10 SAA problems as a candidate solution. It had objective value 32,000, and a 95% confidence interval on its violation probability is [0.016,0.017], and so this has high confidence to be a feasible solution to the chance constraint, and hence provide an upper bound on the cost. This solution staffs 3 nurses in the ER, 3 nurses in intensive care, 5 nurses in surgery, and 1 nurse in the general department. (Note: using a sample size of $N = 500$ I achieved a lower bound of 31,600 and also an approximately feasible solution matching that, so there is a good chance based on that experiment that the optimal value is exaclty 31,600.) $\diamondsuit$

2. Fred owns a forest and has a contract to sell timber he harvests to a local paper company. The forest has trees that may be broken into four classes: Class 1 has trees 0-25 years old, Class 2 has trees 26-50 years old, Class 3 has trees 51-75 years old, and Class 4 has trees older than 75 years. The forest currently has 8,000 acres in Class 1, 10,000 acres in Class 2, 20,000 acres in Class 3, and 60,000 acres in Class 4. Each class of timber has a different yield: Class 1 has no yield, Class 2 yields 250 cubic feet per acre, Class 3 yields 510 cubic feet per acre, and Class 4 yields 710 cubic feet per acre. The "danger" in waiting to harvest trees is that fires or bug-infestations can destroy portions of the forest – effectively returning those stands to Class 1. Each 25-year period, it is equally likely that 5%, 10%, or 15% of the trees will be destroyed; these destruction events are independent between time periods but are common to the entire forest (i.e., all four classes) within a time period. Trees that aren't destroyed by fire or bugs, and aren't harvested, progress to the next class in the next time period (e.g., Class 2 trees become Class 3 trees) except that Class 4 trees remain in Class 4. If an acre of trees in a class is cut or destroyed then it transitions to Class 1 in the next stage. Fred's contract specifies that over the next 100 years the paper company will pay (in present-dollar terms so

no discounting is needed): 10 cents per cubic foot up to 98 million cubic feet, 7 cents per cubic foot between 98-105 million cubic feet, and 5 cents per cubic foot in excess of 105 million cubic feet.

(a) Formulate a four-stage stochastic program that will give a harvesting plan for the next 100 years that maximizes expected profit. (Assume that the forest has no value to Fred's family at the end of the 100-year planning period.) You may use either the node-based or scenario-based formulation. You should include the scenario tree as part of your solution.

Answer:

We use a scenario-based formulation. We define some notation on parameters to write the formulation generically:

- p_k: price paid (in thousands of dollars) per million cubic feet at level k, $k = 1, 2, 3$ ($p_1 = 100$, $p_2 = 70$, $p_3 = 50$)
- u_k: limit on amount that can be sold at price level k, $k = 1, 2$ ($u_1 = 98$, $u_2 = 7$)
- f_{ts}: fraction of trees that survive in stage t under scenario s ($f_{1s} \equiv 1$ because already observed what has survived)
- $w_{i,0,s}$: Data for how many thousands acres of each class are available at $t = 1$ in each class i (same for all scenarios, index s used for consistency in scenario formulation) ($w_{1,0,s} \equiv 8, w_{2,0,s} \equiv 10, w_{3,0,s} \equiv 20, w_{4,0,s} \equiv 60$)
- a_i: yield of wood of class i harvested ($a_2 = 0.250$, $a_3 = 0.510$, $a_4 = 0.710$) (units: millions of cubic feet per thousand acres)

The decision variables are as follows.

- w_{its}: Thousands of acres of class i trees after harvesting in stage t and scenario s (will become class $i + 1$ in next stage if not destroyed), indexed from $i = 0, \ldots, 3$ ($i = 0$ represents total of what is harvested now, so will be clas 1 next stage)
- x_{its}: Thousands of acres of class i trees to harvest in stage t under scenario s, $i = 2, 3, 4$
- y_{kts}: Millions of cubic feet sold at price level k in stage t under scenario s, $k = 1, 2, 3$
- z_{kts}: Millions of cubic feet sold at price level k in stages up to and including stage t under scenario s, $k = 1, 2, 3$

Objective:

$$\min \sum_{t=1}^{5} \sum_{s=1}^{81} \sum_{k=1}^{3} p_k y_{kts}$$

The constraints, excluding the nonanticipativity constraints, are as follows:

$$w_{its} = w_{i-1,t-1,s} f_{ts} - x_{its} \qquad i = 2, 3, t = 1, \dots, 5, \forall s \tag{9}$$
$$w_{4ts} = (w_{3,t-1,s} + w_{4,t-1,s}) f_{ts} - x_{4ts} \qquad t = 2, \dots, 5, \forall s \tag{10}$$
$$w_{1ts} = w_{0,t-1,s} + (1 - f_{ts}) \Big(\sum_{i=1}^{4} w_{i,t-1,s} \Big) \qquad t = 2, \dots, 5, \forall s \tag{11}$$
$$w_{0ts} = \sum_{i=2}^{4} x_{its} \qquad t = 1, 2, \dots, 5 \tag{12}$$
$$\sum_{k=1}^{3} y_{kts} = \sum_{i=2}^{4} a_i x_{its}, \qquad t = 1, \dots, 5, \forall s \tag{13}$$
$$z_{kts} = z_{k,t-1,s} + y_{kts} \qquad k = 1, 2, 3, t = 1, \dots, 5, \forall s \tag{14}$$
$$w_{its} \geq 0, \qquad i = 1, 2, 3, 4, t = 1, \dots, 5, \forall s \tag{15}$$
$$x_{its} \geq 0, \qquad i = 2, 3, 4, t = 1, \dots, 5, \forall s \tag{16}$$
$$0 \leq z_{kts} \leq u_k \qquad k = 1, 2, 3, t = 1, \dots, 5, \forall s \tag{17}$$
$$y_{kts} \geq 0, \qquad k = 1, 2, 3, t = 1, \dots, 5, \forall s. \tag{18}$$

I write the nonanticipativity constraints generically below. One observation is that they only need to be enforced for the state variables (h and z). It is fine, however, to also add them for the other decision variables.

$$w_{i,t(n),s} = \frac{1}{|S(n)|} \sum_{s' \in S(n)} w_{i,t(n),s} \qquad \forall i = 0, 1, 2, 3, 4, n \in \mathcal{T} \tag{19}$$
$$z_{i,t(n),s} = \frac{1}{|S(n)|} \sum_{s' \in S(n)} z_{i,t(n),s} \qquad \forall i = 0, 1, 2, 3, 4, n \in \mathcal{T} \tag{20}$$

$\Diamond$

(b) Implement and solve your model. Report the expected income, and explain the solution.

Answer:
See the implementation file on the course web site. The expected value of the optimal policy is 10.717 million dollars. At the current time (stage $t = 1$) the policy recommends harvesting 20 acres of class 3 forest and 60 acres of class 4 forest. Future decisions will be made based on re-optimizing the model given the observed outcomes in the future (so it is not necessary to state them when stating the solution). However, as a general observation abou the policy, in the following periods, the policy generally is to only to harvest small amounts of class 3 forest in stages 2 and 3, and nothing else until stage 5, when everything is harvested. As an aside, I comment that in practical forestry planning problems, constraints are often imposed on the "connectivity" of remaining forest, to keep it suitable for wildlife, and such constraints would prevent the "clear cut" solution that happens in the initial and final stages here. $\Diamond$