Chapter 2
Uncertainty and Modeling Issues

In the previous chapter, we gave several examples of stochastic programming models. These formulations fit into different categories of stochastic programs in terms of the characteristics of the model. This chapter presents those basic characteristics by describing the fundamentals of any modeling effort and some of the standard forms detailed in later chapters.

Before beginning general model descriptions, however, we first describe the probability concepts that we will assume in the rest of the book. Familiarity with these concepts is essential in understanding the structure of a stochastic program. This presentation is made simple enough to be understood by readers unfamiliar with the field and, thus, leaves aside some questions related to measure theory. Sections 2.2 through 2.7 build on these fundamentals and give the general forms in various categories. Section 2.8 provides a detailed discussion of a modeling exercise. Sections 2.9 and 2.10 give alternative characterizations of stochastic optimization problems and some background on the relationship of stochastic programming to other areas of decision making under uncertainty. Section 2.11 briefly reviews the main optimization concepts used in the book.

2.1 Probability Spaces and Random Variables

Several parameters of a problem can be considered uncertain and are thus represented as random variables. Production and distribution costs typically depend on fuel costs, which are random. Future demands depend on uncertain market conditions. Crop returns depend on uncertain weather conditions.

Uncertainty is represented in terms of random experiments with outcomes denoted by ω. The set of all outcomes is represented by Ω. In a transport and distribution problem, the outcomes range from political conditions in the Middle East to general trade situations, while the random variable of interest may be the fuel cost. The relevant set of outcomes is clearly problem-dependent. Also, it is usually not

J.R. Birge and F. Louveaux, *Introduction to Stochastic Programming*, Springer Series in Operations Research and Financial Engineering, DOI 10.1007/978-1-4614-0237-4_2,

very important to be able to define those outcomes accurately because the focus is mainly on their impact on some (random) variables.

The outcomes may be combined into subsets of Ω called *events*. We denote by $\mathscr{A}$ a collection of random events. As an example, if Ω contains the six possible results of the throw of a die, $\mathscr{A}$ also contains combined outcomes such as an odd number, a result smaller than or equal to four, etc. If Ω contains weather conditions for a single day, $\mathscr{A}$ also contains combined events such as "a day without rain," which might be the union of a sunny day, a partly cloudy day, a cloudy day without showers, etc.

Finally, to each event $A \in \mathscr{A}$ is associated a value $\mathrm{P}(A)$, called a *probability*, such that $0 \leq \mathrm{P}(A) \leq 1$, $\mathrm{P}(\emptyset) = 0$, $\mathrm{P}(\Omega) = 1$ and $\mathrm{P}(A_1 \cup A_2) = \mathrm{P}(A_1) + \mathrm{P}(A_2)$ if $A_1 \cap A_2 = \emptyset$. The triplet $(\Omega, \mathscr{A}, P)$ is called a *probability space* that must satisfy a number of conditions (see, e.g., Chung [1974]). It is possible to define several random variables associated with a probability space, namely, all variables that are influenced by the random events in $\mathscr{A}$. If one takes as elements of Ω events ranging from the political situation in the Middle East to the general trade situations, they allow us to describe random variables such as the fuel costs and the interest rates and inflation rates in some Western countries. If the elements of Ω are the weather conditions from April to September, they influence random variables such as the production of corn, the sales of umbrellas and ice cream, or even the exam results of undergraduate students.

In terms of stochastic programming, there exists one situation where the description of random variables is closely related to Ω: in some cases indeed, the elements $\omega \in \Omega$ are used to describe a few *states of the world* or *scenarios*. All random elements then jointly depend on these finitely many scenarios. Such a situation frequently occurs in strategic models where the knowledge of the possible outcomes in the future is obtained through experts' judgments and only a few scenarios are considered in detail. In many situations, however, it is extremely difficult and pointless to construct Ω and $\mathscr{A}$; the knowledge of the random variables is sufficient.

For a particular random variable $\boldsymbol{\xi}$, we define its cumulative distribution $F_{\boldsymbol{\xi}}(x) = \mathrm{P}(\boldsymbol{\xi} \leq x)$, or more precisely $F_{\boldsymbol{\xi}}(x) = \mathrm{P}(\{\omega \mid \boldsymbol{\xi} \leq x\})$. Two major cases are then considered. A discrete random variable takes a finite or countable number of different values. It is best described by its probability distribution, which is the list of possible values, ξ^k, $k \in K$, with associated probabilities,

$$f(\xi^k) = \mathrm{P}(\boldsymbol{\xi} = \xi^k) \quad \text{s. t.} \sum_{k \in K} f(\xi^k) = 1 \,.$$

Continuous random variables can often be described through a so-called *density* function $f(\xi)$. The probability of ξ being in an interval $[a, b]$ is obtained as

$$\mathrm{P}(a \leq \boldsymbol{\xi} \leq b) = \int_a^b f(\boldsymbol{\xi}) d\boldsymbol{\xi} \,,$$

or equivalently

$$\mathrm{P}(a \leq \boldsymbol{\xi} \leq b) = \int_a^b dF(\boldsymbol{\xi}) \,,$$

where $F(\cdot)$ is the cumulative distribution as earlier. Contrary to the discrete case, the probability of a single value $\mathrm{P}(\boldsymbol{\xi} = a)$ is always zero for a continuous random variable. The distribution $F(\cdot)$ must be such that $\int_{-\infty}^{\infty} dF(\boldsymbol{\xi}) = 1$.

The *expectation* of a random variable is computed as $\mu = \sum_{k \in K} \xi^k f(\xi^k)$ or $\mu = \int_{-\infty}^{\infty} \boldsymbol{\xi} dF(\boldsymbol{\xi})$ in the discrete and continuous cases, respectively. The *variance* of a random variable is $\mathrm{E}[(\boldsymbol{\xi} - \mu)^2]$. The expectation of $\boldsymbol{\xi}^r$ is called the *r th moment* of $\boldsymbol{\xi}$ and is denoted $\bar{\xi}^{(r)} = \mathrm{E}[\boldsymbol{\xi}^r]$. A point η is called the α -quantile of $\boldsymbol{\xi}$ if and only if for $0 < \alpha < 1$, $\eta = \min\{x \mid F(x) \geq \alpha\}$.

The appendix lists the distributions used in the textbook and their expectations and variances. The concepts of probability distribution, density, and expectation easily extend to the case of multiple random variables. Some of the sections in the book use probability measure theory which generalizes these concepts. These sections contain a warning to readers unfamiliar with this field.

2.2 Deterministic Linear Programs

A deterministic linear program consists of finding a solution to

$$\begin{aligned} \min\, & z = c^T x \\ \text{s. t. } & Ax = b \,, \\ & x \geq 0 \,, \end{aligned}$$

where x is an $(n \times 1)$ vector of decisions and c , A and b are known data of sizes $(n \times 1)$, $(m \times n)$, and $(m \times 1)$, respectively. The value $z = c^T x$ corresponds to the objective function, while $\{x \mid Ax = b \,,\, x \geq 0\}$ defines the set of feasible solutions. An optimum x^* is a feasible solution such that $c^T x \geq c^T x^*$ for any feasible x . Linear programs typically search for a minimal-cost solution under some requirements (demand) to be met or for a maximum profit solution under limited resources. There exists a wide variety of applications, routinely solved in the industry. As introductory references, we cite Chvátal [1980], Dantzig [1963], and Murty [1983]. We assume the reader is familiar with linear programming and has some knowledge of basic duality theory as in these textbooks. A short review is given in Section 2.11.

2.3 Decisions and Stages

Stochastic linear programs are linear programs in which some problem data may be considered uncertain. *Recourse programs* are those in which some decisions or recourse actions can be taken after uncertainty is disclosed. To be more precise,

data uncertainty means that some of the problem data can be represented as random variables. An accurate probabilistic description of the random variables is assumed available, under the form of the probability distributions, densities or, more generally, probability measures. As usual, the particular values the various random variables will take are only known after the random experiment, i.e., the vector $\xi = \xi(\omega)$ is only known after the experiment.

The set of decisions is then divided into two groups:

- A number of decisions have to be taken before the experiment. All these decisions are called *first-stage decisions* and the period when these decisions are taken is called the *first stage*.
- A number of decisions can be taken after the experiment. They are called *second-stage decisions*. The corresponding period is called the *second stage*.

First-stage decisions are represented by the vector x, while second-stage decisions are represented by the vector y or $y(\omega)$ or even $y(\omega,x)$ if one wishes to stress that second-stage decisions differ as functions of the outcome of the random experiment and of the first-stage decision. The sequence of events and decisions is thus summarized as

$$x \rightarrow \xi(\omega) \rightarrow y(\omega,x) \ .$$

Observe here that the definitions of first and second stages are only related to before and after the random experiment and may in fact contain sequences of decisions and events. In the farming example of Section 1.1, the first stage corresponds to planting and occurs during the whole spring. Second-stage decisions consist of sales and purchases. Selling extra corn would probably occur very soon after the harvest while buying missing corn will take place as late as possible.

A more extreme example is the following. A traveling salesperson receives one item every day. She visits clients hoping to sell that item. She returns home when a buyer is found or when all clients are visited. Clients buy or do not buy in a random fashion. The decision is not influenced by the previous days' decisions. The salesperson wishes to determine the order in which to visit clients, in such a way as to be at home as early as possible (seems reasonable, does it not?). Time spent involves the traveling time plus some service time at each visited client.

To make things simple, once the sequence of clients to be visited is fixed, it is not changed. Clearly the first stage consists of fixing the sequence and traveling to the first client. The second stage is of variable duration depending on the successive clients buying the item or not. Now, consider the following example. There are two clients with probability of buying 0.3 and 0.8, respectively and traveling times (including service) as in the graph of Figure 1.

Assume the day starts at 8 A.M. If the sequence is $(1,2)$, the first stage goes from 8 to 9:30. The second stage starts at 9:30 and finishes either at 11 A.M. if 1 buys or 4:30 P.M. otherwise. If the sequence is $(2,1)$, the first stage goes from 8 to 12:00, the second stage starts at 12:00 and finishes either at 4:00 P.M. or at 4:30 P.M. Thus, the first stage if sequence $(2,1)$ is chosen may sometimes end after the second stage is finished when $(1,2)$ is chosen if Client 1 buys the item.

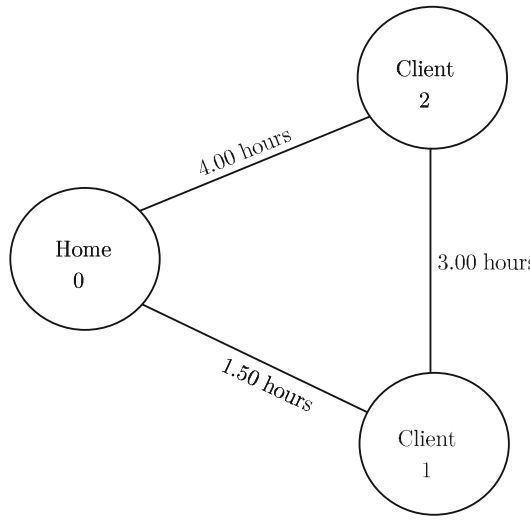

Fig. 1 Traveling salesperson example.

2.4 Two-Stage Program with Fixed Recourse

The classical two-stage stochastic linear program with fixed recourse (originated by Dantzig [1955] and Beale [1955]) is the problem of finding

$$\min z = c^T x + \mathrm{E}_{\boldsymbol{\xi}}[\min q(\omega)^T y(\omega)] \tag{4.1}$$

$$\text{s. t.} \qquad Ax = b\ , \tag{4.2}$$

$$T(\omega)x + Wy(\omega) = h(\omega)\ , \tag{4.3}$$

$$x \geq 0\ , y(\omega) \geq 0\ . \tag{4.4}$$

As in the previous section, a distinction is made between the first stage and the second stage. The first-stage decisions are represented by the $n_1 \times 1$ vector x. Corresponding to x are the first-stage vectors and matrices c, b, and A, of sizes $n_1 \times 1$, $m_1 \times 1$, and $m_1 \times n_1$, respectively. In the second stage, a number of random events $\omega \in \Omega$ may realize. For a given realization ω, the second-stage problem data $q(\omega)$, $h(\omega)$ and $T(\omega)$ become known, where $q(\omega)$ is $n_2 \times 1$, $h(\omega)$ is $m_2 \times 1$, and $T(\omega)$ is $m_2 \times n_1$.

Each component of q, T, and h is thus a possible random variable. Let $T_{i\cdot}(\omega)$ be the ith row of $T(\omega)$. Piecing together the stochastic components of the second-stage data, we obtain a vector $\xi^T(\omega) = (q(\omega)^T, h(\omega)^T, T_{1\cdot}(\omega), \ldots, T_{m_2\cdot}(\omega))$, with potentially up to $N = n_2 + m_2 + (m_2 \times n_1)$ components. As indicated before, a single random event ω (or state of the world) influences several random variables, here, all components of ξ.

Let also $\Xi \subset \Re^N$ be the *support* of ξ, that is, the smallest closed subset in $\Re^N$ such that $\mathrm{P}(\Xi) = 1$. As just said, when the random event ω is realized, the second-stage problem data, q, h, and T, become known. Then, the second-stage decision $y(\omega)$ or $(y(\omega,x))$ must be taken. The dependence of y on ω is of a completely different nature from the dependence of q or other parameters on ω. It is not functional but simply indicates that the decisions y are typically not the same under different realizations of ω. They are chosen so that the constraints (4.3) and (4.4) hold *almost surely* (denoted *a.s.*), i.e., for all $\omega \in \Omega$ except perhaps for sets with zero probability. We assume random constraints to hold in this way throughout this book unless a specific probability is given for satisfying constraints.

The objective function of (4.1) contains a deterministic term $c^T x$ and the expectation of the second-stage objective $q(\omega)^T y(\omega)$ taken over all realizations of the random event ω. This second-stage term is the more difficult one because, for each ω, the value $y(\omega)$ is the solution of a linear program. To stress this fact, one sometimes uses the notion of a deterministic equivalent program. For a given realization ω, let

$$Q(x,\xi(\omega)) = \min_y \{q(\omega)^T y \mid Wy = h(\omega) - T(\omega)x, y \geq 0\} \tag{4.5}$$

be the second-stage value function. Then, define the expected second-stage value function

$$\mathscr{Q}(x) = \mathrm{E}_{\xi} Q(x,\xi(\omega)) \tag{4.6}$$

and the *deterministic equivalent program* (DEP)

$$\min z = c^T x + \mathscr{Q}(x) \tag{4.7}$$

$$\text{s. t. } Ax = b\,, \tag{4.8}$$

$$x \geq 0\,.$$

This representation of a stochastic program clearly illustrates that the major difference from a deterministic formulation is in the second-stage value function. If that function is given, then a stochastic program is just an ordinary nonlinear program.

Formulation (4.1)–(4.4) is the simplest form of a stochastic two-stage program. Extensions are easily modeled. For example, if first-stage or second-stage decisions are to be integers, constraint (4.4) can be replaced by a more general form:

$$x \in X\,, \quad y(w) \in Y\,,$$

where $X = Z_+^{n_1}$ and $Y = Z_+^{n_2}$. Similarly, nonlinear first-stage and second-stage objectives or constraints can easily be incorporated.

Examples of recourse formulation and interpretations

The definition of first stage versus second stage is not only problem dependent but also context dependent. We illustrate different examples of recourse formulations for one class of problems: *the location problem*.

Let $i = 1,\ldots,m$ index clients having demand d_i for a given commodity. The firm can open a facility (such as a plant or a warehouse) in potential sites $j = 1,\ldots,n$. Each client can be supplied from an open facility where the commodity is made available (i.e., produced or stored). The problem of the firm is to choose the number of facilities to open, their locations, and market areas to maximize profit or minimize costs.

Let us first present the deterministic version of the so-called simple plant location or uncapacitated facility location problem. Let x_j be a binary variable equal to one if facility j is open and zero otherwise. Let c_j be the fixed cost for opening and operating facility j and let v_j be the variable operating cost of facility j. Let y_{ij} be the fraction of the demand of client i served from facility j and t_{ij} be the unit transportation cost from j to i.

All costs and profits should be taken in conformable units, typically on a yearly equivalent basis. Let r_i denote the unit price charged to client i and $q_{ij} = (r_i - v_j - t_{ij})d_i$ be the total revenue obtained when all of client i's demand is satisfied from facility j. Then the simple plant location problem or uncapacitated facility location problem (UFLP) reads as follows:

$$\text{UFLP: } \max_{x,y} z(x,y) = -\sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m}\sum_{j=1}^{n} q_{ij} y_{ij} \tag{4.9}$$

$$\text{s. t. } \sum_{j=1}^{n} y_{ij} \le 1\,, \qquad i = 1,\ldots,m\,, \tag{4.10}$$

$$0 \le y_{ij} \le x_j\,, \qquad i = 1,\ldots,m\,,\ j = 1,\ldots,n\,, \tag{4.11}$$

$$x_j \in \{0,1\}\,, \qquad j = 1,\ldots,n\,. \tag{4.12}$$

Constraints (4.10) ensure that the sum of fractions of clients i's demand served cannot exceed one. Constraints (4.11) ensure that clients are served only through open plants.

It is customary to present the uncapacitated facility location in a different canonical form that minimizes the sum of the fixed costs of opening facilities and of the transportation costs plus possibly the variable operating costs. (There are several ways to arrive at this canonical representation. One is to assume that unit prices are much larger than unit costs in such a way that demand is always fully satisfied.) This presentation more clearly stresses the link between the deterministic and stochastic cases.

In the UFLP, a trade-off is sought between opening more plants, which results in higher fixed costs and lower transportation costs and opening fewer plants with the opposite effect. Whenever the optimal solution is known, the size of an open

facility is computed as the sum of demands it serves. (In the deterministic case, it is always optimal to have each y_{ij} equal to either zero or one.) The market areas of each facility are then well-defined.

The notation x_j for the location variables and y_{ij} for the distribution variables is common in location theory and is thus not meant here as first stage and second stage, respectively, although in some of the models it is indeed the case.

Several parameters of the problem may be uncertain and may thus have to be represented by random variables. Production and distribution costs may vary over time. Future demands for the product may be uncertain.

As indicated in the introduction of the section, we will now discuss various situations of recourse. It is customary to consider that the location decisions x_j are first-stage decisions because it takes some time to implement decisions such as moving or building a plant or warehouse. The main modeling issue is on the distribution decisions. The firm may have full control on the distribution, for example, when the clients are shops owned by the firm. It may then choose the distribution pattern after conducting some random experiments. In other cases, the firm may have contracts that fix which plants serve which clients, or the firm may wish fixed distribution patterns in view of improved efficiency because drivers would have better knowledge of the regions traveled.

a. Fixed distribution pattern, fixed demand, r_i, v_j, t_{ij} stochastic

Assume the only uncertainties are in production and distribution costs and prices charged to the client. Assume also that the distribution pattern is fixed in advance, i.e., is considered first stage. The second stage then just serves as a measure of the cost of distribution. We now show that the problem is in fact a deterministic problem in which the total revenue $q_{ij} = (r_i - v_j - t_{ij})d_i$ can be replaced by its expectation. To do this, we formally introduce extra second-stage variables w_{ij}, with the constraint $w_{ij}(\omega) = y_{ij}$ for all ω. We obtain

$$\max \; -\sum_{j=1}^{n} c_j x_j + \mathrm{E}_{\boldsymbol{\xi}} \sum_{i=1}^{m} \sum_{j=1}^{n} q_{ij}(\omega) w_{ij}(\omega)$$

s.t. (4.10), (4.11), (4.12), and

$$w_{ij}(\omega) = y_{ij}, \quad i = 1, \ldots, m\,, \qquad j = 1, \ldots, n \quad \forall \omega\,. \tag{4.13}$$

By (4.13), the second-stage objective function can be replaced by

$$\mathrm{E}_{\boldsymbol{\xi}} \sum_{i=1}^{m} \sum_{j=1}^{n} q_{ij}(\omega) y_{ij}$$

or

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{E}_{\boldsymbol{\xi}} q_{ij}(\omega) y_{ij} \ ,$$

because y_{ij} is fixed and summations and expectation can be interchanged. The problem is thus the deterministic problem

$$\max \ -\sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m}\sum_{j=1}^{n} (\mathrm{E}_{\boldsymbol{\xi}} q_{ij}(\omega)) y_{ij}$$

s.t. (4.10), (4.11), (4.12).

Although there exists uncertainty about the distribution costs and revenues, the only possible action is to plan in view of the expected costs.

b. Fixed distribution pattern, uncertain demand

Assume now that demand is uncertain, but, for some of the reasons cited earlier, the distribution pattern is fixed in the first stage. Depending on the context, the distribution costs and revenues (v_j, t_{ij}, r_i) may or may not be uncertain.

We define $y_{ij} =$ quantity transported from j to i, a quantity no longer defined as a function of the demand d_i, because demand is now stochastic. For simplicity, we assume that a penalty q_i^+ is paid per unit of demand d_i which cannot be satisfied from all quantities transported to i (they might have to be obtained from other sources) and a penalty q_i^- is paid per unit on the products delivered to i in excess of d_i (the cost of inventory, for example). We thus introduce second-stage variables: $w_i^-(\omega) =$ amount of extra products delivered to i in state ω; $w_i^+(\omega) =$ amount of unsatisfied demand to i in state ω.

The formulation becomes

$$\max -\sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m}\sum_{j=1}^{n} (\mathrm{E}_{\boldsymbol{\xi}}(-v_j - t_{ij})) y_{ij} + \mathrm{E}_{\boldsymbol{\xi}}[-\sum_{i=1}^{m} q_i^+ w_i^+(\omega) - \sum_{i=1}^{m} q_i^- w_i^-(\omega)] + \mathrm{E}_{\boldsymbol{\xi}} \sum_{i=1}^{m} r_i d_i(\omega) \tag{4.14}$$

$$\text{s. t.} \ \sum_{i=1}^{m} y_{ij} \le M x_j \ , \qquad j = 1, \ldots, n \ , \tag{4.15}$$

$$w_i^+(\omega) - w_i^-(\omega) = d_i(\omega) - \sum_{j=1}^{n} y_{ij} \ , \qquad i = 1, \ldots, m \ , \tag{4.16}$$

$$x_j \in \{0,1\} \ , \ 0 \le y_{ij} \ , \ w_i^+(\omega) \ge 0 \ , w_i^-(\omega) \ge 0 \ , \quad i = 1, \ldots, m \ , \ j = 1, \ldots, n \ . \tag{4.17}$$

This model is a location extension of the transportation model of Williams [1963]. The objective function contains the investment costs for opening plants, the expected

production and distribution costs, the expected penalties for extra or insufficient demands, and the expected revenue. This last term is constant because it is assumed that all demands must be satisfied by either direct delivery or some other means reflected in the penalty for unmet demand. The problem only makes sense if q_i^+ is large enough, for example, larger than $\mathrm{E}_{\boldsymbol{\xi}}(v_j + t_{ij})$ for all j, although weaker conditions may sometimes suffice. Constraint (4.15) guarantees that distribution only occurs from open plants, i.e., plants such that $x_j = 1$. The constant M represents the maximum possible size of a plant.

Observe that here the variables y_{ij} are first-stage variables. Also observe that in the second stage, the constraints (4.16), (4.17) have a very simple form, as $w_i^+(\omega) = \mathbf{d}_i - \sum_{j=1}^n y_{ij}$ if this quantity is non-negative and $w_i^-(\omega) = \sum_{j=1}^n y_{ij} - \mathbf{d}_i$ otherwise. This is an example of a *second stage with simple recourse*.

Also note that in Cases a and b, the size or capacity of plant j is simply obtained as the sum of the quantity transported from j, namely, $\sum_{i=1}^m d_i y_{ij}$ in Case a and $\sum_{i=1}^m y_{ij}$ in Case b.

c. Uncertain demand, variable distribution pattern

We now consider the case where the distribution pattern can be adjusted to the realization of the random event. This might be the case when uncertainty corresponds to long-term scenarios, of which only one is realized. Then the distribution pattern can be adapted to this particular realization. This also implies that the sizes of the plants cannot be defined as the sum of the quantity distributed, because those quantities depend on the random event. We thus define as before:

$$x_j = \begin{cases} 1 & \text{if plant } j \text{ is open,} \\ 0 & \text{otherwise.} \end{cases}$$

We now let y_{ij} depend on ω with $y_{ij}(\omega) =$ fraction of demand $d_i(\omega)$ served from j and define new variables $w_j =$ size (capacity) of plant j, with unit investment cost g_j.

The model now reads

$$\max \; -\sum_{j=1}^n c_j x_j - \sum_{j=1}^n g_j w_j + \mathrm{E}_{\boldsymbol{\xi}} \max \sum_{i=1}^m \sum_{j=1}^n q_{ij}(\omega) y_{ij}(\omega) \tag{4.18}$$

$$\text{s. t.} \quad x_j \in \{0,1\}\,,\; w_j \geq 0\,, \qquad j = 1,\ldots,n\,, \tag{4.19}$$

$$\sum_{j=1}^n y_{ij}(\omega) \leq 1\,, \qquad i = 1,\ldots,m\,, \tag{4.20}$$

$$\sum_{i=1}^m d_i(\omega) y_{ij}(\omega) \leq w_j\,, \qquad j = 1,\ldots,n\,, \tag{4.21}$$

$$0 \le y_{ij}(\omega) \le x_j\,, \qquad i = 1,\ldots,m\,,\ j = 1,\ldots,n\,, \tag{4.22}$$

where $q_{ij}(\omega) = (r_i - v_j - t_{ij})d_i(\omega)$ now includes the demand $d_i(\omega)$.

Constraint (4.20) indicates that no more than 100% of i's demand can be served, but that the possibility exists that not all demand is served. Constraint (4.21) imposes that the quantity distributed from plant j does not exceed the capacity w_j decided in the first stage. For the sake of clarity, one could impose a constraint $w_j \le Mx_j$, but this is implied by (4.21) and (4.22). For a discussion of algorithmic solutions of this problem, see Louveaux and Peeters [1992].

d. Stages versus periods; Two-stage versus multistage

In this section, we highlight again the difference in a stochastic program between *stages* and *periods* of times. Consider the case of a distribution firm that makes its plans for the next 36 months. It may formulate a model such as (4.18)–(4.22). The location of warehouses would be first-stage decisions, while the distribution problem would be second-stage decisions. The duration of the first stage would be something like six months (depending on the type of warehouse) and the second stage would run over the 30 remaining months. Although we may think of a problem over 36 periods, a two-stage model is totally relevant. In this case, the only moment where the number of periods is important is when the precise values of the objective coefficients are computed.

In this example, a multistage model becomes necessary if the distribution firm foresees additional periods where it is ready to change the location of the warehouses. In this example, suppose the firm decides that the opening of new warehouses can be decided after one year. A three-stage model can be constructed. The first stage would consist of decisions on warehouses to be built now. The second stage would consist of the distribution patterns between months 7 and 18 as well and new openings decided in month 12. The third stage would consist of distribution patterns between months 19 and 36.

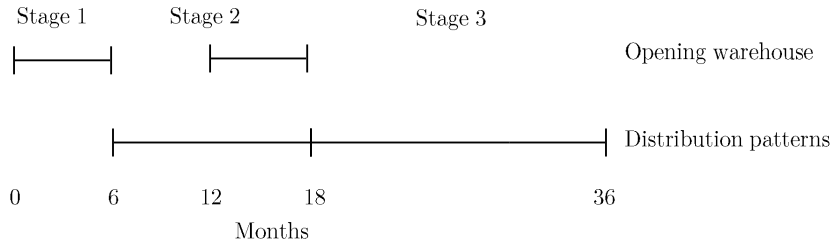

Fig. 2 Three-stage model decisions and times.

Let x^1 and $x^2(\omega_2)$ be the binary vectors representing opening warehouses in stages 1 and 2, respectively. Let $y^2(\omega_2)$ and $y^3(\omega_3)$ be the vectors representing the distribution decisions in stages 2 and 3, respectively, where ω_2 and ω_3 are the states of the world in stages 2 and 3. Assuming each warehouse can only have a fixed size M, the following model can be built:

$$\begin{aligned}
\max \; & -\sum_{j=1}^{n} c_j x_j^1 + \mathrm{E}_{\boldsymbol{\xi}_2} \max\{\sum_{i=1}^{m}\sum_{j=1}^{n} q_{ij}^2(\omega_2) y_{ij}^2(\omega_2) - \sum_{j=1}^{n} c_j^2(\omega_2) x_j^2(\omega_2) \\
& \qquad + \mathrm{E}_{\boldsymbol{\xi}_3|\boldsymbol{\xi}_2} \max[\sum_{i=1}^{m}\sum_{j=1}^{n} q_{ij}^3(\omega_3) y_{ij}^3(\omega_3)]\} \\
\text{s. t.} \; & \sum_{j=1}^{n} y_{ij}^2(\omega_2) \le 1 , && i = 1,\ldots,m , \\
& \sum_{i=1}^{m} d_i(\omega_2) y_{ij}^2(\omega_2) \le M x_j^1 , && j = 1,\ldots,n , \\
& \sum_{j=1}^{n} y_{ij}^3(\omega_3) \le 1 , && i = 1,\ldots,m , \\
& \sum_{i=1}^{m} d_i(\omega_3) y_{ij}^3(\omega_3) \le M(x_j^1 + x_j^2(\omega_2)) , && j = 1,\ldots,n , \\
& x_j^1 + x_j^2(\omega_2) \le 1 , && j = 1,\ldots,n , \\
& x_j^1, x_j^2(\omega_2) \in \{0,1\} , && j = 1,\ldots,n , \\
& y_{ij}^2(\omega_2), y_{ij}^3(\omega_3) \ge 0 , && i = 1,\ldots,m , \; j = 1,\ldots,n .
\end{aligned}$$

Multistage programs will be further studied in Section 3.4.

2.5 Random Variables and Risk Aversion

In our view, one can often classify random events and random variables in two major categories. In the first category, we would place uncertainties that recur frequently on a short-term basis. As an example, uncertainty may correspond to daily or weekly demands. This normally leads to a model similar to the one in Section 2.4, Case b (4.b), where allocation cannot be adjusted every time period. It follows that the expectation in the second stage somehow represents a mean over possible values of the random variables, of which many will occur. Thus, the expectation takes into account realizations that might not occur and many realizations that will occur. To fix ideas here, if in Model 4.b the units in the objective function are in a yearly basis and the randomness involves daily or weekly demands, one may expect that the value of the objective of stochastic model will closely match the realized total yearly revenue.

As one interesting example of a real-world application of a location model of this first category, we may recommend the paper by Psaraftis, Tharakan, and Ceder [1986]. It deals with the optimal location and size of equipment to fight oil spills. Occurrence and sizes of spills are random. The sizes of the spills are represented by a discrete random variable taking three possible values, corresponding to small, medium, or large spills. Sadly enough, spills are sufficiently frequent that the expectation may be considered close enough to the mean cost, as just described. Occurrence of spills at a given site is also random. It is described by a Poisson process. By making the assumption of non-concomitant occurrence of spills, all equipment is made available for each spill, which simplifies the second-stage descriptions compared to (4.14)–(4.17).

As a common example, consider revenue management decisions such as those considered in Problem 1.1 for an airline that must determine reservation controls for hundreds of daily flights. This area has become one of the most widespread applications of analytical methods to determining optimal choices under uncertain conditions (see Talluri and van Ryzin [2005]). Airlines routinely solve thousands of these stochastic programs each month and can reasonably expect to receive close to the expected revenue from their decisions each month (if not each day). Risk aversion has little affect in that case.

In the second category, we would place uncertainties that can be represented as scenarios, of which basically only one or a small number are realized. An example in a similar situation to the airline might be the problem of the organizers of the World Cup championship soccer game, which only occurs once every four years, to choose prices and seat allocations to maximize revenues but also to protect against possible losses. This consideration would also be the case in long-term models where scenarios represent the general trend or path of the variables. As already indicated, this is the spirit in which Model 4.c is built. In the second stage, among all scenarios over which expectation is taken, only one is realized. The objective function with only expected values may then be considered a poor representation of risk aversion, which is typically assumed in decision making (if we exclude gambling).

Starting from the von Neumann and Morgenstern [1944] theory of utility, this field of modeling preferences has been developed by economics. Models such as the mean-variance approach of Markowitz [1959] have been widely used. Other methods have been proposed based on mixes of mean-variance and other approaches (see, e.g, Ben-Tal and Teboulle [1986]). From a theoretical point of view, considering a nonlinear utility function transforms the problems into stochastic nonlinear programs, which can require more computational effort than linear versions. In practice, risk aversion is often captured with a piecewise-linear representation, as in the financial planning example in Section 1.2, to maintain a linear problem structure.

One interesting alternative to nonlinear utility models is to include risk aversion in a linear utility model under the form of a linear constraint, called *downside risk* (Eppen, Martin, and Schrage [1989]). The problem there is to determine the type and level of production capacity at each of several locations. Plants produce various types of cars and may be open, closed, or retooled. The demand for each type of car

in the medium term is random. The decisions about the locations and configurations of plants have to be made before the actual demands are known.

Scenarios are based on pessimistic, neutral, or optimistic realizations of demands. A scenario consists of a sequence of realizations for the next five years. The stochastic model maximizes the present value of expected discounted cash flows. The linear constraint on risk is as follows: the downside risk of a given scenario is the amount by which profit falls below some given target value. It is thus zero for larger profits. The expected downside risk is simply the expectation of the downside risk over all scenarios. The constraint is thus that the expected downside risk must fall below some level.

To give an idea of how this works, consider a two-stage model similar to (4.1)–(4.4) but in terms of profit maximization, by

$$\max z = c^T x + \mathrm{E}_{\boldsymbol{\xi}}[\max q^T(\omega) y(\omega)]$$

s.t. (4.2)–(4.4).

Then define the target level g on profit. The downside risk $u(\xi)$ is thus defined by two constraints:

$$u(\xi(\omega)) \geq g - q^T(\omega) y(\omega) \tag{5.1}$$

$$u(\xi(\omega)) \geq 0\,. \tag{5.2}$$

The constraint on expected downside risk is

$$\mathrm{E}_{\boldsymbol{\xi}} u(\boldsymbol{\xi}) \leq l\,, \tag{5.3}$$

where l is some given level. For a problem with a discrete random vector $\boldsymbol{\xi}$, constraint (5.3) is linear. Observe that (5.3) is in fact a first-stage constraint as it runs over all scenarios. It can be used directly in the extensive form. It can also be used indirectly in a sequential manner, by imposing such a constraint only when needed. This can be done in a way similar to the induced constraints for feasibility that we will study in Chapter 5.

2.6 Implicit Representation of the Second Stage

This book is mainly concerned with stochastic programs of the form (4.1)–(4.4), assuming that an adequate and computationally tractable representation of the recourse problem exists. This is not always the case. Two possibilities then exist that still permit some treatment of the problem:

- A closed form expression is available for the expected value function $\mathscr{Q}(x)$.
- For a given first-stage decision x, the expected value function $\mathscr{Q}(x)$ is computable.

These possibilities are described in the following sections.

a. A closed form expression is available for $\mathscr{Q}(x)$

We may illustrate this case by the *stochastic queue median* model (SQM) first proposed by Berman, Larson, and Chiu [1985] from which we take the following in a simplified form. The problem consists of locating an emergency unit (such as an ambulance). When a call arrives, there is a certain probability that the ambulance is already busy handling an earlier demand for ambulance service. In that event, the new service demand is either referred to a backup ambulance service or entered into a queue of other waiting "customers." Here, the first-stage decision consists of finding a location for the ambulance. The second stage consists of the day-to-day response of the system to the random demands. Assuming a first-in, first-out decision rule, decisions in the second stage are somehow automatic. On the other hand, the quality of response, measured, e.g., by the expected service time, depends on the first-stage decision. Indeed, when responding to a call, an ambulance typically goes to the scene and returns to the home location before responding to the next call. The time when it is unavailable for another call is clearly a function of the home location.

Let λ be the total demand rate, $\lambda \geq 0$. Let p_i be the probability that a demand originates from demand region i, with $\sum_{i=1}^{m} p_i = 1$. Let also $t(i,x)$ denote the travel time between location x and call i. On-scene service time is omitted for simplicity. Given facility location x, the expected response time is the sum of the mean-in-queue delay $w(x)$ and the expected travel time $\bar{t}(x)$,

$$\mathscr{Q}(x) = w(x) + \bar{t}(x) \ , \tag{6.1}$$

where

$$w(x) = \begin{cases} \frac{\lambda \bar{t}^{(2)}(x)}{2(1-\lambda \bar{t}(x))} & \text{if } \lambda \bar{t}(x) < 1 \ , \\ 0 & \text{otherwise,} \end{cases} \tag{6.2}$$

$$\bar{t}(x) = \sum_{i=1}^{m} p_i t(i,x) \ , \tag{6.3}$$

and

$$\bar{t}^{(2)}(x) = \sum_{i=1}^{m} p_i t^2(i,x) \ . \tag{6.4}$$

The global problem is then of the form:

$$\min_{x \in X} \mathscr{Q}(x) \ , \tag{6.5}$$

where the first-stage objective function is usually taken equal to zero and X represents the set of possible locations, which typically consists of a network.

It should be clear that no possibility exists to adequately describe the exact sequence of decisions and events in the so-called second stage and that the expected recourse $\mathcal{Q}(x)$ represents the result of a computation assuming the system is in steady state.

b. For a given x, $\mathcal{Q}(x)$ is computable

The deterministic traveling salesperson problem (TSP) consists of finding a Hamiltonian tour of least cost or distance. Following a Hamiltonian tour means that the traveling salesperson starts from her home location, visits all customers, (say $i = 1, \ldots, m$) exactly, and returns to the home location.

Now, assume each customer has a probability p_i of being present. A full optimization that would allow the salesperson to decide the next customer to visit at each step would be a difficult multistage stochastic program. A simpler two-stage model, known as *a priori optimization* is as follows: in the first-stage, an a priori Hamiltonian tour is designed. In the second stage, the a priori tour is followed by skipping the absent customers. The problem is to find the tour with minimal expected cost (Jaillet [1988]).

The exact representation of such a second-stage recourse problem as a mathematical program with binary decision variables might be possible in theory but would be so cumbersome that it would be of no practical value. On the other hand, the expected length of the tour (and thus $\mathcal{Q}(x)$) is easily computed when the tour (x) is given.

Let c_{ij} be the distance between i and j. Assume for simplicity of notation that the given tour is $\{0, 1, 2, \ldots, n, 0\}$ where 0 is the depot.

Define $t(k)$ as the expected length from k till the depot if k is present. Thus we search for $Q(x) = t(0)$.

Start with $t(n+1) = 0$ and $t(n) = c_{n0}$. Let $p_0 = 1$ and $c_{in+1} = c_{i0}$. Then

$$t(k) = \sum_{r=0}^{n-k} \prod_{j=1}^{r} (1 - p_{k+j})\, p_{k+r+1} (c_{kk+r+1} + t(k+r+1)) \quad \text{for } k = n-1, \ldots, 0,$$

where the condensed product is equal to 1 if $r = 0$.

This calculation is a backward recursion: assuming k is present, it considers the next present customer to be $k+r+1$ (and thus $k+1$ to $k+r$ being absent) for all possible successors ($k+1$ to $n+1 := 0$).

2.7 Probabilistic Programming

In probabilistic programming, some of the constraints or the objective are expressed in terms of probabilistic statements about first-stage decisions. The description of second-stage or recourse actions is thus avoided. This is particularly useful when the cost and benefits of second-stage decisions are difficult to assess.

For some probabilistic constraints, it is possible to derive a deterministic linear equivalent. A first example was given in Section 1.3. We now detail two other examples where a deterministic linear equivalent is obtained and one where it is not.

a. Deterministic linear equivalent: a direct case

Consider Exercise 1.6.1. An airline wishes to partition a plane of 200 seats into three categories: first, business, economy. Now, assume the airline wishes a special guarantee for its clients enrolled in its loyalty program. In particular, it wants 98% probability to cover the demand of first-class seats and 95% probability to cover the demand of business class seats (by clients of the loyalty program). First-class passengers are covered if they get a first-class seat. Business class passengers are covered if they get either a business or a first-class seat (upgrade). Assume weekday demands of loyalty-program passengers are normally distributed, say $\xi_F \sim N(16,16)$ and $\xi_B \sim N(30,48)$ for first-class and business, respectively. Also assume that the demands for first-class and business class seats are independent.

Let x_1 be the number of first-class seats and x_2 the number of business seats. The probabilistic constraints are simply

$$\mathrm{P}(x_1 \geq \xi_F) \geq 0.98, \tag{7.1}$$

$$\mathrm{P}(x_1 + x_2 \geq \xi_F + \xi_B) \geq 0.95\,. \tag{7.2}$$

Given the assumptions on the random variables, these probabilistic constraints can be transformed into a deterministic linear equivalent.

Constraint (7.1) can be written as $F_F(x_1) \geq 0.98$, where $F_F(\cdot)$ denotes the cumulative distribution of ξ_F. Now, the 0.98 quantile of the normal distribution is 2.054. As $\xi_F \sim N(16,16)$, $F_F(x_1) \geq 0.98$ is the same as $(x_1 - 16)/4 \geq 2.054$ or $x_1 \geq 24.216$. Thus, the probabilistic constraint (7.1) is equivalent to a simple bound.

Similarly, constraint (7.2) can be written as $F_{FB}(x_1 + x_2) \geq 0.95$, where $F_{FB}(\cdot)$ denotes the cumulative distribution of $\xi_F + \xi_B$. By the independence assumption and the properties of the normal distribution, $\xi_F + \xi_B \sim N(46,64)$. The 0.95 quantile of the standard normal distribution is 1.645. Thus, $F_{FB}(x_1 + x_2) \geq 0.95$ is the same as $(x_1 + x_2 - 46)/8 \geq 1.645$ or $x_1 + x_2 \geq 59.16$.

Thus, the probabilistic constraint (7.2) is equivalent to a linear constraint. We say that (7.2) has a linear deterministic equivalent. This is the desired situation with probabilistic constraints.

b. Deterministic linear equivalent: an indirect case

We now provide an example where finding the deterministic equivalent requires some transformation.

Consider the following covering location problem. Let $j = 1,\ldots,n$ be the potential locations with, as usual, $x_j = 1$ if site j is open and 0 otherwise, and c_j the investment cost. Let $i = 1,\ldots,m$ be the clients. Client i is served if there exists an open site within distance t_i. The distance between i and j is t_{ij}. Define $N_i = \{j \mid t_{ij} < t_i\}$ as the set of eligible sites for client i. The deterministic covering problem is

$$\min \sum_{j=1}^{n} c_j x_j \tag{7.3}$$

$$\text{s. t.} \sum_{j \in N_i} x_j \geq 1\,, \qquad i = 1,\ldots,m\,, \tag{7.4}$$

$$x_j \in \{0,1\}\,, \qquad j = 1,\ldots,n\,. \tag{7.5}$$

Taking again the case of an ambulance service, one site may cover more than one region or demand area. When a call is placed, the emergency units may be busy serving another call. Let q be the probability that no emergency unit is available at site j. For simplicity, assume this probability is the same for every site (see Toregas et al. [1971]). Then, the deterministic covering constraint (7.4) may be replaced by the requirement that P (at least one emergency unit from an open eligible site is available) $\geq \alpha$ where α is some confidence level, typically 90 or 95%. Here, the probability that none of the eligible sites has an available emergency unit is q to the power $\sum_{j \in N_i} x_j$, so that the probabilistic constraint is

$$1 - q^{\sum_{j \in N_i} x_j} \geq \alpha\,, \qquad i = 1,\ldots,m \tag{7.6}$$

or

$$q^{\sum_{j \in N_i} x_j} \leq 1 - \alpha\,.$$

Taking the logarithm on both sides, one obtains

$$\sum_{j \in N_i} x_j \geq b \tag{7.7}$$

with

$$b = \left\lceil \frac{\ln(1-\alpha)}{\ln q} \right\rceil\,, \tag{7.8}$$

where $\lceil a \rceil$ denotes the smallest integer greater than or equal to a. Thus, the probabilistic constraint (7.6) has a linear deterministic equivalent (7.7).

c. Deterministic nonlinear equivalent: the case of random constraint coefficients

The diet problem is a classical example of linear programming (discussed in Dantzig [1963] for the case in Stigler [1945]) . It consists of selecting a number of foods in order to get the cheapest menus that meet the daily requirements in the main nutrients (energy, protein, vitamins,...). Consider the data in the introductory example of Chvátal (1980). Polly wants to choose among six foods (oatmeal, chicken, eggs, whole milk, cherry pie and pork with beans). Each food has a given serving size; for instance, a serving of eggs is two large eggs and a serving of pork with beans is 260 grams. Each food has therefore a known content of nutrients. If we take the case of protein, the content is 4 , 32 , 13 , 8 , 4 and 14 grams (grams) of proteins, respectively, for the given serving sizes.

Let $x_1,\ldots,x_6$ represent the number of servings of each product per day. As Polly is a girl of 18 years of age, she needs 55 grams of protein per day. The protein constraint reads as follows:

$$4x_1 + 32x_2 + 13x_3 + 8x_4 + 4x_5 + 14x_6 \geq 55 \ .$$

(We omit here the other constraints and the objective function, which are very important to Polly but not central to our discussion.)

The same book later on contains an interesting discussion on the difficulty to get precise reliable RDA (recommended daily allowances) as well as precise nutrient contents per serving (Chvátal [Chapter 11, pp. 182–187]). Let us concentrate on this second aspect. It is indeed very unlikely that every large egg has exactly 6.5 grams of protein, or every serving of 260 grams of pork with beans has exactly 14 grams of protein. This implies that the nutrient content of each serving is in fact a random variable. Let $a_1,\ldots,a_6$ be the random content in proteins for the six products. The probabilistic constraint reads as follows:

$$\mathrm{P}(a_1\, x_1 + a_2\, x_2 + a_3\, x_3 + a_4\, x_4 + a_5\, x_5 + a_6\, x_6 \geq 55) \geq \alpha \ . \tag{7.9}$$

Let us now assume the contents of the products are normally distributed, say $a_i \sim N(\mu_i, \sigma_i^2)$, $i = 1,\ldots,6$. We can clearly assume independence between the six products. Then $a_1\, x_1 + a_2\, x_2 + a_3\, x_3 + a_4\, x_4 + a_5\, x_5 + a_6\, x_6 \sim N(\mu, \sigma^2)$ with $\mu = \mu_1\, x_1 + \mu_2\, x_2 + \mu_3\, x_3 + \mu_4\, x_4 + \mu_5\, x_5 + \mu_6\, x_6$ and $\sigma^2 = \sigma_1^2\, x_1^2 + \sigma_2^2\, x_2^2 + \sigma_3^2\, x_3^2 + \sigma_4^2\, x_4^2 + \sigma_5^2\, x_5 + \sigma_6^2\, x_6$.

Classical probabilistic analysis of the normal distribution implies that (7.9) is equivalent to

$$(55 - \mu)/\sigma \leq z_{1-\alpha}$$

with $z_{1-\alpha}$ the $(1-\alpha)$-quantile of the normal distribution. Taking $\alpha = 0.98$, the constraint reads $(55-\mu)/\sigma \leq -2.054$ or $\mu \geq 55 + 2.054 \cdot \sigma$. As $\sigma^2 = \sigma_1^2\, x_1^2 + \sigma_2^2\, x_2^2 + \sigma_3^2\, x_3^2 + \sigma_4^2\, x_4^2 + \sigma_5^2\, x_5^2 + \sigma_6^2\, x_6^2$, this constraint is non-linear and convex.

2.8 Modeling Exercise

In this section, we propose a modeling exercise and comment on a number of possible answers.

a. Presentation

Consider a production or assembly problem. It consists of producing two products, say A and B. They are obtained by assembling two components, say $C1$ and $C2$, in fixed quantities. The following table shows the components usage for the two products:

Components usage	A	B
$C1$	6	10
$C2$	8	5

Components are produced within the plant. Material (and / or operating) costs for $C1$ and $C2$ are 0.4 and 1.2, respectively. The level of production, or capacity, is related to the work-force and the equipment. Each unit of capacity costs 150 and 180 and can produce batches of 60 and 90 components, respectively for $C1$ and $C2$. Current capacity level is $(40, 20)$ batches and cannot be decreased. The total number of batches must not exceed 120. An integer number of batches is not requested here.

In the deterministic case, the demands and unit selling prices are certain and are as follows:

	A	B
Demand	500	200
Unit selling price	50	60

Unmet demand results in lost sales. This does does not imply any additional penalty.

1. Select adequate units for each data. Formulate and solve the deterministic problem.

Then, consider a number of stochastic variants. For the sake of comparison, in all cases, the random variables have expectations which are the corresponding deterministic values.

2. Stochastic prices (known demand).

The selling prices of A and B are described by a random vector, say $\zeta^T = (\zeta_1, \zeta_2)$. The rest of the data is unchanged. Formulate a recourse model in the following cases:

(a) ζ^T takes on the values $(54,56)$, $(50,60)$, and $(46,64)$ with probability 0.3, 0.4 and 0.3 respectively.
(b) ζ_1 takes on the values $(46,50,54)$ with probability 0.3, 0.4 and 0.3; ζ_2 takes on the values $(56,60,64)$ with probability 0.3, 0.4 and 0.3; ζ_1 and ζ_2 are independent.
(c) ζ_1 has a continuous uniform distribution in the range $[46,54]$; ζ_2 has a continuous uniform distribution in the range $[56,64]$; ζ_1 and ζ_2 are independent.
(d) ζ^T takes on the values $(70,50)$, $(50,60)$, $(30,70)$ with probability 0.3, 0.4 and 0.3.
(e) ζ_1 takes on the values $(30,50,70)$ with probability 0.3, 0.4 and 0.3; ζ_2 takes on the values $(50,60,70)$ with probability 0.3, 0.4 and 0.3; ζ_1 and ζ_2 are independent.

3. Stochastic demands (known prices).

The demand levels of A and B are described by a random vector, say $\eta^T = (\eta_1, \eta_2)$. The rest of the data is as in the deterministic model.

(a) Formulate and solve a recourse model when η^T takes on the values $(400,100)$, $(500,200)$, $(600,300)$ with probability 0.3, 0.4 and 0.3.
(b) Assume η_1 and η_2 are independent random variables with normal distributions, $\eta_1 \sim N(500,6000)$ and $\eta_2 \sim N(200,12000)$. Find the optimal solution of the recourse problem if the production of A and B is decided in the first-stage and there is no restriction at all on the number of batches of $C1$ and $C2$.
(c) Consider case (b). Add the restriction that the total number of batches must not exceed 120. Also ensure that the probability that the demand of B is covered must be larger than 80%.

4. Stochastic prices and demands.

Demands and prices are described by three scenarios $S1$, $S2$ and $S3$, as follows.

Demand level	$S1$	$S2$	$S3$
A	700	500	300
B	100	200	300
Unit selling price			
A	45	50	55
B	70	60	50

Formulate and solve a recourse model assuming the three scenarios have probability 0.3, 0.4 and 0.3 respectively.

5. Obtain *EVPI* and *VSS* for some relevant cases among these alternatives.

b. Discussion of solutions

1. Choice of units and deterministic model.

Units are as follows. First, define the unit of time. We may assume here data are given per day for example. Then, demand is the number of units of A and B per day. Selling prices are given as \$ per unit of A and B. The level of production is given by the number of batches (of 60 $C1$ and 90 $C2$) per day. Capacity cost must include work-force cost, operating costs, and depreciation per day. Material costs are \$ per component. The distinction among these costs is important for the stochastic model.

There is more than one formulation for the deterministic problem. The following formulation (M1) is useful in view of later stochastic models. Let

- $x_1 =$ number of batches of $C1$ available for production;
- $x_2 =$ number of batches of $C2$ available for production;
- $x_3 =$ number of units of A produced and sold per day;
- $x_4 =$ number of units of B produced and sold per day.

For batches of $C1$ and $C2$, the objective contains the daily capacity cost. For products A and B, it contains the selling price minus the material costs. (Each unit of A, e.g. has a selling price of \$50. It requests 6 units of C_1 and 8 units of C_2 for a total material cost of \$12. The difference is the objective coefficient 38.) The first two constraints state that the usage of components is smaller than the availability. The third constraint is the upper limit on the number of batches. Demand and capacity bounds follow.

$$\begin{aligned} \text{(M1)}\quad z = \ & \max -150x_1 - 180x_2 + 38x_3 + 50x_4 \\ & \text{s. t.}\ \ 6x_3 + 10x_4 \le 60x_1, \\ & \qquad 8x_3 + 5x_4 \le 90x_2, \\ & \qquad x_1 + x_2 \le 120, \\ & \qquad 40 \le x_1,\ 20 \le x_2,\ 0 \le x_3 \le 500,\ 0 \le x_4 \le 200. \end{aligned}$$

The optimal solution of (M1) is $z = 5800$, $x_1 = 220/3$, $x_2 = 140/3$, $x_3 = 400$, $x_4 = 200$. Product B is at the maximum corresponding to its demand. All 120 batches of capacity are used. The rest of the solutions follow.

A shorter formulation (M2) is to define two variables:

- $x_1 =$ number of units of A produced and sold per day;
- $x_2 =$ number of units of B produced and sold per day.

This formulation requires computing the margins of A and B. Each unit of A obtains the selling price of \$50. It requires 6 components $C1$ and 8 components $C2$ for a total material cost of \$12. It also requires 6/60 batches of capacity for $C1$ and 8/90 batches for $C2$ at a cost of \$31. The net margin for A is thus \$7 per unit. Similarly, the net margin for B is \$15 per unit. Note that this calculation of the margins of A and B is only valid if there is no unused capacity or unsold product, which is not always the case in a stochastic model. The first two constraints correspond to maintaining at least the existing capacity levels of 40 and 20 respectively. The third constraint corresponds to a maximal capacity level of 120 (each unit of A requires $6/60$ of $C1$ and $8/90$ of $C2$, or $17/90$ capacity units; each unit of B requires $10/60$ of $C1$ and $5/90$ of $C2$ or $20/90$ capacity units). The model also includes the demand constraints and reads as follows:

$$\begin{array}{lll} \text{(M2)} & z = & \max 7x_1 + 15x_2 \\ & \text{s. t.} & 6x_1 + 10x_2 \geq 2400, \\ & & 8x_1 + 5x_2 \geq 1800, \\ & & 17x_1 + 20x_2 \leq 10800, \\ & & 0 \leq x_1 \leq 500\,,\ 0 \leq x_2 \leq 200. \end{array}$$

This model has the same optimal solution, $z = 5800$, $x_1 = 400$, $x_2 = 200$, as previously. It is clear in (M2) that the margin of B is larger than that of A. Thus, product B is at the maximum corresponding to its demand. Product A is then reduced from the limit of 120 batches of capacity. The number of batches for $C1$ and $C2$ can be computed from the production of A and B, and are equal to $220/3$ and $140/3$, respectively.

2. Stochastic prices.

The essential modeling question concerns the timing of the decisions. Typically, the capacity decisions are made in the long run. They are first-stage decisions. Sales occur when the price is known. They are always second-stage decisions. Depending on the flexibility of the production process, the decision on the quantity to be produced may be first- or second-stage. We may thus distinguish between two formulations: production is first-stage (M3) or second-stage (M4).

2.1. Production is first-stage.

Let

- $x_1 =$ number of batches of $C1$ available for production;
- $x_2 =$ number of batches of $C2$ available for production;
- $x_3 =$ number of units of A produced per day;
- $x_4 =$ number of units of B produced per day;
- $y_1 =$ number of units of A sold per day;
- $y_2 =$ number of units of B sold per day;

$$
\begin{aligned}
z = \max\ & -150x_1 - 180x_2 - 12x_3 - 10x_4 \\
& + \mathrm{E}_{\boldsymbol{\xi}}(q_1(\omega)\, y_1(\omega) + q_2(\omega)\, y_2(\omega)) \\
\text{s. t. } & 6x_3 + 10x_4 \le 60x_1, \\
& 8x_3 + 5x_4 \le 90x_2, \\
& x_1 + x_2 \le 120 \\
& y_1(\omega) \le x_3,\ y_2(\omega) \le x_4, \\
& 40 \le x_1,\ 20 \le x_2,\ 0 \le x_3,\ 0 \le x_4, \\
& 0 \le y_1(\omega) \le 500,\ 0 \le y_2(\omega) \le 200,
\end{aligned}
$$

where $\boldsymbol{\xi}^T(\omega) = (q_1(\omega), q_2(\omega)) = \boldsymbol{\zeta}^T(\omega)$ corresponds to the selling prices.

In practice, it is customary to use a simplified notation where the dependence of y and $\boldsymbol{\xi}$ on ω is not made explicit. This (abuse of) notation is used here.

$$
\begin{aligned}
\text{(M3)}\quad z = \max\ & -150x_1 - 180x_2 - 12x_3 - 10x_4 \\
& + \mathrm{E}_{\boldsymbol{\xi}}(q_1\, y_1 + q_2\, y_2) \\
\text{s. t. } & 6x_3 + 10x_4 \le 60x_1, \\
& 8x_3 + 5x_4 \le 90x_2, \\
& x_1 + x_2 \le 120, \\
& y_1 \le x_3\ ,\ y_2 \le x_4, \\
& 40 \le x_1\ ,\ 20 \le x_2\ ,\ 0 \le x_3\ ,\ 0 \le x_4\ , \\
& 0 \le y_1 \le 500\ ,\ 0 \le y_2 \le 200,
\end{aligned}
$$

where $\boldsymbol{\xi}^T = (q_1, q_2) = \boldsymbol{\zeta}^T$.

We now transform (M3) as in Section 2.4a. Assuming q_1 and q_2 are never negative (a much needed assumption for the producer to survive), we obtain

$$
\begin{aligned}
\text{(M3')}\quad z = \max\ & -150x_1 - 180x_2 - 12x_3 - 10x_4 \\
& + \mathrm{E}_{\boldsymbol{\xi}}(q_1 \min\{x_3, 500\} + q_2 \min\{x_4, 200\}) \\
\text{s. t. } & 6x_3 + 10x_4 \le 60x_1, \\
& 8x_3 + 5x_4 \le 90x_2, \\
& x_1 + x_2 \le 120, \\
& 40 \le x_1\ ,\ 20 \le x_2\ ,\ 0 \le x_3\ ,\ 0 \le x_4,
\end{aligned}
$$

or

$$
\begin{aligned}
\text{(M3'')}\quad z = \max\ & -150x_1 - 180x_2 - 12x_3 - 10x_4 \\
& + \mu_1 \min\{x_3, 500\} + \mu_2 \min\{x_4, 200\} \\
\text{s. t. } & 6x_3 + 10x_4 \le 60x_1 \\
& 8x_3 + 5x_4 \le 90x_2 \\
& x_1 + x_2 \le 120 \\
& 40 \le x_1\ ,\ 20 \le x_2\ ,\ 0 \le x_3\ ,\ 0 \le x_4
\end{aligned}
$$

where (μ_1, μ_2) is the expectation of $\boldsymbol{\xi}^T$.

As (μ_1, μ_2) is equal to the deterministic selling prices $(50, 60)$, it is easy to show that (M3'') has the same optimal solution as the model (M1). This is true for each of the considered cases (a) to (e). To put it another way, if production is decided in the first-stage, the stochastic model where only the selling prices are

random can be replaced by a deterministic model with the random prices replaced by their expectations.

2.2. Production is second-stage

Let x_1 and x_2 be as in (M3) and

- $y_1 =$ number of units of A produced and sold per day;
- $y_2 =$ number of units of B produced and sold per day.

$$\begin{array}{lll} \text{(M4)} & z = & \max -150x_1 - 180x_2 + \mathrm{E}_{\boldsymbol{\xi}}(q_1\, y_1 + q_2\, y_2) \\ & \text{s. t.} & x_1 + x_2 \leq 120, \\ & & 6y_1 + 10y_2 \leq 60x_1, \\ & & 8y_1 + 5y_2 \leq 90x_2, \\ & & 40 \leq x_1\ ,\ 20 \leq x_2\ ,\ 0 \leq y_1 \leq 500\ ,\ 0 \leq y_2 \leq 200, \end{array}$$

where $\boldsymbol{\xi}^T = (q_1, q_2) = \boldsymbol{\zeta}^T - (12, 10)$ corresponds to selling prices minus material costs.

Before using formulation (M4), consider the deterministic formulation (M2). As long as the margin of B is larger than the margin of A and the margin of A remains positive, it is optimal to produce and sell $400\,A$ and $200\,B$. If this holds for all realizations of the selling prices, the same optimal solution is obtained for all realizations of $\boldsymbol{\zeta}$. It is thus the optimal solution of the stochastic model. (This will be elaborated in the comments after Proposition 5 of Chapter 4.) The expected margin is simply $\mathrm{E}_{\boldsymbol{\zeta}}(400\zeta_1 + 200\zeta_2 - 26,200)$ where $26,200$ is the total of the material and capacity costs for the daily production of $400\,A$ and $200\,B$. As (ζ_1, ζ_2) has expectation $(50, 60)$ as in the deterministic model, the expected margin is again the same as in the deterministic model. This situation occurs in cases (a), (b) and (c) of this exercise: the margin of A is $\zeta_1 - 43$, the margin of B is $\zeta_2 - 45$ and the relation $\zeta_2 - 45 \geq \zeta_1 - 43 \geq 0$ holds.

If at some point, the margin of A becomes negative or exceeds that of B, then (M4) is a truly stochastic model. For cases (d) and (e), there are values of the selling prices where the margin of A exceeds that of B. The stochastic model (M4) has to be solved.

In case (d), $\boldsymbol{\zeta}^T$ takes on the values $(70, 50)$, $(50, 60)$, $(30, 70)$ with probability 0.3, 0.4 and 0.3, respectively. First-stage optimal capacity decisions are $(x_1, x_2) = (69.167, 50.833)$. Second-stage optimal production and sale decisions (x_3, x_4) are $(500, 115)$, $(500, 115)$ and $(358.333, 200)$ for the three possible scenarios. The optimal objective value is $z = 5990$.

In case (e), the two random variables ζ_1 and ζ_2 are independent, taking three different values each. Thus, the second-stage must consider 9 realizations. The optimal solution is the same as in the deterministic case: first-stage decisions are $(x_1, x_2) = (73.333, 46.667)$, second-stage decisions are $(x_3, x_4) = (400, 200)$, with objective value $z = 5800$.

3. Stochastic demands.

(a) As in Question 2, the first modeling question is the timing of the decisions. Capacity decisions are made in the long run and are first-stage decisions. Sales occur when price is known and are second-stage. The decisions on the quantities to be produced may be first- or second-stage.

(a.1) Production is first-stage.

If production is first-stage, lost sales occur when demand exceeds production. What happens when production exceeds demand is problem dependent. In some situations, excess production may be held in inventory. This would be the case when the randomness represents day-to-day variations in demand. Then excess production is used later to compensate for possible lost sales. Randomness only results in inventory costs. On the other hand, for products such as perishable goods, production is lost ($C1$ and $C2$ could be flour and eggs, A and B could be bread and pastry, e.g.) and lost sales cannot be compensated. The same is true when the randomness describes a set of scenarios of which only one is realized. The scenarios could represent the uncertainty about the success of a new product. If a product is not successful, extra production is lost. If it is very successful, sales are lost to competitors if the production level is insufficient. Or, alternative actions are needed such as subcontracting or overtime.

We now present a formulation (M5) corresponding to a scenario situation (excess production is lost, lost sales are not compensated). The decision variables are the same as in (M3).

$$
\begin{array}{lll}
\text{(M5)} & z = & \max -150x_1 - 180x_2 - 12x_3 - 10x_4 \\
 & & \qquad + \mathrm{E}_{\boldsymbol{\xi}}(50y_1 + 60y_2) \\
 & \text{s. t.} & 6x_3 + 10x_4 \le 60x_1, \\
 & & 8x_3 + 5x_4 \le 90x_2, \\
 & & x_1 + x_2 \le 120, \\
 & & y_1 \le x_3\ ,\ y_2 \le x_4, \\
 & & 40 \le x_1\ ,\ 20 \le x_2\ ,\ 0 \le x_3\ ,\ 0 \le x_4\ , \\
 & & 0 \le y_1 \le d_1\ ,\ 0 \le y_2 \le d_2,
\end{array}
$$

where $\boldsymbol{\xi}^T = (d_1, d_2) = \boldsymbol{\eta}^T$ correspond to the demand level.

The first-stage optimal capacity decisions are $(x_1, x_2) = (56.667, 41.111)$. The second-stage optimal production and sale decisions (x_3, x_4) are $(400, 100)$ in the three possible scenarios. The optimal objective value is $z = 4300$. Observe that the production is set to meet the lowest possible demand.

(a.2) Production is second-stage.

If production is second-stage, lost sales occur when the available production capacities are insufficient to cover the demand. Excess production does not occur as the level of production can be adjusted to the downside. The decision variables are the same as in (M4). Formulation (M6) reads as follows:

$$\begin{array}{lll} \text{(M6)} & z = & \max -150x_1 - 180x_2 + \mathrm{E}_{\boldsymbol{\xi}}(38y_1 + 50y_2) \\ & \text{s. t.} & x_1 + x_2 \leq 120, \\ & & 6y_1 + 10y_2 \leq 60x_1, \\ & & 8y_1 + 5y_2 \leq 90x_2, \\ & & 40 \leq x_1\ ,\ 20 \leq x_2\ ,\ 0 \leq y_1 \leq d_1\ ,\ 0 \leq y_2 \leq d_2, \end{array}$$

where $\boldsymbol{\xi}^T = (d_1, d_2) = \boldsymbol{\eta}^T$ corresponds to the demand level.

The first-stage optimal capacity decisions are $(x_1, x_2) = (67.083, 41.111)$. The second-stage optimal production and sale decisions (x_3, x_4) are $(400, 100)$, $(337.5, 200)$ and $(337.5, 200)$ for the three possible scenarios. The optimal objective value is $z = 4575$. Observe that the capacity limit of 120 batches is not fully used.

(b) We consider a variant of formulation (M5) where the only constraints on x_1 and x_2 are the components usage:

$$\begin{array}{lll} \text{(M7)} & z = & \max -150x_1 - 180x_2 - 12x_3 - 10x_4 \\ & & \qquad + \mathrm{E}_{\boldsymbol{\xi}}(50\min\{x_3, d_1\} + 60\min\{x_4, d_2\}) \\ & \text{s. t.} & 6x_3 + 10x_4 \leq 60x_1, \\ & & 8x_3 + 5x_4 \leq 90x_2, \\ & & 0 \leq x_1\ ,\ 0 \leq x_2\ ,\ 0 \leq x_3\ ,\ 0 \leq x_4, \end{array}$$

where $\boldsymbol{\xi}^T = (d_1, d_2) = \boldsymbol{\eta}^T$ corresponds to the demand level.

Clearly, the two constraints are always tight. Replacing x_1 by $(6x_3 + 10x_4)/60$ and x_2 by $(8x_3 + 5x_4)/90$, the model becomes

$$\begin{aligned} z = \max\{ & -43x_3 - 45x_4 + \mathrm{E}_{\boldsymbol{\xi}}(50\min\{x_3, d_1\} \\ & + 60\min\{x_4, d_2\}) \mid 0 \leq x_3,\ 0 \leq x_4\}\ , \end{aligned}$$

or

$$\begin{aligned} \text{(M7')}\quad z = \max\{ & -43x_3 + 50\mathrm{E}_{\boldsymbol{\xi}_1}\min\{x_3, \boldsymbol{\xi}_1\} - 45x_4 \\ & + 60\mathrm{E}_{\boldsymbol{\xi}_2}\min\{x_4, \boldsymbol{\xi}_2\} \mid 0 \leq x_3\ ,\ 0 \leq x_4\}. \end{aligned}$$

This optimization is separable in x_3 and x_4. Both variables will be nonzero. So, we are searching twice for the unconstrained minimum of an expression of the form $-a\,x + b\mathscr{Q}(x)$, with $\mathscr{Q}(x) = \mathrm{E}_{\boldsymbol{\xi}}\min\{x, \boldsymbol{\xi}\}$ and $\boldsymbol{\xi} \sim N(\mu, \sigma^2)$. From Exercise 2.8.2, we obtain that $\mathscr{Q}'(x) = 1 - F(x)$. As $\mathscr{Q}''(x) = -f(x)$, the second-order conditions are satisfied. Thus the unconstrained minimum is obtained for $\mathscr{Q}'(x) = a/b$, i.e. $1 - F(x) = a/b$.

Denote by $F_i(\cdot)$ the cumulative distribution of $\boldsymbol{\xi}_i$, $i = 1, 2$. For x_3, the unconstrained optimum satisfies $1 - F_1(x_3) = 43/50$, or $F_1(x_3) = 0.14$. It corresponds to a quantile $q = -1.08$ and a decision $x_3 = 500 - 1.08\sqrt{6000} = 416.34$. For x_4, we have $1 - F_2(x_4) = 45/60$, or $F_2(x_4) = 0.25$. It corresponds to a

quartile $q = -0.675$ and a decision $x_4 = 200 - 0.675\sqrt{12000} = 126.06$. For the sake of comparison, we may compute $x_1 = (6x_3 + 10x_4)/60 = 62.644$ and $x_2 = (8x_3 + 5x_4)/90 = 44.011$. Also, using the closed form expression of $\mathscr{Q}(x)$, (see again Exercise 2.8.2), one can obtain the optimal value of z.

(c) Requesting that the probability that the demand of B is covered must be larger than 80% is $\mathrm{P}(x_4 \geq \boldsymbol{\xi}_2) \geq 0.8$ or $F_2(x_4) \geq 0.8$. The 0.8 quantile is 0.84. Thus, $F_2(x_4) \geq 0.8$ is equivalent to $(x_4 - \mu_2)/\sigma_2 \geq 0.8$, or $x_4 \geq 200 + 0.84\sqrt{12000}$, or $x_4 \geq 292.02$.

The model to solve is:

$$\begin{aligned}(\text{M8}) \quad z = \max\{&-43x_3 + 50\mathrm{E}_{\boldsymbol{\xi}_1} \min\{x_3, \boldsymbol{\xi}_1\} - 45x_4 \\ &+ 60\mathrm{E}_{\boldsymbol{\xi}_2} \min\{x_4, \boldsymbol{\xi}_2\} \mid 0 \leq x_3, \\ &292.02 \leq x_4\ ,\ 17x_3 + 20x_4 \leq 10800\},\end{aligned}$$

where the constraint on the 120 batches has been transformed as in (M2).

By applying the Karush-Kuhn-Tucker conditions (see Review Section 2.11c.), one can show that $(x_3, x_4) = (291.74, 292.02)$ is the optimal solution.

4. Just as in the previous cases, there are two possible formulations as the production decisions may be first- or second-stage. Model (M9) corresponds to first-stage production while (M10) corresponds to second-stage production.

$$\begin{aligned}(\text{M9}) \quad z = \ &\max -150x_1 - 180x_2 - 12x_3 - 10x_4 \\ &+ \mathrm{E}_{\boldsymbol{\xi}}(q_1\, y_1 + q_2\, y_2) \\ \text{s. t.} \ \ &6x_3 + 10x_4 \leq 60x_1, \\ &8x_3 + 5x_4 \leq 90x_2, \\ &x_1 + x_2 \leq 120, \\ &y_1 \leq x_3\ ,\ y_2 \leq x_4, \\ &40 \leq x_1\ ,\ 20 \leq x_2\ ,\ 0 \leq x_3\ ,\ 0 \leq x_4\ , \\ &0 \leq y_1 \leq d_1\ ,\ 0 \leq y_2 \leq d_2,\end{aligned}$$

where $\boldsymbol{\xi}^T = (q_1, q_2, d_1, d_2)$, with q_1 and q_2 the selling prices and d_1 and d_2 the demands jointly defined in a scenario. Thus $\boldsymbol{\xi}^T = (45, 70, 700, 100), (50, 60, 500, 200)$ and $(55, 50, 300, 300)$ with probability 0.3, 0.4, and 0.3 respectively. The optimal solution is $z = 3600$, $(x_1, x_2) = (46.667, 32.222)$ with corresponding $(x_3, x_4) = (300, 100)$. The second-stage decisions are $(y_1, y_2) = (300, 100)$ in all three scenarios. As the production cannot be adapted to the demand, the optimal solution is to plan for the lowest demand and the expected margin is low.

$$\begin{aligned}(\text{M10}) \quad z = \ &\max -150x_1 - 180x_2 + \mathrm{E}_{\boldsymbol{\xi}}(q_1\, y_1 + q_2\, y_2) \\ \text{s. t.} \ \ &x_1 + x_2 \leq 120 \\ &6y_1 + 10y_2 \leq 60x_1, \\ &8y_1 + 5y_2 \leq 90x_2, \\ &40 \leq x_1\ ,\ 20 \leq x_2\ ,\ 0 \leq y_1 \leq d_1\ ,\ 0 \leq y_2 \leq d_2,\end{aligned}$$

where $\boldsymbol{\xi}^T = (q_1, q_2, d_1, d_2)$ with q_1 and q_2 the selling prices minus the material costs and d_1 and d_2 the demands. Thus, $\boldsymbol{\xi}^T = (33, 60, 700, 100)$, $(38, 50, 500, 200)$ and $(43, 40, 300, 300)$ with probability 0.3, 0.4, and 0.3. The optimal solution is $z = 4048.75$, $(x_1, x_2) = (73.333, 46.667)$. The second-stage decisions are $(y_1, y_2) = (462.5, 100)$, $(400, 200)$ and $(300, 260)$ in the three scenarios. While

obtaining the optimal solution of (M10) with your favorite LP solver, you may observe that there is a high shadow price for the maximum number of batches.

Exercises

1. Consider Exercise 1 of Section 1.6.

   (a) Show that this is a two-stage stochastic program with first-stage integer decision variables. Observe that, for a random variable with integer realizations, the second-stage variables can be assumed continuous because the optimal second-stage decisions are automatically integer. Assume that Northam revises its seating policy every year. Is a multistage program needed?
   (b) Assume that the data in Exercise 1 correspond to the demand for seat reservations. Assume that there is a 50% probability that all clients with a reservation effectively show up and that 10 or 20% no-shows occur with equal probability. Model this situation as a three-stage program, with first-stage decisions as before, second-stage decisions corresponding to the number of accepted reservations, and third-stage decisions corresponding to effective seat occupation. Show that the third stage is a simple recourse program with a reward for each occupied seat and a penalty for each denied reservation.
   (c) Consider now the situation where the number of seats has been fixed to 12, 24, and 140 for the first class, business class, and economy class, respectively. Assume the top management estimates the reward of an occupied seat to be 4, 2, and 1 in the first class, business class, and economy class, respectively, and the penalty for a denied reservation is 1.5 times the reward. Model the corresponding problem as a recourse program. Find the optimal acceptance policy with the data of Exercise 1 in Section 1.6 and no-shows as in (b) of the current exercise. To simplify, assume that passengers with a denied reservation are not seated in a higher class even if a seat is available there.

2. Let $\mathscr{Q}(x) = \mathrm{E}_{\boldsymbol{\xi}} \min\{x, \boldsymbol{\xi}\}$.

   (a) Obtain a closed form expression for $\mathscr{Q}(x)$ when $\boldsymbol{\xi}$ follows a Poisson distribution.
   (b) Obtain a closed form expression for $\mathscr{Q}(x)$ when $\boldsymbol{\xi}$ follows a normal distribution. (Hint: for a normal distribution, the relation $\xi f(\boldsymbol{\xi}) = \mu f(\boldsymbol{\xi}) - \sigma^2 f'(\boldsymbol{\xi})$ holds for any given $\boldsymbol{\xi}$.)
   (c) Assume $\boldsymbol{\xi}$ has a continuous distribution. Show that $\mathscr{Q}'(x) = 1 - F(x)$.

3. Consider an airplane with x seats. Assume passengers with reservations show up with probability 0.90, independently of each other.

(a) Let $x = 40$. If 42 passengers receive a reservation, what is the probability that at least one is denied a seat.
(b) Let $x = 50$. How many reservations can be accepted under the constraint that the probability of seating all passengers who arrive for the flight is greater than 90% ?

4. Consider the design problem in Section 1.4. Suppose the design decision does not completely specify x in (1.4.1), but the designer only knows that if a value $\hat{x}$ is specified then $x \in [.99\hat{x}, 1.01\hat{x}]$. Suppose a uniform distribution for x is assumed initially on this interval. How would the formulation in Section 1.4 be modified to account for information as new parts are produced?
5. Consider the example in Section 2.7a.

(a) One may feel uncomfortable with the deterministic linear equivalent yielding a non-integer number of seats. Show how to cope with this.
(b) One may also feel uncomfortable with the demands represented by normal distributions. Show that deterministic linear equivalents are also obtained if $\xi_F \sim P(3)$ and $\xi_B \sim P(4)$ for example.

2.9 Alternative Characterizations and Robust Formulations

While the main focus of this book is on problems that can be represented in the form in (4.1–4.4) as stochastic linear programs, this formulation can still represent a wide range of risk preferences. As observed in Section 2.5, an expected von Neumann-Morgenstern concave utility objective can be represented as a piecewise-linear function. For example, if the utility function is $U(-q(\omega)^T y(\omega) - \gamma)$ where γ is a scaling parameter for fitting the function, then an additional set of variables $y'(\omega)_j$ with bounds u_j and slopes $-q'_j$ such that $0 \le y'(\omega)_j \le u_j$, $-q'_j \ge -q'_{j+1}$, and for $j = 0, \ldots, J$ can be defined with an additional linear constraint as:

$$-y'_0 + \sum_{j=1} y'_j(\omega) - q(\omega)^T y(\omega) = \gamma, \tag{9.1}$$

and with a new recourse function objective to minimize

$$-q'_0 y_0(\omega) + \sum_{j=1}^{J} q'_j y(\omega). \tag{9.2}$$

The parameters γ, q', and u' can be chosen to fit the utility function U as closely as desired while maintaining the same linear optimization form as in (4.1–4.4).

Other risk–measures may be included in the objective and as fixed or probabilistic constraints. A common use of these constraints in financial applications is to maximize expected return subject to a constraint on *value–at–risk* (*VaR*), the greatest loss in portfolio value that can occur with a given probability α, defined as

$$VaR_\alpha(q(\omega)^T y(\omega)) = \min\{t | \mathrm{P}(q(\omega)^T y(\omega) \le t) \ge \alpha\}. \tag{9.3}$$

A *VaR* constraint to limit losses to be no greater than $\bar{t}$ with probability at most α can then be written as

$$\mathrm{P}(q(\omega)^T y(\omega) \le \bar{t}) \ge \alpha, \tag{9.4}$$

since this ensures that $VaR_\alpha(q(\omega)^T y(\omega)) \le \bar{t}$.

A criticism of *VaR* as a measure of risk is that it does not have the useful property of subadditivity such that the *VaR* of the sum of two random variables is at most the the sum of the *VaR*'s of each individual random variable. The subadditive property is part of the set of axioms that define coherent risk measures (see Artzner, Delbaen, Eber, and Heath [1999]), such that $R(\cdot)$ is a *coherent risk measure* if the following hold:

Definition 2.1. 1. *subadditivity*: $R(\xi+\zeta) \le R(\xi)+R(\zeta)$ for any random variables ξ and ζ;
2. *positive homogeneity (of degree one)*: $R(\lambda\xi) = \lambda R(\xi)$ for all $\lambda \ge 0$;
3. *monotonicity*: $R(\xi) \le R(\zeta)$ whenever $\xi \preceq \zeta$, where $\preceq$ indicates first-order stochastic dominance,i.e., $\mathrm{P}(\xi \le t) \ge \mathrm{P}(\zeta \le t), \forall t$;
4. *translation invariance*: $R(\xi+t) = R(\xi)+t$ for any $t \in \Re$.

A related risk measure to *VaR*, called the *conditional value-at-risk* (*CVaR*), can be defined to avoid the potential problems of a non-subadditive risk measure by taking the conditional expectations over losses in excess of *VaR*. For random loss ξ with distribution function P, the α-confidence level is then defined as

$$CVaR_\alpha(\xi) = \mathrm{E}_{P_\alpha}[\xi], \tag{9.5}$$

where P_α is the distribution function defined by

$$P_\alpha(t) = \begin{cases} 0 & \text{if } t < VaR_\alpha(\xi); \\ \frac{P(t)-\alpha}{1-\alpha} & \text{if } t \ge VaR_\alpha(\xi). \end{cases} \tag{9.6}$$

As shown by Rockafellar and Uryasev [2000,2002], *CVaR* satisfies all of the axioms for a coherent risk measure (Exercise 3) and has a convenient representation as the solution to the following optimization problem:

$$CVaR_\alpha(\xi) = \min_t \; t + \frac{1}{1-\alpha}\mathrm{E}_P[(\xi - t)^+], \tag{9.7}$$

which can also be written as the linear program:

$$\min \; t + \frac{1}{1-\alpha}\mathrm{E}_P[y(\omega)] \tag{9.8}$$

$$\text{s. t. } \xi(\omega) - y(\omega) \le t, \text{ a. s.} \tag{9.9}$$

$$y(\omega) \ge 0, \text{ a. s.} \tag{9.10}$$

With the representation in (9.8), a risk constraint to limit $CVaR_\alpha$ to be less than $\bar{t}$ can be constructed similarly to the probabilistic constraint in (9.4) or the downside risk constraint in (5.3) with additional linear constraints and variables $y'(\omega)$ as follows:

$$t + \frac{1}{1-\alpha}\mathrm{E}[y'(\omega)] \leq \bar{t} \tag{9.11}$$

$$-t + q(\omega)^T y(\omega) - y'(\omega) \leq 0,\ a.s., \tag{9.12}$$

$$y'(\omega) \geq 0, a.s. \tag{9.13}$$

The use of coherent risk measures has another useful interpretation that R is a coherent risk measure if and only if there is a class of probability measure $\mathscr{P}$ such that $R(\xi)$ equals the highest expectation of ξ with respect to members of this class (see Huber [1981]):

$$R(\xi) = \sup_{P \in \mathscr{P}} \mathrm{E}_P[\xi]. \tag{9.14}$$

This representation provides a worst-case view of the risk, which is discussed in more detail in Chapter 8.

One worst-case version of the approach in (9.14) is to let $\mathscr{P}$ correspond to any distribution with support in a given range or uncertainty set. This worst-case type of risk-measure is called *robust* so that optimization models including a robust risk-measure of this form are *robust optimization* models. A robust version of the two–stage stochastic program can then be written as:

$$\min_x \max_{\xi \in \Xi} c^T x + Q(x, \xi) \tag{9.15}$$

$$\text{s. t.} \qquad Ax = b,$$

$$x \geq 0.$$

Depending on the properties of Ξ, robust optimization models can be tractable linear or conic optimization models. A variety of results in the area appear in Bertsimas and Sim [2006], Ben-Tal and Nemirovski [2002] with multi-period extensions also appearing, for example, in Ben-Tal, Boyd, and Nemirovski [2006] and Bertsimas, Iancu, and Parrilo [2010].

Exercises

1. Give an example of random variables ξ and ζ where $VaR_\alpha(\xi+\zeta) > VaR_\alpha(\xi) + VaR_\alpha(\zeta)$ for some $0 < \alpha < 1$.
2. Show that VaR satisfies the axioms of positive homogeneity, monotonicity, and translation independence.
3. Show that $CVaR$ satisfies all of the axioms for a coherent risk measure.
4. Give a class of probability distribution $\mathscr{P}$ such that $CVaR$ solves (9.14).

5. Find the robust formulation of the two-stage model (9.15) when uncertainty is only in the right-hand side $h \in \Xi = [l,u]$, a rectangular region.
6. Find the robust formulation of the two-stage model (9.15) when uncertainty is only in the right-hand side $h \in \Xi = \{h|(h-\mu)^T V(h-\mu) \leq 1\}$, an ellipsoidal region.

2.10 Relationship to Other Decision-Making Models

The stochastic programming models considered in this section illustrate the general form of a stochastic program. While this form can apply to virtually all decision-making problems with unknown parameters, certain characteristics typify stochastic programs and form the major emphasis of this book. In general, stochastic programs are generalizations of deterministic mathematical programs in which some uncontrollable data are not known with certainty. The key features are typically many decision variables with many potential values, discrete time periods for decisions, the use of expectation functionals for objectives, and known (or partially known) distributions. The relative importance of these features contrasts with similar areas, such as statistical decision theory, decision analysis, dynamic programming, Markov decision processes, and stochastic control. In the following subsections, we consider these other areas of study and highlight the different emphases.

a. Statistical decision theory and decision analysis

Wald [1950] developed much of the foundation of optimal statistical decision theory (see also DeGroot [1970] and Berger [1985]). The basic motivation was to determine best levels of variables that affect the outcome of an experiment. With variables x in some set X, random outcomes, $\omega \in \Omega$, an associated distribution, $F(\omega)$, and a reward or loss associated with the experiment under outcome ω of $r(x,\omega)$, the basic problem is to find $x \in X$ to

$$\max \mathrm{E}_{\omega}[r(x,\omega)|F] = \max \int_{\omega} r(x,\omega)dF(\omega). \tag{10.1}$$

The problem in (10.1) is also the fundamental form of stochastic programming. The major differences in emphases between the fields stem from underlying assumptions about the relative importance of different aspects of the problem.

In stochastic programming, one generally assumes that difficulties in finding the form of the function r and changes in the distribution F as a function of actions are small in comparison to finding the expectations with known distributions and an optimal value x with all other information known. The emphasis is on finding a solution after a suitable problem statement in the form (10.1) has been found.

For example, in the simple farming example in Section 1.1, the number of possible planting configurations (even allowing only whole-acre lots) is enormous. Enumerating the possibilities would be hopeless. Stochastic programming avoids such inefficiencies through an optimization process.

We might suppose that the fields or crop varieties are new and that the farmer has little direct information about yields. In this case, the yield distribution would probably start as some prior belief but would be modified as time went on. This modification and possible effects of varying crop rotations to obtain information are the emphases from statistical decision theory. If we assumed that only limited variation in planting size (such as 50-acre blocks) was possible, then the combinatorial nature of the problem would look less severe. Enumeration might then be possible without any particular optimization process. If enumeration were not possible, the farmer might still update the distributions and objectives and use stochastic programming procedures to determine next year's crops based on the updated information.

In terms of (10.1)), statistical decision theory places a heavy emphasis on changes in F to some updated distribution $\hat{F}_x$ that depends on a partial choice of x and some observations of ω. The implied assumption is that this part of the analysis dominates any solution procedure, as when X is a small finite set that can be enumerated easily.

Decision analysis (see, e.g., Raiffa [1968]) can be viewed as a particular part of optimal statistical decision theory. The key emphases are often on acquiring information about possible outcomes, on evaluating the utility associated with various outcomes, and on defining a limited set of possible actions (usually in the form of a decision tree). For example, consider the capacity expansion problem in Section 1.3. We considered a wide number of alternative technology levels and production decisions. In that model, we assumed that demand in each period was independent of the demand in the previous period. This characteristic gave the block separability property that can allow efficient solutions for large problems.

A decision analytic model might apply to the situation where an electric utility's demand depends greatly on whether a given industry locates in the region. The decision problem might then be broken into separate stochastic programs depending on whether the new industry demand materializes and whether the utility starts on new plants before knowing the industry decision. In this framework, the utility first decides whether to start its own projects. The utility then observes whether the new industry expands into the region and faces the stochastic program form from Section 1.4 with four possible input scenarios about the available capacity when the industry's location decision is known (see Figure 3).

The two stochastic programs given each initial decision allow for the evaluation of expected utility given the two possible outcomes and two possible initial decisions. The actual initial decision taken on current capacity expansion would then be made by taking expectations over these two outcomes.

Separation into distinct possible outcomes and decisions and the realization of different distributions depending on the industry decision give this model a decision analysis framework. In general, a decision analytic approach would probably also consider multiple attributes of the capacity decisions (for example, social costs for a

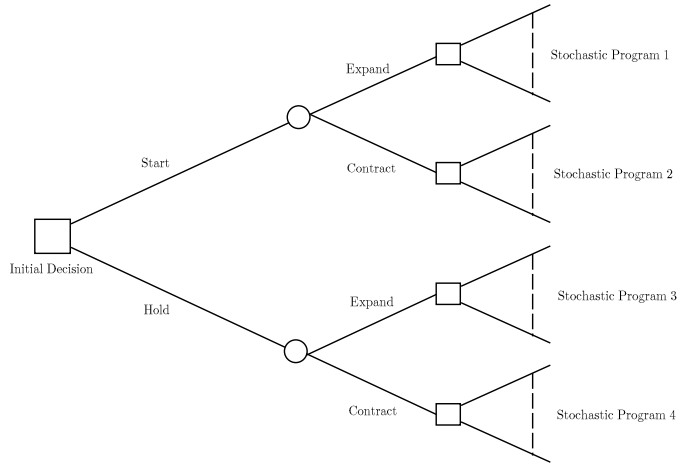

Fig. 3 Decision tree for utility with stochastic programs on leaves.

given location) and would concentrate on the value of risk in the objective. It would probably also entail consideration of methods for obtaining information about the industry's decision and contingent decisions based on the outcomes of these investigations. Of course, these considerations can all be included in a stochastic program, but they are not typically the major components of a stochastic programming analysis.

b. Dynamic programming and Markov decision processes

Much of the literature on stochastic optimization considers dynamic programming and Markov decision processes (see, e.g., Heyman and Sobel [1984], Bellman [1957], Ross [1983], and Kall and Wallace [1994] for a discussion relating to stochastic programming). In these models, one searches for optimal actions to take at generally discrete points in time. The actions are influenced by random outcomes and carry one from some state at some stage t to another state at stage $t+1$. The emphasis in these models is typically in identifying finite (or, at least, low-dimensional) state and action spaces and in assuming some Markovian structure (so that actions and outcomes only depend on the current state).

With this characterization, the typical approach is to form a backward recursion resulting in an optimal decision associated with each state at each stage. With large state spaces, this approach becomes quite computationally cumbersome although it does form the basis of many stochastic programming computation schemes as given in Chapter 6. Another approach is to consider an infinite horizon and use discounting

to establish a stationary policy (see Howard [1960] and Blackwell [1965]) so that one need only find an optimal decision associated with a state for any stage.

A typical example of this is in investment. Suppose that instead of saving for a specific time period in the example of Section 1.2, you wish to maximize a discounted expected utility of wealth in all future periods. In this case, the state of the system is the amount of wealth. The decision or action is to determine what amount of the wealth to invest in stock and bonds. We could discretize to varying wealth levels and then form a problem as follows:

$$\max \sum_{t=1}^{\infty} \rho^t \mathrm{E}[q\mathbf{y}(t) - r\mathbf{w}(t)] \tag{10.2}$$

$$\begin{aligned} \text{s. t.} \quad x(1,1) + x(2,1) &= b, \\ \boldsymbol{\xi}(1,t)\mathbf{x}(1,t) + \boldsymbol{\xi}(2,t)\mathbf{x}(2,t) - \mathbf{y}(t) + \mathbf{w}(t) &= G, \\ \boldsymbol{\xi}(1,t)\mathbf{x}(1,t) + \boldsymbol{\xi}(2,t)\mathbf{x}(2,t) &= \mathbf{x}(1,t+1) + \mathbf{x}(2,t+1), \\ \mathbf{x}(i,t), \mathbf{y}(t), \mathbf{w}(t) \geq 0, \quad \mathbf{x} &\in \mathscr{N}, \end{aligned}$$

where $\mathscr{N}$ is the space of nonanticipative decisions and ρ is some discount factor. This approach could lead to finding a stationary solution to

$$z(b) = \max_{x(1)+x(2)=b} \{\mathrm{E}[-q(G - \boldsymbol{\xi}(1)x(1) - \boldsymbol{\xi}(2)x(2))^-$$

$$-r(G - \boldsymbol{\xi}(1)x(1) - \boldsymbol{\xi}(2)x(2))^+ + \rho \mathrm{E}[z(\boldsymbol{\xi}(1)x(1) + \boldsymbol{\xi}(2)x(2))]\}. \tag{10.3}$$

Again, problem (10.2) fits the general stochastic programming form, but particular solutions as in (10.3) are more typical of Markov decision processes. These are not excluded in stochastic programs, but stochastic programs generally do not include the Markovian assumptions necessary to derive (10.3).

c. Machine learning and online optimization

While Markov decision problems have the general character of stochastic programs of including a distribution over some set of uncertain parameters, online optimization problems involve a changing objective (perhaps chosen adversarially) without knowledge of the choice and only considering the history of observations. The objective is then to choose $x^1, x^2, \ldots$ sequentially to minimize

$$\sum_{t=1}^{H} f^t(x^t), \tag{10.4}$$

where H may increase without bound and each x^t is chosen only with knowledge of $x^1, \ldots, x^{t-1}$ and $f^1(x^1), \ldots, f^{t-1}(x^{t-1})$. Performance is measured in terms of regret, which refers to the difference relative to best possible choices taken *ex post*,

i.e.,

$$\text{regret}_H = \sum_{t=1}^{H} f^t(x^t) - \min_{x \in X} \sum_{t=1}^{H} f^t(x), \tag{10.5}$$

where X is some feasible region.

The emphasis in this stream of literature is on algorithms with provable regret bounds. For convex objectives, stochastic search methods (as in Chapter 9) can obtain bounds on regret_H, such as $O(H^{3/4})$, $O(\sqrt{H})$, and $O(\log H)$ depending on properties of f^t and observability of the function (see, respectively, Hazan, Kalai, Kale, and Agarwal [2006], Zinkerich [2003], Flaxman, Kalai, and McMahon [2004]).

d. Optimal stochastic control

Stochastic control models are often similar to stochastic programming models. The differences are mainly due to problem dimension (stochastic programs would generally have higher dimension), emphases on control rules in stochastic control, and more restrictive constraint assumptions in stochastic control. In many cases, the distinction is, however, not at all clear.

As an example, suppose a more general formulation of the financial model in Section 1.2. There, we considered a specific form of the objective function, but we could also use other forms. For example, suppose the objective was generally stated as minimizing some cost $r_t(\mathbf{x}(t), \mathbf{u}(t))$ in each time period t, where $\mathbf{u}(t)$ are the controls $u((i,j),t,s)$ that correspond to actual transactions of exchanging asset i into asset j in period t under scenario s. In this case, problem (1,2.2) becomes:

$$\begin{aligned} \min z = \sum_s p(s) \Big(\sum_{t=1}^{H} & r_t(x(t,s), u(t,s), s) \Big) \\ \text{s. t.} \quad & x(0,s) = b, \\ & x(t,s) + \xi(s)^T u(t,s) = x(t+1,s), t = 0, \ldots, H, \\ & x(s), u(s) \text{ nonanticipative}, \end{aligned} \tag{10.6}$$

where $\xi(s)$ represents returns on investments minus transaction costs. Additional constraints may be incorporated into the objective of (10.6) through penalty terms.

Problem (10.6) is fairly typical of a discrete time control problem governed by a linear system. The general emphasis in control approaches to such problems is for *linear, quadratic, Gaussian* (LQG) models (see, for example, Kushner [1971], Fleming and Rishel [1975], and Dempster [1980]), where we have a linear system as earlier, but where the randomness is Gaussian in each period (for example, ξ is known but the state equation for $x(t+1,s)$ includes a Gaussian term), and r_t is quadratic. In these models, one may also have difficulty observing x so that an additional observation variable $y(t)$ may be present.

LQG models can also include forms of risk aversion as, for example, in Whittle [1990]. In this model, instead of an additively time-separable model as generally used here, the objective to minimize becomes:

$$\frac{2}{\theta}\log \mathrm{E}\left[e^{\theta \sum_{t=1}^{H}(\mathbf{x}^t)^T Q^t \mathbf{x}^t + (\mathbf{u}^t)^T R^t \mathbf{u}^t}\right], \tag{10.7}$$

where $x^{t+1} = A^t x^t + B^t u^t + \varepsilon^t$. A useful property is that this objective avoids some of the issues with time-additive utility functions that do not appear consistent with preferences (as, for example, discussed in Kreps and Porteus [1979], Epstein and Zinn [1989]). A minimizing solution also has a min-max characterization as in robust optimization models and the max-min utility function proposed in Gilboa and Schmeidler [1989] (see Exercise 3 and Hansen and Sargent [1995]).

The LQG problem leads to Kalman filtering solutions (see, for example, Kalman [1969]). Various extensions of this approach are also possible, but the major emphasis remains on developing controls with specific decision rules to link observations directly into estimations of the state and controls. In stochastic programming models, general constraints (such as non-negative state variables) are emphasized. In this case, most simple decision rules forms (such as when u is a linear function of state) fail to obtain satisfactory solutions (see, for example, Gartska and Wets [1974]). For this reason, stochastic programming procedures tend to search for more general solution characteristics.

Stochastic control procedures may, of course, apply but stochastic programming tends to consider more general forms of interperiod relationships and state space constraints. Other types of control formulations, such as robust control, may also be considered specific forms of a stochastic program that are amenable to specific techniques to find control policies with given characteristics.

Continuous time stochastic models (see, e.g., Harrison [1985]) are also possible but generally require more simplified models than those considered in stochastic programming. Again, continuous time formulations are consistent with stochastic programs but have not been the main emphasis of research or the examples in this book. In certain examples again, they may be quite relevant (see, for example, Harrison and Wein [1990] for an excellent application in manufacturing) in defining fundamental solution characteristics, such as the optimality of control limit policies.

In all these control problems, the main emphasis is on characterizing solutions of some form of the dynamic programming Bellman-Hamilton-Jacobi equation or application of Pontryagin's maximum principle. Stochastic programs tend to view all decisions from beginning to end as part of the procedure. The dependence of the current decision on future outcomes and the transient nature of solutions are key elements. Section 3.5 provides some further explanation by describing these characteristics in terms of general optimality conditions.

e. Summary

Stochastic programming is simply another name for the study of optimal decision making under uncertainty. The term *stochastic programming* emphasizes a link to mathematical programming and algorithmic optimization procedures. These considerations dominate work in stochastic programming and distinguish stochastic programming from other fields of study. In this book, we follow this paradigm of concentrating on representation and characterizations of optimal decisions and on developing procedures to follow in determining optimal or approximately optimal decisions. This development begins in the next chapter with basic properties of stochastic program solution sets and optimal values.

Exercises

1. Consider the design problem in Section 1.4. Suppose the design decision does not completely specify x in (1.4.1), but the designer only knows that if a value $\hat{x}$ is specified then $x \in [.99\hat{x}, 1.01\hat{x}]$. Suppose a uniform distribution for x is assumed initially on this interval and that the designer can alter the design once after manufacturing and testing N axles out of a total predicted demand of 1,000 axles. The designer assumes that her posterior distribution on the actual mean relative to $\hat{x}$ would not change if she adjusts the target diameter $\hat{x}$ after observing the first N axle diameters. With these assumptions, formulate a Bayesian model to determine an initial specification $\hat{x}^1$ and N followed by a second specification $\hat{x}^2$ for the remaining $1000 - N$ axles.
2. From the example in Section 1.2, suppose that a goal in each period is to realize a 16% return in each period with penalties $q = 1$ and $r = 4$ as before. Formulate the problem as in (10.2).
3. Consider the risk-sensitive model in (10.7) given initial state x^1, $\theta > 0$, $H = 2$, and $\varepsilon^1 \sim N(\mu, \Sigma)$, the multivariate normal distribution with mean μ and variance-covariance matrix, Σ. Show that solving (10.7) is equivalent to solving the min-max problem:

$$\min_{u^1} \max_{\varepsilon^1} \theta[((u^1)^T R^1 u^1 + x^2(x^1, u^1, \varepsilon^1)^T Q^2 x^2(x^1, u^1, \varepsilon^1)^T) + (\varepsilon^1 - \mu)^T \Sigma^{-1} (\varepsilon^1 - \mu)], \quad (10.8)$$

i.e., u^1 optimal in (10.8) is also optimal in (10.7) and vice versa as long as both problems have finite optimal values. To do this, first show that $\int e^{-Q(x,y)} dy = ke^{-\min_y Q(x,y)}$ for some constant k (independent of x) for any positive definite quadratic function $Q(x, y)$.

2.11 Short Reviews

a. Linear programming

Consider a linear program (LP) of the form

$$\max\{c^T x \mid Ax = b, x \geq 0\} , \tag{11.1}$$

where A is an $m \times n$ matrix, x and c are $n \times 1$ vectors, and b is an $m \times 1$ vector. If needed, any inequality constraint can be transformed into an equality by the addition of *slack variables*:

$$a_{i\cdot} x \leq b_i \quad \text{becomes} \quad a_{i\cdot} x + s_i = b_i ,$$

where s_i is the slack variable of row i and $a_{i\cdot}$ is the i th row of matrix A.

A *solution* to (11.1) is a vector x that satisfies $Ax = b$. A *feasible solution* is a solution x with $x \geq 0$. An *optimal solution* x^* is a feasible solution such that $c^T x^* \geq c^T x$ for all feasible solutions x. A *basis* is a choice of n linearly independent columns of A. Associated with a basis is a submatrix B of the corresponding columns, so that, after a suitable rearrangement, A can be partitioned into $A = [B, N]$. Associated with a basis is a *basic solution*, $x_B = B^{-1} b$, $x_N = 0$, and $z = c_B^T B^{-1} b$, where $[x_B, x_N]$ and $[c_B, c_N]$ are partitions of x and c following the basic and nonbasic columns. We use B^{-1} to denote the inverse of B, which is known to exist because B has linearly independent columns and is square.

In geometric terms, basic solutions correspond to *extreme points* of the polyhedron, $\{x \mid Ax = b, x \geq 0\}$. A basis is feasible (optimal) if its associated basic solution is feasible (optimal). The conditions for feasibility are $B^{-1} b \geq 0$. The conditions for optimality are that in addition to feasibility, the inequalities, $c_N^T - c_B^T B^{-1} N \leq 0$, hold.

Linear programs are routinely solved by widely distributed, easy-to-use LP solvers. Access to such a solver would be useful for some exercises in this book. For a better understanding, some examples and exercises also use manual solutions of linear programs.

Finding an optimal solution is equivalent to finding an optimal *dictionary*, a definition of individual variables in terms of the other variables. In the *simplex algorithm*, starting from a feasible dictionary, the next one is obtained by selecting an *entering variable* (any nonbasic variable whose increase leads to an increase in the objective value), then finding a *leaving variable* (the first to become negative as the entering variable increases), then realizing a *pivot* substituting the entering for the leaving variable in the dictionary. An optimal solution is reached when no entering variable can be found.

A linear program is *unbounded* if an entering variable exists for which no leaving variable can be found. In some cases, a feasible initial dictionary is not available at once. Then, *phase one* of the simplex method consists of finding such an initial dictionary. A number of artificial variables are introduced to make the dictionary

feasible. The phase one procedure minimizes the sum of artificials using the simplex method. If a solution with a sum of artificials equal to zero exists, then the original problem is feasible and *phase two* continues with the true objective function. If the optimal solution of the phase one problem is nonzero, then the original problem is *infeasible*.

As an example, consider the following linear program:

$$\begin{aligned} &\max -x_1 + 3x_2 \\ &\text{s. t. } 2x_1 + x_2 \geq 5\,, \\ &\qquad x_1 + x_2 \leq 3\,, \\ &\qquad x_1, x_2 \geq 0\,. \end{aligned}$$

Adding slack variables s_1 and s_2, the two constraints read

$$\begin{aligned} 2x_1 + x_2 - s_1 &= 5\,, \\ x_1 + x_2 + s_2 &= 3\,. \end{aligned}$$

The natural choice for the initial basis is (s_1, s_2). This basis is infeasible as s_1 would obtain the value -5. An *artificial variable* (a_1) is added to row one to form:

$$2x_1 + x_2 - s_1 + a_1 = 5\,.$$

The phase-one problem consists of minimizing a_1, i.e., finding $-\max -a_1$. Let $z = -a_1$ be the phase one objective, which after substituting for a_1 gives the initial dictionary in phase one:

$$\begin{aligned} z &= -5 + 2x_1 + x_2 - s_1\,, \\ a_1 &= 5 - 2x_1 - x_2 + s_1\,, \\ s_2 &= 3 - x_1 - x_2\,, \end{aligned}$$

corresponding to the initial basis (a_1, s_2). Entering candidates are x_1 and x_2 as they both increase the objective value. Choosing x_1, the leaving variable is a_1 (because it becomes zero for $x_1 = 2.5$ while s_2 becomes zero only for $x_1 = 3$). Substituting x_1 for a_1, the second dictionary becomes:

$$\begin{aligned} z &= -a_1\,, \\ x_1 &= 2.5 - 0.5x_2 + 0.5s_1 - 0.5a_1\,, \\ s_2 &= 0.5 - 0.5x_2 - 0.5s_1 + 0.5a_1\,. \end{aligned}$$

This dictionary is an optimal dictionary for phase one. (No nonbasic variable would possibly increase x.) This means the original problem is feasible. (In fact, the basis (x_1, s_2) is feasible with solution $x_1 = 2.5$, $x_2 = 0.0$.)

We now turn to phase two. We replace the phase one objective with the original objective:

$$z = -x_1 + 3x_2 = -2.5 + 3.5x_2 - 0.5s_1\,.$$

By removing the artificial variable a_1 (as it is not needed anymore), we obtain the following first dictionary in phase two:

$$\begin{aligned} z &= -2.5 + 3.5x_2 - 0.5s_1 \,, \\ x_1 &= 2.5 - 0.5x_2 + 0.5s_1 \,, \\ s_2 &= 0.5 - 0.5x_2 - 0.5s_1 \,. \end{aligned}$$

The next entering variable is x_2 with leaving variable s_2. After substitution, we obtain the final dictionary:

$$\begin{aligned} z &= 1 - 4s_1 - 7s_2 \,, \\ x_1 &= 2 + s_1 + s_2 \,, \\ x_2 &= 1 - s_1 - 2s_2 \,, \end{aligned}$$

which is optimal because no nonbasic variable is a valid entering variable. The optimal solution is $x^* = (2,1)^T$ with $z^* = 1$.

b. Duality for linear programs

The *dual* of the so-called primal problem (11.1) is:

$$\min\{\pi^T b \mid \pi^T A \geq c^T, \pi \text{ unrestricted}\} \,. \tag{11.2}$$

Variables π are called *dual variables*. One such variable is associated with each constraint of the primal. When the primal constraint is an equality, the dual variable is *free* (unrestricted in sign). Dual variables are sometimes called *shadow prices* or *multipliers* (as in nonlinear programming). The dual variable π_i may sometimes be interpreted as the marginal value associated with resource b_i.

If the dual is unbounded, then the primal is infeasible. Similarly, if the primal is unbounded, then the dual is infeasible. Both problems can also be simultaneously infeasible.

If x is primal feasible and π is dual feasible, then $c^T x \leq \pi^T b$. The primal has an optimal solution x^* if and only if the dual has an optimal solution π^*. In that case, $c^T x^* = (\pi^*)^T b$ and the primal and dual solutions satisfy the *complementary slackness* conditions:

$$(a_{i\cdot})x^* = b_i \text{ or } \pi_i^* = 0 \text{ or both, for any } i = 1, \ldots, m \,,$$

$$(\pi^*)^T a_{\cdot j} = c_j \text{ or } x_j^* = 0 \text{ or both, for any } j = 1, \ldots, n \,,$$

where $a_{\cdot j}$ is the j-th column of A and, as before, $a_{i\cdot}$ is the i-th row of A.

An alternative presentation is to say that $s_i^* \pi_i^* = 0$, where s_i is the slack variable of the ith constraint, i.e., either the slack or the dual variable associated with a constraint is zero, and similarly for the second condition. Thus, the optimal solution of the dual can be recovered from the optimal solution for the primal, and vice versa.

The optimality conditions can also be interpreted to say that either there exists some *improving direction*, w, from a current feasible solution, $\hat{x}$, so that $c^T w > 0$, $w_j \geq 0$ for all $j \in N$, $N = \{j \mid \hat{x}_j = 0\}$, and $a_{i\cdot} w = 0$ for all $i \in I$, $I = \{i \mid a_{i\cdot}\hat{x} = b_i\}$ (hence, for $Ax = b$ in the primal system of (11.1), $I = \{1,\ldots,m\}$) or there exists some π such that $\sum_{i\in I} \pi_i a_{ij} \geq c_j$ for all $j \in N$, $\sum_{i\in I} \pi_i a_{ij} = c_j$ for all $j \notin N$, but both cannot occur. This result is equivalent to the *Farkas lemma*, which gives alternative systems with or without solutions.

The *dual simplex method* replicates on the primal solution what the iterations of the simplex method would be on the dual problem: it first finds the leaving variable (one that is strictly negative) then the entering variable (the first one that would become positive in the objective line). The dual simplex is particularly useful when a solution is already available to the original primal problem and some extra constraint or bound is added to the problem. The reader is referred to Chvátal [1980, pp. 152–157] for a detailed presentation.

Other material not covered in this section is meant to be restrictive to a given topic area. The next section discusses more of the mathematical properties of solutions and functions.

c. Nonlinear programming and convex analysis

When objectives and constraints may contain nonlinear functions, the optimization problem becomes a *nonlinear program*. The nonlinear program analogous to (11.1) has the form

$$\min\{f(x) \mid g(x) \leq 0, h(x) = 0\}\ , \tag{11.3}$$

where $x \in \Re^n$, $f : \Re^n \to \Re$, $g : \Re^n \to \Re^m$, and $h : \Re^n \to \Re^l$. We may also assume that the range of f may include ∞ to allow the objective to include constraints directly through an *indicator function*:

$$\delta(x \mid X) = \begin{cases} 0 & \text{if } g(x) \leq 0\ ,\ h(x) = 0\ , \\ +\infty & \text{otherwise,} \end{cases}$$

where X is the set of x satisfying the constraints in (11.3), i.e., the *feasible region*.

In this book, the feasible region is usually a *convex set* so that X contains any *convex combination*,

$$\sum_{i=1}^{s} \lambda^i x^i, \sum_{i=1}^{s} \lambda^i = 1, \lambda^i \geq 0\ ,\ i = 1,\ldots,s\ ,$$

of points, x^i, $i = 1,\ldots,s$, that are in the feasible region. Extreme points of the region are points that cannot be expressed as a convex combination of two distinct points also in the region. The set of all convex combinations of a given set of points is its *convex hull*.

The feasible region is also most generally *closed* so that it contains all limits of infinite sequences of points in the region. The region is also generally *connected*, so that, for any x^1 and x^2 in the region, there exists some path of points in the feasible region connecting x^1 to x^2 by a function, $\eta : [0,1] \rightarrow \Re^n$ that is continuous with $\eta(0) = x^1$ and $\eta(1) = x^2$. For certain results, we may also assume the region is *bounded* so that a *ball* of radius M, $\{x \mid \|x\| \leq M\}$, contains the entire set of feasible points. Otherwise, the region is *unbounded*. Note that a region may be unbounded while the optimal value in (11.1) or (11.3) is still bounded. In this case, the region often contains a *cone*, i.e., a set S such that if $x \in S$, then $\lambda x \in S$ for all $\lambda \geq 0$. When the region is both closed and bounded, then it is *compact*.

The set of equality constraints, $h(x) = 0$, is often *affine*, i.e., they can be expressed as linear combinations of the components of x and some constant. In this case, each constraint, $h_i(x) = 0$, is a *hyperplane*, $a_i x - b_i = 0$, as in the linear program constraints. In this case, $h(x) = 0$, defines an *affine space*, a *translation* of the *parallel subspace*, $Ax = 0$. The affine space *dimension* is the same as its parallel subspace, i.e., the maximum number of linearly independent vectors in the subspace.

With nonlinear constraints and inequalities, the region may not be an affine space, but we often consider the lowest-dimension affine space containing them, i.e., the *affine hull* of the region. The affine hull is useful in optimality conditions because it distinguishes *interior* points that can be the center of a ball entirely within the region from the *relative interior* (ri), which can be the center of a ball whose intersection with the affine hull is entirely within the region. When a point is not in a feasible region, we often take its *projection* into the region using an operator, Π. If the region is X, then the projection of x onto X is $\Pi(x) = \operatorname{argmin}\{\|w - x\| \mid w \in X\}$.

In this book, we generally assume that the objective function f is a *convex function*, i.e., such that

$$f(\lambda x^1 + (1-\lambda)x^2) \leq \lambda f(x^1) + (1-\lambda) f(x^2),$$

$0 \leq \lambda \leq 1$. If f also is never $-\infty$ and is not $+\infty$ everywhere, then f is a *proper convex function*. The region where f is finite is called the *effective domain* of f (dom f). We can also define convex functions in terms of the *epigraph* of f, $\operatorname{epi}(f) = \{(x,\beta) \mid \beta \geq f(x)\}$. In this case, f is convex if and only if its epigraph is convex. If $-f$ is convex, then f is *concave*.

Often, we assume that f has *directional derivatives*, $f'(x;w)$, that are defined as:

$$f'(x;w) = \lim_{\lambda \downarrow 0} \frac{f(x + \lambda w) - f(x)}{\lambda}.$$

When these limits exist and do not vary in all directions, then f is *differentiable*, i.e., there exists a *gradient*, ∇f, such that

$$\nabla f^T w = f'(x;w)$$

for all directions $w \in \Re^n$. We sometimes distinguish this standard form of differentiability from stricter forms as *Gâteaux* or *G-differentiability*. The stricter forms impose more conditions on the directional derivative such as uniform convergence over compact sets (*Hadamard* derivatives).

We also consider *Lipschitz* continuous or *Lipschitzian* functions such that $|f(x) - f(w)| \leq M\|x - w\|$ for any x and w and some $M < \infty$. If this property holds for all x and w in a set X, then f is Lipschitzian *relative to* X. When this property only holds *locally*, i.e., for $\|w - x\| \leq \varepsilon$ for some $\varepsilon > 0$, then f is *locally Lipschitz* at x.

Among differentiable functions, we often use *quadratic* functions that have a *Hessian* matrix of second derivatives, D, and can be written as

$$f(x) = c^T x + \frac{1}{2} x^T D x \, .$$

Many functions are not, however, differentiable. In this case, we express optimality in terms of *subgradients* at a point x, or vectors, η, such that

$$f(w) \geq f(x) + \eta^T (w - x)$$

for all w. In this case, $\{(x, \beta) \mid \beta = f(x) + \eta^T (w - x)\}$ is a *supporting hyperplane* of f at x. The set of subgradients at a point x is the *subdifferential* of f at x, written $\partial f(x)$.

Other useful properties include that f is *piecewise linear*, i.e., such that $f(x)$ is linear over regions defined by linear inequalities. When f is *separable* so that $f(x) = \sum_{i=1}^n f_i(x_i)$, then other advantages are possible in computation.

Given f convex and a convex feasible region in (11.3), we can define conditions that an optimal solution x^* and associated multipliers (π^*, ρ^*) must satisfy. In general, these conditions require some form of *regularity* condition. A common form is that there exists some $\hat{x}$ such that $g(\hat{x}) < 0$ and h is affine. This is generally called the *Slater condition*.

Given a regularity condition of this type, if the constraints in (11.3) define a feasible region, then x^* is optimal if and only if the *Karush-Kuhn-Tucker* conditions hold so that $x^* \in X$ and there exists $\pi^* \geq 0, \rho^*$ such that

$$\nabla f(x^*) + (\pi^*)^T \nabla g(x^*) + (\rho^*)^T \nabla h(x^*) = 0, \nabla g(x^*)^T \pi^* = 0 \, . \tag{11.4}$$

Optimality can also be expressed in terms of the *Lagrangian*:

$$l(x, \pi, \rho) = f(x) + \pi^T g(x) + \rho^T h(x) \, ,$$

so that sequentially minimizing over x and maximizing over π (in both orders) produces the result in (11.4). This occurs through a *Lagrangian dual problem* to (11.3) as

$$\max_{\pi \geq 0, \rho} \inf_x f(x) + \pi^T g(x) + \rho^T h(x) \, , \tag{11.5}$$

which is always a lower bound on the objective in (11.3) (*weak duality*), and, under the regularity conditions, yields equal optimal values in (11.3) and (11.4) (*strong duality*). In many cases, the Lagrangian can also be interpreted with the *conjugate* function of f, defined as

$$f^*(\pi) = \sup_x \{\pi^T x - f(x)\} ,$$

which is also a convex function if f is convex.

Our algorithms often apply to the Lagrangian to obtain *convergence*, i.e., a sequence of solutions, $x^\nu \to x^*$. In some cases, we also approximate the function so that $f^\nu \to f$ in some way. If this convergence is *pointwise*, then $f^\nu(x) \to f(x)$ for each x individually. If the convergence is *uniform* on a set X, then, for any $\varepsilon > 0$, there exists $N(\varepsilon)$ such that for all $\nu \geq N(\varepsilon)$ and all $x \in X$, $|f^\nu(x) - f(x)| < \varepsilon$.

http://www.springer.com/978-1-4614-0236-7

Introduction to Stochastic Programming

Birge, J.R.; Louveaux, F.
2011, XXV, 485 p. 44 illus., Hardcover
ISBN: 978-1-4614-0236-7