Design and analysis of algorithms for stochastic integer programming

LEEN

STOUGIE

Design and analysis of algorithms for stochastic integer programming

(Ontwerp en analyse van algoritmen voor
stochastische geheeltallige programmering)

Proefschrift

ter verkrijging van de graad van doctor
in de economische wetenschappen
aan de Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus
Prof.dr. M.W. van Hof
en volgens besluit van het College van Dekanen.
De openbare verdediging zal plaatsvinden op
donderdag 30 mei 1985 te 16.00 uur

door

Leendert Stougie

geboren te Oud-Beijerland

Contents

Introduction

Many operations management planning and control problems require a series of decisions over time at an increasing level of detail. For example, there are at least two distinct decision making levels in most production operations. At the lowest level, detailed production scheduling decisions determine who will do a particular job on what machine and when. Considerations at this level include minimizing setups and meeting due dates. At a higher level, aggregate planning decisions are made concerning hiring and layoffs, overtime, production levels for product groups, ordering of raw materials, and setting due dates. The time horizon for aggregate decisions may range from several months to one year. At the time aggregate decisions are made, much detailed information is not known with certainty. This may include future product demand, job processing times, machine breakdowns, worker availability, and raw material availability. In addition, other details are deliberately ignored at the aggregate level. For example, the sequence dependent nature of setups is usually ignored, and product groups are used rather than individual stock-keeping units.

Let us consider a *two-level decision situation* in more general terms. At the *aggregate level* one has to decide upon the acquisition of resources. Precise information on what will ultimately be required of them is not yet available. Subsequently, at the *detailed level,* one has to decide on the actual allocation of the resources, when all the relevant information is at hand. The challenge of these *hierarchical planning problems* is to incorporate the initially imperfect detailed level information into the aggregate decision so as to arrive at an overall solution procedure that is optimal or nearly optimal. Specifically, the

costs of acquiring resources at the aggregate level have to be weighed against the benefits of having them available at the detailed level.

The traditional approach to these types of problems is through the design of a *hierarchical planning system*. In such a system, each decision level is treated as a separate mathematical programming model. The various models are linked such that the solution of a higher level model generates part of the input for the model below it.

There are two fundamental reasons for using a hierarchical approach.

Reducing complexity. Breaking a problem into subproblems is a standard method for simplifying the solution process. A tenet of hierarchical planning is that this partitioning can be done so that the interaction effects between subproblems are acceptably weak.

Coping with uncertainty. It is important to realize that the decisions at the various levels in the planning process need to be made at different points in time. For example, aggregate planning decisions are made early enough to implement plans for hiring/layoff, raw materials acquisition, etc.. On the other hand, a decision to assign a particular job to a specific machine can be postponed until the instant before the job begins processing. This is important in light of the fact that much data at the detailed level is uncertain at the time aggregate decisions are made. If detailed and aggregate decisions were combined in a single giant optimization model, as is sometimes proposed, the detailed decisions would be made earlier than necessary and hence would be based on less reliable forecasts of the uncertain data. The hierarchical approach postpones the detailed decisions as long as possible so that they can be based on more timely and hence more accurate data.

A third advantage often attributed to hierarchical systems is that they parallel the hierarchical organization of most firms. While this is certainly an important consideration, we believe that hierarchical planning organizations, as well as hierarchical planning systems, are a response to the nature of the problems being solved, and to the need to reduce complexity and respond to uncertainty cited above.

Past work in hierarchical planning has mainly consisted of building clever systems. The models have so far always been deterministic in nature. A natural question is how the quality of the decisions produced by such systems can be evaluated. We are interested both in comparing different systems and in direct evaluation of a single system. A very favorable and often applied method to compare different systems empirically is *Monte Carlo* simulation. In this approach higher level models are run with forecasts of the lower level data. Lower level models are run with actual data values generated randomly by the Monte Carlo method. We can often evaluate, either by analytic or by empirical methods, the degree of optimality of the solutions to the *submodel* at each level. All of these evaluation methods fail to answer the question of how good a particular hierarchical system performs when compared to an optimal system. To answer this we need a measure of optimality for the entire system, not just for each subproblem. But then, first of all, we need a rigorous formulation of the optimization problem that the hierarchical system is

supposed to solve.

A little thought should make it clear that no *deterministic* mathematical programming model can be appropriate if we wish to capture the uncertainty that exists at lower levels of the overall decision problem accurately. It would be more appropriate to represent this uncertainty by a *stochastic* model. This leads us into the theory of *stochastic programming,* which is the subarea of mathematical programming that considers problems with parameters represented by random variables.

The stochastic program related to a hierarchical planning problem models lower levels accurately, using stochastic parameters for which probability distributions are specified. The objective at each level is to minimize known costs at that level plus the *expected* objective value of an optimal lower level solution. For example, consider the form such a model would take for a *hierarchical job shop scheduling problem.* At the time machines are acquired, only probabilistic information is available on the jobs to be processed. A two-stage stochastic programming model of this problem would select the number or the types of the machines so as to minimize the acquisition costs of the machines plus the expected cost of processing the jobs optimally on the acquired machines.

For those who are familiar with stochastic programming theory, we notice that our concept of a multi-stage stochastic program is broader than what is common in the literature. According to the traditional interpretation each next stage reflects a *recourse* decision to correct infeasibilities due to the decision at the previous stage. Here we do not so much correct infeasibilities but we have to pay extra if our aggregate level decision differs from one that would have been optimal if all detailed level information would have been available before the aggregate decision was made.

Unfortunately, the formulation of a hierarchical planning problem as a multi-stage stochastic programming problem does not bring us any closer to its optimal solution, because of the generally recognized computational difficulty of stochastic programming. The evaluation of a stochastic programming objective function in one point of its domain asks for the computation of the expected optimal solution value of the detailed level problem. In the case the problem parameters have a continuous distribution this amounts to the integration of a function, of which one evaluation requires the solution of a deterministic mathematical programming model.

In this book we focus on hierarchical planning problems, of which the lower decision level is of a *combinatorial* nature. It therefore involves the solution of an *integer* rather than a *linear programming* problem as is common in the existing literature on stochastic programming. This adds another serious computational difficulty.

Whereas for solving deterministic linear programming problems truly efficient methods (like Karmarkar's method) have been developed, no such method has been found so far for integer programming. None of the methods proposed in the sixties turned out to be able to solve any but the smallest problems within a reasonable amount of time. Even today, when linear

programming problems with thousands of variables are solved on a routine basis, integer programming problems with one hundred variables may already present insurmountable problems.

The computational difficulties associated with integer programming appear to be fundamental. For a while, optimists could keep hoping that some totally new approach could provide a breakthrough to a truly efficient integer programming method. *Computational complexity theory,* however, put an end to that illusion in the early seventies, by showing that the computational difficulties encountered in solving integer programming problems are likely to be caused by the inherent complexity of the problem and not by the intellectual limitations of the researchers studying it. This theory associates the notion of an easy or *well-solved* problem with the existence of an algorithm whose running time is bounded by a *polynomial* function of the problem size (defined as the number of bits needed to encode a problem instance). In this sense the general integer programming problem is highly unlikely to be easy: it belongs to a class of notoriously difficult combinatorial optimization problems, the *NP-hard* problems, for which strong evidence exists that any solution method has, in the worst case, a running time that is a superpolynomial function of the problem size.

The importance of the distinction between these two types of running times is revealed when large problem instances are considered. Table 1 (cf. [Garey & Johnson 1977]) illustrates the differences in growth rates among several running time functions. We note the explosive growth rates for the two exponential functions.

Even more revealing is an examination of the effect of improved technology on algorithms having these running time functions. Table 2 shows how the size of the largest problem solvable in one hour would be affected if we had a computer 100 or 1000 times faster than our present one. We observe that with the 2^n running time function a thousand fold increase in speed would only add 10 to the size of the largest problem that can be solved in one hour, whereas with the n^5 algorithm this size almost quadruples.

Thus an improvement in technology will not really help us in solving larger problems with algorithms that have an exponential running time. Our only hope is therefore through understanding the structure of the problems in order to arrive at faster algorithms.

The intractability of both integer programming and stochastic programming justifies some pessimism about the optimal solution of *stochastic integer programming problems.* It would indeed be foolhardy to aim for the design of a solution method that solves any instance of these problems to optimality. As in integer programming one possible attitude towards these problems is to abolish the ideal of optimization and to settle for an approximation of the optimal solution. Thus, *approximation methods* or *heuristics* are looked for. Essentially, the hierarchical planning systems that we mentioned before are nothing but that: *heuristics for stochastic integer programming.* So we have

running time function	Size n					
	10	20	30	40	50	60
n	.00001 second	.00002 second	.00003 second	.00004 second	.00005 second	.00006 second
n^2	.0001 second	.0004 second	.0009 second	.0016 second	.0025 second	.0036 second
n^3	.001 second	.008 second	.027 second	.064 second	.125 second	.216 second
n^5	.1 second	3.2 seconds	24.3 seconds	1.7 minutes	5.2 minutes	13.0 minutes
2^n	.001 second	1.0 second	17.9 minutes	12.7 days	35.7 years	366 centuries
3^n	.059 second	58 minutes	6.5 years	3855 centuries	2×10^8 centuries	1.3×10^{13} centuries

Comparison of several polynomial and exponential running time functions. (Copied with permission from M.R. Garey and D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness. W.H. Freeman and Company. Copyright © 1979, Figure 1.2, p 7.)

Size of Largest Problem Instance
Solvable in 1 Hour

running time function	With present computer	With computer 100 times faster	With computer 1000 times faster
n	N_1	$100N_1$	$1000N_1$
n^2	N_2	$10N_2$	$31.6N_2$
n^3	N_3	$4.64N_3$	$10N_3$
n^5	N_4	$2.5N_4$	$3.98N_4$
2^n	N_5	$N_5 + 6.64$	$N_5 + 9.97$
3^n	N_6	$N_6 + 4.19$	$N_6 + 6.29$

TABLE 2. Effect of improved technology on several polynomial and exponential time algorithms. (Copied with permission from M.R. Garey and D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman and Company. Copyright © 1979. Figure 1.3, p 8.)

returned to where we started. But we have gained something on our way. The formulation of hierarchical planning problems as stochastic integer programming problems provides a proper framework for a *theoretical* analysis of hierarchical planning systems, as opposed to an empirical one. Whereas an empirical analysis involves the evaluation of (necessarily arbitrary) computational experiments, we will now aim for a rigid estimate of the *error* of the heuristic, i.e., the absolute or relative difference between the heuristic solution value and the optimal one. As both must depend on realizations of the random parameters of the stochastic programming model, our performance analyses and the resulting quality statements are necessarily of a probabilistic nature.

The design and analysis of hierarchical planning systems viewed as heuristics for stochastic integer programming problems are the subject of the first three chapters. In Chapter 1 a *general approach* is outlined. Stochastic models for hierarchical planning problems with two decision levels are formulated. We also indicate how to construct hierarchical planning systems for their solution. It is evident that a heuristic for scheduling problems will differ from one for vehicle routing problems. Therefore the hierarchical planning systems are unavoidably problem specific. However, the differences are mainly reflected in the part of the systems concerned with the solution of the detailed level decision problem. There is enough similarity between the various hierarchical planning systems with respect to the ways in which the aggregate level decision is derived and in which the detailed level heuristic is embedded in the system, to allow general design principles to be formulated.

Also the performance analyses of our hierarchical planning systems have enough in common to allow for the application of general devices. We review various performance measures and exhibit some relations between them. For two stochastic integer programming problems of a general nature we design hierarchical planning systems along the above lines and derive sufficient conditions under which they satisfy various quality statements based on the performance measures.

In Chapters 2 and 3 we consider some specific examples. In Chapter 2 we study hierarchical *scheduling* problems, i.e., problems, in which the detailed level involves the scheduling of jobs on machines. In Chapter 3 we study hierarchical *routing* problems, in which the detailed level asks for the routing of vehicles located at a central depot through customers, and hierarchical *location* problems, in which depots or service centers are to be located at the detailed level. For these problems hierarchical planning systems are designed and analyzed through application of the general principles outlined in Chapter 1.

The remaining part of this book is devoted to *optimization methods* rather than approximation methods for stochastic integer programming problems. As mentioned above it is virtually hopeless to aim for methods that solve any stochastic integer programming problem efficiently. As in integer programming, the only way to obtain computational success seems to be through the exploitation of special structure. One might investigate if some stochastic integer programming problems have enough structure to allow for the design of enumerative solution methods whose empirical behavior is satisfactory although they are not efficient in the formal sense.

Only a few results are available in this direction. In Chapter 4 optimization methods are designed for stochastic integer programming problems whose special structure is induced by the (discrete) distribution of the parameters which is concentrated on a small number of points. The relations that exist between the various feasible solutions of these problems can efficiently be exploited by *dynamic programming* recursions. Such methods have a running time that is bounded by a function that is polynomial in the problem size, but exponential in the above mentioned number of points with positive density.

Finally, in Chapter 5, we aim for more general results. For this we need theoretical insight into the properties of stochastic integer programming objective functions. Unfortunately, certain theoretical properties of linear programming that have contributed to the design of successful stochastic linear programming algorithms are typically lacking for integer programming. For example, properties of optimal linear programming solution values viewed as functions of the parameters imply convexity of stochastic linear programming objective functions in the aggregate level decision variables. Stochastic integer programming objective functions, however, are generally non-convex, and if the random parameters have discrete distributions, they are even discontinuous. The main results that we derive in this chapter are that, for a general class of stochastic integer programming models, continuous distributions for the random parameters induce continuous but not necessarily convex objective functions, while discrete distributions lead to discontinuous objective functions. These results should be regarded as some initial theoretical insight in the structure of stochastic integer programming objective functions. They are still far removed from a well implemented stochastic integer programming algorithm. Directions for future research will be discussed.

A framework for the design and probabilistic analysis

of hierarchical planning systems

In the introduction we formulated hierarchical planning problems as multi-stage stochastic integer programming problems. We explained their intractability and proposed hierarchical planning systems as heuristics for their solution. The stochastic programming model provides a proper framework for the analytical evaluation of the performance of such heuristics. In Chapters 2 and 3 we will see that precise statements about the behavior of hierarchical scheduling, routing and location systems can be derived, such as asymptotic optimality in expectation, in probability or with probability 1.

Although the probabilistic analyses of these heuristics are different, the statements that can be derived are similar. Also, the hierarchical planning systems constructed have many features in common. The purpose of this chapter is to outline a general approach to the design and analysis of hierarchical planning systems.

In Section 1.1 we will formulate two basic stochastic programming models for a hierarchical planning problem with two decision levels. In Section 1.2 we will indicate how to construct heuristics for its solution. We will review the various ways to measure the performance of such heuristics in Section 1.3 and exhibit relations between these measures in Section 1.4. We will use these concepts in Section 1.5 in analyzing a general two-level planning problem, of which many of the problems in Chapters 2 and 3 are special cases. In Section 1.6 we analyze a general two-stage decision situation in which there is a possibility to adjust the first stage decision at a certain cost, when a realization of the stochastic parameters has become known.

1.1. Stochastic programming models

Consider the typical two-stage decision situation outlined in the introduction.

At the aggregate level, one has to decide upon the acquisition of resources. The first stage decision will be denoted by X, the set of feasible decisions by $\mathcal{X}$, and the direct cost associated with X by $f(X)$, where $f: \mathcal{X} \rightarrow \mathbb{R}$ is a real function. Probabilistic information about future resource requirements is represented by a n-dimensional vector $\mathbf{w}$. We will indicate random variables by boldface characters. We denote the set of all possible realizations of $\mathbf{w}$ by $\mathcal{W}$.

The input to the detailed level consists of the first stage decision X and a realization w of the random vector $\mathbf{w}$. The objective at the second stage is to decide upon a certain allocation of the resources acquired so as to minimize a cost $y(X,w)$, where $y: \mathcal{X} \times \mathbb{R}^n \rightarrow \mathbb{R}$ is a real function. The optimal value of $y(X,\mathbf{w})$ will be denoted by $y^*(X,\mathbf{w})$. We notice that, for a given X, $y^*(X,\mathbf{w})$ is a (complicated) function of the random vector $\mathbf{w}$ and hence a random variable in itself. The total cost of the acquisition decision X and the optimal allocation decision will be denoted by $z^*(X,\mathbf{w}) = f(X) + y^*(X,\mathbf{w})$.

The stochastic programming formulation that naturally captures the uncertainty existing at the lower level is the *two-stage decision model*. Each of the two decision levels of the hierarchical planning problem corresponds to a stage in this model. The objective at the first stage is to determine a decision $X^* \in \mathcal{X}$ such that the *expected* total cost $Ez^*(X,\mathbf{w}) = f(X) + Ey^*(X,\mathbf{w})$ is minimized:

$$Ez^*(X^*,\mathbf{w}) = \min_{X \in \mathcal{X}} \{Ez^*(X,\mathbf{w})\}.$$

In stochastic programming terminology, the first stage decision is made 'here and now', given *imperfect* information about the second stage, and it is therefore independent of a particular realization w of $\mathbf{w}$.

We also formulate the *distribution model*. Contrary to the two-stage decision model, the objective here is to determine a *function* $X^\circ: \mathbb{R}^n \rightarrow \mathcal{X}$ such that for each realization w of $\mathbf{w}$ the *actual* cost is minimized:

$$z^*(X^\circ(w),w) = \min_{X \in \mathcal{X}} \{z^*(X,w)\}, \ \forall w \in \mathcal{W}.$$

Thus, before the aggregate decision is taken, we 'wait and see' until *perfect* information about the second stage is available. Solving this problem can be regarded as characterizing the minimum total cost achievable as a function of $\mathbf{w}$. Although this model does not reflect the actual decision process in a hierarchical planning problem, it allows us to obtain more insight into the performance qualities of methods for its solution.

1.2. Two-stage stochastic programming heuristics

As has been argued in the introduction, there is little hope to develop efficient optimization algorithms for the above stochastic programs. As to the two-stage decision model, the determination of $y^*(X,w)$ is often an NP-hard problem, so that a heuristic must be used at the second stage. Even if $y^*(X,w)$ can easily be determined, it seems impossible to obtain a tractable representation of $Ey^*(X,\mathbf{w})$, and the use of a heuristic at the first stage is generally unavoidable. The distribution model is at least as hard to solve to optimality. We will outline a two-stage heuristic approach; the heuristics at the first and second stage will be denoted by H_1 and H_2, respectively.

At the first stage, we replace $Ey^*(X,\mathbf{w})$ by an approximation $y^{H_1}(X)$ and determine an $X^{H_1} \in \mathcal{X}$ such that the approximate total cost $z^{H_1}(X) = f(X) + y^{H_1}(X)$ is minimized:

$$z^{H_1}(X^{H_1}) = \min_{X \in \mathcal{X}}\{z^{H_1}(X)\}.$$

In some cases, even this approximate first stage problem is NP-hard and another heuristic device is needed to solve it (see Sections 2.1.2 and 3.1.2).

At the second stage, we allocate the resources acquired, achieving an approximate cost $y^{H_2}(X^{H_1},w)$. In some cases, the second stage problem does not require any approximation and H_2 denotes a polynomial-time optimization algorithm. The total cost of the entire heuristic for a realization w of $\mathbf{w}$ will be denoted by $z^{H_2}(X^{H_1},w) = f(X^{H_1}) + y^{H_2}(X^{H_1},w)$.

The success of this heuristic approach evidently depends on the quality of $y^{H_1}(X)$ as an approximation of $Ey^*(X,\mathbf{w})$ and of $y^{H_2}(X,w)$ as an approximation of $y^*(X,w)$. In this context, use can be made of the existing literature on probabilistic analyses of combinatorial optimization problems and heuristics to solve them. Specifically, we can use results from *probabilistic value analysis* of combinatorial optimization problems. Although these problems may be difficult to solve, their optimal value often allows for a simple probabilistic description in terms of the problem parameters. Results in this direction were achieved for routing problems in [Beardwood et al. 1959] and [Steele 1981] and for location problems in [Hochbaum & Steele 1981] and [Zemel 1984].

We will also use results from *probabilistic error analysis* of heuristics for combinatorial optimization problems as were outlined in [Karp 1977] for routing problems and in [Fisher & Hochbaum 1980], [Hochbaum & Steele 1981], [Papadimitriou 1981] and [Zemel 1984] for location problems.

1.3. Performance measures

Before defining a number of ways to measure the performance of stochastic programming heuristics, we recall some concepts of stochastic convergence. A sequence of random variables $\mathbf{x}_1, \mathbf{x}_2, \ldots$ is said to converge to a random variable $\mathbf{x}$

(a) *in expectation* if $\lim_{n \to \infty} E \mid x_n - x \mid = 0$
    [notation: $E \mid x_n - x \mid \to 0$];
(b) *in probability* if $\lim_{n \to \infty} Pr \{ \mid x_n - x \mid \leq \epsilon \} = 1$ for every $\epsilon > 0$
    [notation: $x_n \to x$ (ip)];
(c) *with probability* 1 or *almost surely* if $Pr \{ \lim_{n \to \infty} x_n = x \} = 1$
    [notation: $x_n \to x$ (wp1)].

Some well-known relations between these types of convergence are given in Section 1.4.

The quality of a solution provided by a two-stage heuristic (H_1, H_2) can be measured by comparing it with optimal solutions to the two-stage decision model and to the distribution model. The quality statements are of an asymptotic nature, i.e., they are concerned with problems of growing size reflected in the dimension of the random vector **w**. In the notation the index indicating the problem size is suppressed if it is clear from the context.

In the context of the first model, one is primarily interested in the asymptotic behavior of the ratio of the expected costs

$$\frac{Ez^{H_2}(X^{H_1}, \mathbf{w})}{Ez^*(X^*, \mathbf{w})}.$$

We notice that the ratios defined in this section are well-defined since for each X and for each $w \in \mathcal{W}$ $z^*(X, w) > 0$. If the above ratio tends to 1 as the problem size tends to infinity, then we say that the approximation algorithm (H_1, H_2) is *asymptotically expectation-optimal*. If the heuristic depends on a given number $\epsilon > 0$ and has the property that, for each ϵ, the ratio tends to a number less than $1 + \epsilon$, then $(H_1(\epsilon), H_2(\epsilon))$ is said to be an *asymptotically expectation-optimal approximation scheme*.

Other obvious ideas are to investigate the asymptotic behavior of the ratio of the actual costs

$$\frac{z^{H_2}(X^{H_1}, \mathbf{w})}{z^*(X^*, \mathbf{w})} \quad \text{and} \quad \frac{z^{H_2}(X^{H_1}, \mathbf{w})}{z^*(X^\circ(\mathbf{w}), \mathbf{w})}.$$

If the first of both ratios tends to 1 (or, for each $\epsilon > 0$, to a number less than $1 + \epsilon$) in expectation, in probability or with probability 1, then we say that the approximation algorithm (or scheme) is *asymptotically optimal in expectation, in probability* or *with probability* 1. If the second ratio satisfies analogous properties, then the heuristic is said to be *asymptotically clairvoyant* rather than asymptotically optimal: in addition to the inaccuracy due to approximating the two-stage decision model, the relative loss caused by imperfect information also disappears in the limit.

Still other measures are based on a comparison of the aggregate decisions X^{H_1}, X^* and $X^\circ(\mathbf{w})$. In case $\mathcal{X}$ is a set of numbers, one can directly investigate the limiting behavior of the ratios

$$\frac{X^{H_1}}{X^*} \quad \text{and} \quad \frac{X^{H_1}}{X^\circ(\mathbf{w})}$$

(see e.g. Sections 2.1.1, 2.3, 3.1.1). The first of these ratios is a deterministic

variable, but the second one is random and its convergence analysis results in probabilistic statements. Sometimes it may even be possible to obtain good bounds on the differences $X^{H_1} - X^*$ and $X^{H_1} - X^\circ(\mathbf{w})$. In case $\mathcal{X}$ is a family of subsets, one possibility is to convert each set $X \in \mathcal{X}$ into a number $W(X)$ by taking a weighted sum over its elements and to consider the ratios of or the differences between $W(X^{H_1})$, $W(X^*)$ and $W(X^\circ(\mathbf{w}))$ (cf. Sections 2.1.2, 3.1.2).

1.4. Relations between performance measures

Lemmas 1.1, 1.2 and 1.3 give fundamental relations between the three types of convergence of a sequence of random variables $\mathbf{x}_1, \mathbf{x}_2, \dots$ to a random variable $\mathbf{x}$. We refer to [Serfling 1980] for proofs and for examples which show that the inverse implications do not hold in general.

LEMMA 1.1. $\mathbf{x}_n \to \mathbf{x}$ (wp1) $\Rightarrow \mathbf{x}_n \to \mathbf{x}$ (ip). $\square$

LEMMA 1.2. $E \, | \, \mathbf{x}_n - \mathbf{x} \, | \to 0 \Rightarrow \mathbf{x}_n \to \mathbf{x}$ (ip). $\square$

Preliminary to Lemma 1.3 we give the following definition. Let $1_{\{x \in A\}}$ be the indicator function of x which has value 1 if $x \in A$ and value 0 otherwise.

DEFINITION. *A sequence of random variables* $\mathbf{x}_1, \mathbf{x}_2, \dots$ *is uniformly integrable if*

$$\lim_{c \to \infty} \sup_n E(\, | \, \mathbf{x}_n \, | \, 1_{|\mathbf{x}_n| > c}) = 0.$$

LEMMA 1.3. *Suppose the sequence* $\mathbf{x}_1, \mathbf{x}_2, \dots$ *is uniformly integrable. Then* $\mathbf{x}_n \to \mathbf{x}$ (ip) $\Rightarrow E \, | \, \mathbf{x}_n \, | \to E \, | \, \mathbf{x} \, |$ *and* $E\mathbf{x}_n \to E\mathbf{x}$. $\square$

We will now investigate relations between the performance measures introduced in the previous section. To simplify notation, we will write

$$\mathbf{z}^H \text{ for } z^{H_2}(X^{H_1}, \mathbf{w}), \quad \mathbf{z}^* \text{ for } z^*(X^*, \mathbf{w}), \quad \mathbf{z}^\circ \text{ for } z^*(X^\circ(\mathbf{w}), \mathbf{w}).$$

To simplify the analysis, we introduce the following assumptions that are usually satisfied in our applications.

ASSUMPTION 1.1. $\mathbf{z}^H / \mathbf{z}^\circ$ *is uniformly integrable.*

ASSUMPTION 1.2. $\mathbf{z}^H / E\mathbf{z}^\circ$ *is uniformly integrable.*

ASSUMPTION 1.3. *There exists a constant* $c_1 > 0$ *such that for* n *sufficiently large* $\mathbf{z}^* / E\mathbf{z}^* < c_1$ (wp1).

ASSUMPTION 1.4. *There exists a constant* $c_2 > 0$ *such that for* n *sufficiently large* $\mathbf{z}^* / E\mathbf{z}^* > c_2$ (wp1).

ASSUMPTION 1.5. *There exists a constant* $c_3 > 0$ *such that for* n *sufficiently large* $\mathbf{z}^\circ / E\mathbf{z}^\circ > c_3$ (wp1).
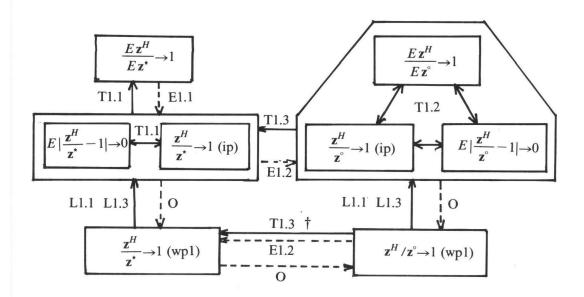
14

$$\frac{E\,z^H}{E\,z^\star}\to 1$$

T1.1   E1.1

$$E\,|\frac{z^H}{z^\star}-1|\to 0 \qquad \text{T1.1} \qquad \frac{z^H}{z^\star}\to 1 \ (\text{ip})$$

T1.3

$$\frac{E\,z^H}{E\,z^\circ}\to 1$$

T1.2

$$\frac{z^H}{z^\circ}\to 1\ (\text{ip}) \qquad E\,|\frac{z^H}{z^\circ}-1|\to 0$$

E1.2

L1.1 L1.3   O

L1.1 L1.3   O

T1.3 †

$$\frac{z^H}{z^\star}\to 1\ (\text{wp1})$$

E1.2

$$z^H/z^\circ\to 1\ (\text{wp1})$$

O

FIGURE 1.1. Relations between performance measures

$\to$: valid implication; $\dashrightarrow$: invalid implication; O: Obvious; E: Example; L: Lemma; T: Theorem; †: if $z^H/z^\star$ has a finite limit (wp1)

In addition to these assumptions, we will use the basic properties of our models that $z^H\geqslant z^\circ$, $z^\star\geqslant z^\circ$, $z^\circ>0$ and $E\,z^H\geqslant E\,z^\star$; but it need not be true that $z^H\geqslant z^\star$ for every realization $w\in\mathcal{W}$. Note that under Assumption 1.1 also $z^H/z^\star$ is uniformly integrable, and that under Assumption 1.2 also $z^H/E\,z^\star$ and $z^\circ/E\,z^\circ$ are uniformly integrable. In each of the applications that we are considering in the following chapters we will investigate to what extent our assumptions are realistic.

Figure 1.1 shows which relations hold under these assumptions, and which do not. We will first illustrate some of the invalid implications by means of two examples, and next prove the valid implications in three theorems. The examples are of a general non-asymptotic nature, but the variables satisfy the above properties, which are inherent to our models.

EXAMPLE 1.1. $E\mathbf{z}^H/E\mathbf{z}^* \to 1$ but $E|\mathbf{z}^H/\mathbf{z}^* - 1| \not\to 0$ and $\mathbf{z}^H/\mathbf{z}^* \not\to 1$ (ip).

Let $Pr\{\mathbf{z}^H = 1, \mathbf{z}^* = 2\} = \frac{1}{2}$ and $Pr\{\mathbf{z}^H = 2, \mathbf{z}^* = 1\} = \frac{1}{2}$. We have $E\mathbf{z}^H/E\mathbf{z}^* = 1$ but $E|\mathbf{z}^H/\mathbf{z}^* - 1| = \frac{3}{4}$ and $\mathbf{z}^H/\mathbf{z}^* \in \{\frac{1}{2}, 2\}$. $\square$

EXAMPLE 1.2. $\mathbf{z}^H/\mathbf{z}^\circ \to 1$ (wp1) but $\mathbf{z}^H/\mathbf{z}^* \not\to 1$ (wp1).

Let $\mathbf{x}$ be uniformly distributed on the unit interval $[0,1]$ and let n denote problem size. For each $n \in \mathbb{N}$, define

$$\mathbf{z}_n^H = \begin{cases} 2 & (x \in I_n^H) \\ 1 & (x \notin I_n^H) \end{cases},$$

$$\mathbf{z}_n^* = \begin{cases} 2 & (x \in I_n^*) \\ 1 & (x \notin I_n^*) \end{cases},$$

$$\mathbf{z}_n^\circ = 1,$$

where the intervals I_n^H and I_n^* are defined by

$$l(n) = 2^{[\log_2 n]}, \quad I_n^H = \left[0, \frac{1}{l(n)}\right], \quad I_n^* = \left[\frac{n - l(n)}{l(n)}, \frac{n - l(n) + 1}{l(n)}\right]$$

$$n = 4,5,6,7;$$
$$l(n) = 4$$

$$I_n^H$$

$$0 \quad I_4^* \quad \frac{1}{4} \quad I_5^* \quad \frac{1}{2} \quad I_6^* \quad \frac{3}{4} \quad I_7^*$$

FIGURE 1.2. Illustration of the intervals in Example 1.2

(cf. Figure 1.2). We observe that $\lim_{n\to\infty} \mathbf{z}_n^H = 1$ (wp1) so that $\lim_{n\to\infty} \mathbf{z}_n^H/\mathbf{z}_n^\circ = 1$ (wp1) as well; however, with probability 1 $\lim_{n\to\infty} \mathbf{z}_n^*$ does not exist and neither does $\lim_{n\to\infty} \mathbf{z}_n^H/\mathbf{z}_n^*$. In probabilistic terms, we therefore have that $\mathbf{z}_n^H/\mathbf{z}_n^\circ \to 1$ with probability 1 but $\mathbf{z}_n^H/\mathbf{z}_n^* \to 1$ only in probability.

This example is due to H.C.P. Berbee. It will be shown in Theorem 1.3 (ii) that, if $\mathbf{z}_n^H/\mathbf{z}_n^*$ has a finite limit (wp1), then the implication is valid. $\square$

Theorems 1.1 and 1.2 collect the implications between the various convergence properties in the context of the two-stage decision model and the distribution model, respectively.

THEOREM 1.1.

(i) $E|z^H/z^\star - 1| \to 0 \Rightarrow z^H/z^\star \to 1$ (ip);

(ii) $z^H/z^\star \to 1$ (ip) $\Rightarrow E|z^H/z^\star - 1| \to 0$ *under Assumption* 1.1;

(iii) $E|z^H/z^\star - 1| \to 0 \Rightarrow Ez^H/Ez^\star \to 1$ *under Assumption* 1.3.

PROOF. (i) This is immediate from Lemma 1.2.

(ii) It is obvious that under Assumption 1.1, $z^H/z^\star - 1$ is also uniformly integrable. Hence, (ii) follows from Lemma 1.3.

(iii) We bound Ez^H from above by

$$Ez^H = \int z^H dF(w) \leqslant \int (|\frac{z^H}{z^\star} - 1| + 1) z^\star dF(w),$$

so that, under Assumption 1.3, for n sufficiently large

$$1 \leqslant \frac{Ez^H}{Ez^\star} \leqslant 1 + \int |\frac{z^H}{z^\star} - 1| \frac{z^\star}{Ez^\star} dF(w)$$

$$\leqslant 1 + c_1 E|z^H/z^\star - 1|.$$

Since $E|z^H/z^\star - 1| \to 0$, we have $Ez^H/Ez^\star \to 1$. $\square$

THEOREM 1.2.

(i) $E|z^H/z^\circ - 1| \to 0 \Rightarrow z^H/z^\circ \to 1$ (ip);

(ii) $Ez^H/Ez^\circ \to 1 \Rightarrow z^H/z^\circ \to 1$ (ip); *under Assumption* 1.5;

(iii) $z^H/z^\circ \to 1$ (ip) $\Rightarrow E|z^H/z^\circ - 1| \to 0$ *under Assumption* 1.1;

(iv) $z^H/z^\circ \to 1$ (ip) $\Rightarrow Ez^H/Ez^\circ \to 1$ *under Assumptions* 1.1 *and* 1.2.

PROOF. (i) This is immediate from Lemma 1.2.

(ii) For every $\epsilon > 0$ we can bound Ez^H/Ez° from below by

$$\frac{Ez^H}{Ez^\circ} = \int \frac{z^H}{z^\circ} \frac{z^\circ}{Ez^\circ} dF(w) \geqslant 1 + \epsilon E(\frac{z^\circ}{Ez^\circ} 1_{z^{H^\circ} > z^\circ(1+\epsilon)}).$$

So that, under Assumption 1.5,

$$\frac{Ez^H}{Ez^\circ} \geqslant 1 + \epsilon c_3 Pr\{z^H > z^\circ(1+\epsilon)\}.$$

It follows from $Ez^H/Ez^\circ \to 1$ that $Pr\{z^H > z^\circ(1+\epsilon)\} \to 0$ for every $\epsilon > 0$, i.e., $z^H/z^\circ \to 1$ (ip).

(iii) Under Assumption 1.1, $z^H/z^\circ - 1$ is uniformly integrable and we can apply Lemma 1.3 to obtain (iii).

(iv) We can bound Ez^H/Ez° by

$$1 \leqslant \frac{Ez^H}{Ez^\circ} = E(z^H/Ez^\circ)$$

$$= E(\frac{z^H}{z^\circ} \frac{z^\circ}{Ez^\circ} 1_{z^H \leqslant (1+\epsilon)z^\circ}) + E(\frac{z^H}{Ez^\circ} 1_{z^H > (1+\epsilon)z^\circ})$$

$$\leq (1+\epsilon) + E\left(\frac{z^H}{E z^\circ} 1_{z^H > (1+\epsilon)z^\circ}\right).$$

Under Assumption 1.2, we have that $z^H / E z^\circ$ is uniformly integrable. Therefore for every $\epsilon > 0$ there exists a constant δ such that uniformly $E(z^H / E z^\circ 1_{z^H / E z^\circ > \delta}) < \epsilon$, and hence

$$\frac{E z^H}{E z^\circ} \leq (1+\epsilon) + E(z^H / E z^\circ 1_{z^H / E z^\circ \leq \delta} 1_{z^H > (1+\epsilon)z^\circ})$$

$$+ E(z^H / E z^\circ 1_{z^H / E z^\circ > \delta} 1_{z^H > (1+\epsilon)z^\circ})$$

$$\leq 1 + \epsilon + \delta Pr\{z^H > (1+\epsilon)z^\circ\} + E(z^H / E z^\circ 1_{z^H / E z^\circ > \delta})$$

$$\leq 1 + 2\epsilon + \delta Pr\{z^H > (1+\epsilon)z^\circ\}.$$

Since $z^H / z^\circ \to 1$ (ip), we have $Pr\{z^H > (1+\epsilon)z^\circ\} \to 0$ for every $\epsilon > 0$. It follows that $E z^H / E z^\circ \to 1$. $\square$

Theorem 1.3 states the relations between the two-stage decision model and the distribution model.

THEOREM 1.3.
(i) $z^H / z^\circ \to 1$ (ip) $\Rightarrow z^H / z^* \to 1$ (ip), *under Assumptions* 1.2 *and* 1.4;
(ii) *Suppose* z^H / z^* *converges to a constant* L (wp1). *Then* $z^H / z^\circ \to 1$ (wp1) $\Rightarrow z^H / z^* \to 1$ (wp1), *under Assumptions* 1.2 *and* 1.4.

PROOF. (i) For every $\epsilon > 0$ we define

$$\mathcal{W}_0(\epsilon) = \{w : \frac{z^H}{z^\circ} > 1 + \epsilon^2\},$$

$$\mathcal{W}_1(\epsilon) = \{w : \frac{z^H}{z^*} > 1 + \epsilon^2\},$$

$$\mathcal{W}_2(\epsilon) = \{w : 1 - \epsilon \leq \frac{z^H}{z^*} \leq 1 + \epsilon^2\},$$

$$\mathcal{W}_3(\epsilon) = \{w : \frac{z^H}{z^*} < 1 - \epsilon\}.$$

$E z^H$ can be bounded from above by

$$E z^H = \int z^H dF(w)$$

$$\leq \int_{\mathcal{W}_1(\epsilon)} z^H dF(w) + (1+\epsilon^2) \int_{\mathcal{W}_2(\epsilon)} z^* dF(w) + (1-\epsilon) \int_{\mathcal{W}_3(\epsilon)} z^* dF(w),$$

so that

$$1 \leq \frac{E z^H}{E z^*} \leq \int_{\mathcal{W}_1(\epsilon)} \frac{z^H}{E z^*} dF(w) + 1 + \epsilon^2 - \epsilon \int_{\mathcal{W}_3(\epsilon)} \frac{z^*}{E z^*} dF(w).$$

18

Under Assumption 1.2, for every $\epsilon > 0$ there exists a $\delta > 0$ such that uniformly $E(\mathbf{z}^H / E\mathbf{z}^\star 1_{\mathbf{z}^H / E\mathbf{z}^\star > \delta}) < \epsilon^2$. Therefore under Assumptions 1.2 and 1.4 we can bound $E\mathbf{z}^H / E\mathbf{z}^\star$ by

$$1 \leqslant E\mathbf{z}^H / E\mathbf{z}^\star \leqslant \epsilon^2 + 1 + \epsilon^2 + \delta Pr\{\mathbf{w} \in \mathcal{W}_1(\epsilon)\} - \epsilon c_2 Pr\{\mathbf{w} \in \mathcal{W}_3(\epsilon)\}$$

that is,

$$Pr\{\mathbf{w} \in \mathcal{W}_3(\epsilon)\} \leqslant \frac{\delta}{\epsilon c_2} Pr\{\mathbf{w} \in \mathcal{W}_1(\epsilon)\} + \frac{2}{c_2}\epsilon.$$

Since $\mathbf{z}^H / \mathbf{z}^\star \to 1$ (ip), we have $Pr\{\mathbf{w} \in \mathcal{W}_0(\epsilon)\} \to 0$ for every $\epsilon > 0$ and, since $\mathcal{W}_0(\epsilon) \supseteq \mathcal{W}_1(\epsilon)$, $Pr\{\mathbf{w} \in \mathcal{W}_1(\epsilon)\} \to 0$ for every $\epsilon > 0$ as well. This result together with the above upper bound implies that for every $\epsilon > 0$, asymptotically

$$Pr\{\mathbf{w} \in \mathcal{W}_3(\epsilon)\} \leqslant \frac{2}{c_2}\epsilon.$$

We observe that $\lim_{\epsilon \to 0} Pr\{w \in \mathcal{W}_3(\epsilon)\} = 0$. Since $Pr\{w \in \mathcal{W}_3(\epsilon)\}$ is a decreasing function of ϵ this implies that $Pr\{\mathbf{w} \in \mathcal{W}_3(\epsilon)\} \to 0$ for every $\epsilon > 0$. It follows that $Pr\{w \in \mathcal{W}_1(\epsilon) \cap \mathcal{W}_3(\epsilon)\} \to 0$ for every $\epsilon > 0$, i.e., $\mathbf{z}^H / \mathbf{z}^\star \to 1$ (ip).

(ii) Since $\mathbf{z}^H / \mathbf{z}^\circ \to 1$ (wp1) and $\mathbf{z}^\circ \geqslant \mathbf{z}^\star$, we know that $L \leqslant 1$. By Theorem 1.1(ii), the uniform integrability of $\mathbf{z}^H / \mathbf{z}^\star$ and the assumption that $\lim_{n \to \infty} (\mathbf{z}^H / \mathbf{z}^\star)$ exists, we have that

$$1 = \lim_{n \to \infty} E(\mathbf{z}^H / \mathbf{z}^\star) = \int \lim_{n \to \infty} (\mathbf{z}^H / \mathbf{z}^\star) dF(w) = L. \quad \square$$

1.5. A general two-level planning problem

We will first consider a specific, simply structured type of the distribution model formulated in Section 1.1. We will next show how results derived for this type of model also hold for a more general model by simple extensions of the arguments.

The first model has $\mathcal{X} = \mathbb{N}$ and $f(X) = cX$ at the aggregate level for a given constant $c > 0$, and the objective is to determine a function $X^\circ : \mathbb{R}^n \to \mathbb{N}$ such that for each realization w of $\mathbf{w}$

$$z^\star(X^\circ(w), w) = \min_{X \in \mathbb{N}} \{cX + y^\star(X, w)\}.$$

Models of this type occur when one has to decide upon the acquisition of a number of identical resources each at a fixed cost c. Such models are studied in Chapters 2 (Sections 2.1.1 and 2.3) and 3 (Sections 3.1.1, 3.2, 3.3). They share some features that allow us to treat them in a general way. This general treatment concerns the design of the first stage heuristic as well as the analysis of the quality of the first stage decision and the entire hierarchical planning system.

First of all, there typically is a lower bound on $y^\star(X, w)$ that can be written as the product of two factors, one depending only on X and the other only on w. More specifically, there exist a constant $\gamma > 0$ and a function $g : \mathbb{R}^n \to \mathbb{R}$ such

that asymptotically for each $X \in \mathbb{N}$

$$\frac{g(\mathbf{w})}{X^\gamma} \leq y^\star(X,\mathbf{w}) \quad (\text{wp1}).$$

Secondly, there often is an approximation v of $g(\mathbf{w})$ depending on the problem size and the probability distribution of $\mathbf{w}$ that is asymptotically accurate with probability 1:

$$\frac{g(\mathbf{w})}{v} \to 1 \quad (\text{wp1}).$$

Such value estimates are available for various combinatorial optimization problems, as has been mentioned already in Section 1.2.

These characteristics lead to a simple heuristic H_1 for the first stage problem. Defining $y^{H_1}(X) = v/X^\gamma$, we have that asymptotically

$$z^{H_1}(X) = cX + \frac{v}{X^\gamma} \leq cX + y^\star(X,\mathbf{w}) = z^\star(X,\mathbf{w}) \; (\text{wp1}). \tag{1.1}$$

Observing that z^{H_1} is an unimodal function, achieving its minimum at

$$\hat{X} = (\frac{\gamma v}{c})^{\frac{1}{\gamma+1}},$$

we conclude that X^{H_1} is determined by minimizing $z^{H_1}(X)$ subject to $X \in \{\lfloor \hat{X} \rfloor, \lceil \hat{X} \rceil\} \cap \mathbb{N}$. ($\lfloor \hat{X} \rfloor$ and $\lceil \hat{X} \rceil$ denote the integer rounddown and roundup of $\hat{X}$ respectively.)

The third common feature is the existence of a second stage heuristic H_2 that produces an upper bound on $y^\star(X^\circ(\mathbf{w}),\mathbf{w})$ which is asymptotically equal to the above probabilistic lower bound with probability 1:

$$\frac{y^{H_2}(X^{H_1},\mathbf{w})}{y^{H_1}(X^{H_1})} \to 1 \quad (\text{wp1}).$$

No general recipe for the design of such a heuristic can be given, since the model considered here allows for a wide variety of problem types at the detailed level. In this situation, it can be proved that the heuristics H_1 and (H_1,H_2) are both asymptotically clairvoyant with probability 1.

THEOREM 1.4. *If $X \in \mathbb{N}$, $f(X) = cX$ ($c > 0$) and H_1 and H_2 are such that*
(A) $y^{H_1}(X) = v/X^\gamma \leq y^\star(X,\mathbf{w})$ *(wp1) asymptotically and*
(B) $y^{H_2}(X^{H_1},\mathbf{w})/y^{H_1}(X^{H_1}) \to 1$ *(wp1),*
then
(i) $z^{H_2}(X^{H_1},\mathbf{w})/z^\star(X^\circ(\mathbf{w}),\mathbf{w}) \to 1$ *(wp1);*
(ii) $X^{H_1}/X^\circ(\mathbf{w}) \to 1$ *(wp1).*

PROOF. (i) We can bound $z^\star(X,\mathbf{w})$ from below (asymptotically with probability 1) and from above (deterministically) by

$$cX + y^{H_1}(X) = z^{H_1}(X) \leq z^\star(X,\mathbf{w}) \leq z^{H_2}(X,\mathbf{w}) = cX + y^{H_2}(X,\mathbf{w}) \; (\text{wp1}),$$

so that

$$cX^{H_1} + y^{H_1}(X^{H_1}) \leqslant z^*(X^\circ(\mathbf{w}),\mathbf{w}) \leqslant cX^{H_1} + y^{H_2}(X^{H_1},\mathbf{w}) \text{ (wp1)}.$$

Condition B then yields the desired result.

(ii) Let n denote problem size. For each $\epsilon > 0$ we define

$$\mathcal{W}(\epsilon) = \{w: \limsup_{n\to\infty} \frac{X^{H_1}}{X^\circ(w)} < \frac{1}{1+\epsilon}\}.$$

The unimodality of z^{H_1} implies that for $\mathbf{w} \in \mathcal{W}(\epsilon)$ asymptotically

$$z^{H_1}((1+\epsilon)X^{H_1}) \leqslant z^{H_1}(X^\circ(\mathbf{w})) \leqslant z^*(X^\circ(\mathbf{w}),\mathbf{w}) \leqslant z^{H_2}(X^{H_1},\mathbf{w}) \text{ (wp1)}.$$

A tedious but straightforward calculation shows that for each n

$$\frac{z^{H_1}((1+\epsilon)X^{H_1})}{z^{H_1}(X^{H_1})} = \frac{(1+\epsilon)\gamma + (1+\epsilon)^{-\gamma}}{\gamma+1} > 1.$$

Hence, we have for $\mathbf{w} \in \mathcal{W}(\epsilon)$ that

$$\liminf_{n\to\infty} \frac{z^{H_2}(X^{H_1},\mathbf{w})}{z^{H_1}(X^{H_1})} > 1 \text{ (wp1)}.$$

On the other hand, we know that this limit is equal to 1 (wp1), so that $Pr\{\mathbf{w} \in \mathcal{W}(\epsilon)\} = 0$ for every $\epsilon > 0$. Similarly, $Pr\{\liminf_{n\to\infty} X^{H_1}/X^\circ(\mathbf{w}) > 1/(1-\epsilon)\} = 0$. It follows that $X^{H_1}/X^\circ(\mathbf{w}) \to 1$ (wp1). $\square$

Obviously, many of the nice features of the above model no longer hold if the resources that are to be acquired are not identical, so that X is a *set* rather than a number and $f(X)$ is a *set function* such as in Sections 2.1.2 and 3.1.2. We will show that extensions of notions from the previous model capture the difficulties inherent to these models and allow for the derivation of the same statements about the heuristic.

Again there is often a lower bound on $y^*(X,\mathbf{w})$ that can be written as the product of a function $w: \mathbb{R}^n \to \mathbb{R}$ and in this case a set function $h: \mathcal{X} \to \mathbb{R}$: i.e., asymptotically

$$y^*(X,\mathbf{w}) \geqslant h(X)g(\mathbf{w}) \text{ (wp1)}.$$

Using an approximation v of $g(\mathbf{w})$, having the same qualities as in the previous model, and defining $y^{H_1}(X) = vh(X)$ leads to the asymptotic lower bound function

$$z^{H_1}(X) = f(X) + vh(X) \leqslant f(X) + y^*(X,\mathbf{w}) = z^*(X,\mathbf{w}) \text{ (wp1)}.$$

The determination of a set X^{H_1} that minimizes $z^{H_1}(X)$ is a *combinatorial selection problem*. Theorem 1.4(i) can be reformulated appropriately such as to hold for this new situation.

THEOREM 1.5. *If $f: \mathcal{X} \to \mathbb{R}$ and H_1 and H_2 are such that*
(A) $y^{H_1}(X) = vh(X) \leqslant y^*(X,\mathbf{w})$ (wp1) *asymptotically and*

(B) $y^{H_2}(X^{H_1},\mathbf{w})/y^{H_1}(X^{H_1})\rightarrow 1$ (wp1),
then

$$z^{H_2}(X^{H_1},\mathbf{w})/z^{\star}(X^{\circ}(\mathbf{w}),\mathbf{w}) \rightarrow 1 \text{ (wp1)}. \quad \square$$

In most cases the above combinatorial selection problem is NP-hard, which suggests that approximating X^{H_1} is the only practical alternative. For the approximation X' we cannot guarantee that

$$z^{H_1}(X') \leqslant \min_{X \in \mathfrak{X}}\{z^{\star}(X,\mathbf{w})\} \text{ (wp1)}.$$

Nevertheless, in Theorem 1.6 we show that the same statement of Theorem 1.5 can be established if we add an extra condition which requires that $z^{H_1}(X')$ is a sufficiently good approximation of $z^{H_1}(X^{H_1})$.

THEOREM 1.6. *If* $f:\mathfrak{X}\rightarrow\mathbb{R}$ *and* H_1 *and* H_2 *are such that*
(A) $y^{H_1}(X)=vh(X)\leqslant y^{\star}(X,\mathbf{w})$ (wp1) *asymptotically*,
(B) $y^{H_2}(X',\mathbf{w})/y^{H_1}(X')\rightarrow 1$ (wp1), *and*
(C) $z^{H_1}(X')/z^{H_1}(X^{H_1})\rightarrow 1$
then

$$z^{H_2}(X',\mathbf{w})/z^{\star}(X^{\circ}(\mathbf{w}),\mathbf{w})\rightarrow 1 \quad \text{(wp1)}.$$

PROOF. Because of Condition A we can bound $z^{\star}(X,w)$ from below (asymptotically with probability 1) and from above (deterministically) by

$$f(X)+y^{H_1}(X) = z^{H_1}(X) \leqslant z^{\star}(X,\mathbf{w})\leqslant z^{H_2}(X,\mathbf{w}) = f(X)+y^{H_2}(X,\mathbf{w}) \text{ (wp1)},$$

so that

$$f(X^{H_1})+y^{H_1}(X^{H_1}) = z^{H_1}(X^{H_1})\leqslant z^{\star}(X^{\circ}(\mathbf{w}),\mathbf{w})\leqslant z^{H_2}(X',\mathbf{w})$$
$$= f(X')+y^{H_2}(X',\mathbf{w}) \text{ (wp1)}.$$

It is obvious that, asymptotically,

$$\frac{z^{H_2}(X',\mathbf{w})}{z^{\star}(X^{\circ}(\mathbf{w}),\mathbf{w})} \leqslant \frac{z^{H_2}(X',\mathbf{w})}{z^{H_1}(X^{H_1})} \quad \text{(wp1)}.$$

The latter term is equal to

$$\frac{z^{H_2}(X',\mathbf{w})}{z^{H_1}(X')} \cdot \frac{z^{H_1}(X')}{z^{H_1}(X^{H_1})}.$$

Conditions B and C together yield the theorem. $\square$

We finally note that, if in Theorems 1.4, 1.5 and 1.6 the probabilistic nature of the conditions is different (in probability or in expectation instead of with probability 1), then the probabilistic statements about the asymptotic clairvoyancy that can be proved differ correspondingly.

1.6. A general recourse problem

In this section we study a two-level planning problem, in which a recourse on the aggregate level decision is possible, after a realization of the random parameters has become known.

As in the first model of the previous section we assume that we have to decide on a number X of resources at the aggregate level at a cost c each, so that $f(X) = cX$. Given a realization w of the random parameters at the detailed level, it is possible to acquire some extra resources $X_1(w)$ at a price $c_1 > c$ each or to sell some of the acquired resources $X_2(w)$ at a price $c_2 < c$. With the resources that are ultimately available we have to solve the detailed level combinatorial problem, the optimal value of which is now denoted by $y^*(X + X_1(w) - X_2(w), w)$. The total cost of the two decisions is defined by $z_R^*(X, w) = cX + X_1(w) - X_2(w) + y^*(X + X_1(w) - X_2(w), w)$.

The two-stage decision problem is to determine a value $X^* \in \mathbb{N}$, for which

$$E z_R^*(X^*, \mathbf{w}) = \min_{X \in \mathbb{N}} \{ E z_R^*(X, \mathbf{w}) \}. \tag{1.2}$$

The optimal solution of the distribution problem is equal to the one of the first model of Section 1.5 as this problem aims for the optimal choice of X given a realization of the random parameters before the aggregate level decision is made. It is obvious that here for each realization w of $\mathbf{w}$ $X_1(w) = X_2(w) = 0$.

Along the lines of Section 1.5 we design and analyze a hierarchical system for the solution of the two-stage decision problem. Again we assume that an almost sure lower bound on $y^*(X, w)$ exists that can for each X be written as

$$\frac{g(\mathbf{w})}{X^\gamma} < y^*(X, \mathbf{w}) \quad (\text{wp1})$$

($\gamma > 0$). This implies that asymptotically

$$\frac{g(\mathbf{w})}{(X + X_1(\mathbf{w}) - X_2(\mathbf{w}))^\gamma} < y^*(X + X_1(\mathbf{w}) - X_2(\mathbf{w})) \quad (\text{wp1}).$$

Hence, asymptotically, we can bound the second stage cost from below by

$$\min_{X_1(\mathbf{w}), X_2(\mathbf{w}) \in \mathbb{N}} \left\{ c_1 X_1(\mathbf{w}) - c_2 X_2(\mathbf{w}) + \frac{g(\mathbf{w})}{(X + X_1(\mathbf{w}) - X_2(\mathbf{w}))^\gamma} \right\} \quad (\text{wp1}).$$

This minimum is determined from the first order Kuhn-Tucker conditions

$$c_1 - \gamma g(\mathbf{w}) / (X + X_1(\mathbf{w}) - X_2(\mathbf{w}))^{\gamma+1} = 0$$

$$c_2 - \gamma g(\mathbf{w}) / (X + X_1(\mathbf{w}) - X_2(\mathbf{w}))^{\gamma+1} = 0.$$

From these conditions we obtain:

$$c_2 X^{\gamma+1} > \gamma g(\mathbf{w}) \Rightarrow X_2(\mathbf{w}) = X - \left(\frac{\gamma g(\mathbf{w})}{c_2} \right)^{\frac{1}{\gamma+1}}, \; X_1(\mathbf{w}) = 0;$$

$$c_2 X^{\gamma+1} \leq \gamma g(\mathbf{w}) \leq c_1 X^{\gamma+1} \Rightarrow X_2(\mathbf{w}) = X_1(\mathbf{w}) = 0;$$

$$c_1 X^{\gamma+1} < \gamma g(\mathbf{w}) \Rightarrow X_2(\mathbf{w}) = 0, \; X_1(\mathbf{w}) = \left(\frac{\gamma g(\mathbf{w})}{c_1}\right)^{\frac{1}{\gamma+1}} - X.$$

Now, a lower bound on the expected second stage cost is given by

$$z_R^{H_1}(X) = cX + \int\limits_{\frac{c_1}{\gamma}X^{\gamma+1}}^{\infty} \left(c_1\left(\left(\frac{\gamma g(w)}{c_1}\right)^{\frac{1}{\gamma+1}} - X\right) + \frac{g(w)}{\left(X + \left(\frac{\gamma g(w)}{c_1}\right)^{\frac{1}{\gamma+1}} - X\right)^{\gamma}}\right) dF(g(w))$$

$$+ \int\limits_{\frac{c_2}{\gamma}X^{\gamma+1}}^{\frac{c_1}{\gamma}X^{\gamma+1}} \frac{g(w)}{X^{\gamma}} dF(g(w))$$

$$- \int\limits_{0}^{\frac{c_2}{\gamma}X^{\gamma+1}} \left(c_2\left(X - \left(\frac{\gamma g(w)}{c_2}\right)^{\frac{1}{\gamma+1}}\right) - \frac{g(w)}{\left(X - \left(X - \left(\frac{\gamma g(w)}{c_2}\right)^{\frac{1}{1+\gamma}}\right)\right)^{\gamma}} dF(g(w))\right.$$

$$= cX + \int\limits_{\frac{c_1}{\gamma}X^{\gamma+1}}^{\infty} \left((\gamma c^{\gamma} + g(w))^{\frac{1}{\gamma+1}} - c_1 X + \left(\left(\frac{c_1}{\gamma}\right)^{\gamma} g(w)\right)^{\frac{1}{\gamma+1}}\right) dF(g(w))$$

$$+ \int\limits_{\frac{c_2}{\gamma}X^{\gamma+1}}^{\frac{c_1}{\gamma}X^{\gamma+1}} \frac{c_2}{\gamma} X^{\gamma+1} g(w)/X^{\gamma} dF(g(w))$$

$$- \int\limits_{0}^{\frac{c_2}{\gamma}X^{\gamma+1}} \left(c_2 X (\gamma c^{\gamma} g(w))^{\frac{1}{\gamma+1}} - \left(\left(\frac{c_2}{\gamma}\right)^{\gamma} g(w)\right)^{\frac{1}{\gamma+1}}\right) dF(g(w)).$$

To obtain the heuristic first stage decision X^{H_1} we minimize $z^{H_1}(X)$. The first order condition is given by

$$c - (\gamma c_1^{\gamma} c_1 X^{\gamma+1}/\gamma)^{\frac{1}{\gamma+1}} f(c_1 X^{\gamma+1}/\gamma) \frac{\gamma+1}{\gamma} c_1 X^{\gamma} - c_1 \int\limits_{c_1 X^{\gamma+1}/\gamma} dF(g(w)) +$$

$$c_1 X f(c_1 X^{\gamma+1}/\gamma) \frac{\gamma+1}{\gamma} c_1 X^{\gamma} - \left(\frac{c_1^{\gamma} c_1 X^{\gamma+1}/\gamma}{\gamma^{\gamma}}\right)^{\frac{1}{\gamma+1}} f(c_1 X^{\gamma+1}/\gamma) \frac{\gamma+1}{\gamma} c_1 X^{\gamma} -$$

$$\frac{\gamma}{X^{\gamma+1}} \int\limits_{c_2 X^{\gamma+1}/\gamma}^{c_1 X^{\gamma+1}/\gamma} g(w) dF(g(w)) +$$

$$\frac{c_1 X^{\gamma+1}}{\gamma X^{\gamma}} f(c_1 X^{\gamma+1}/\gamma) \frac{\gamma+1}{\gamma} c_1 X^{\gamma} - \frac{c_2 X^{\gamma+1}}{\gamma X^{\gamma}} f(c_2 X^{\gamma+1}/\gamma) \frac{\gamma+1}{\gamma} c_2 X^{\gamma} -$$

$$c_2 \int_0^{c_2 X^{\gamma+1}/\gamma} dF(g(w)) - c_2 X f(c_2 X^{\gamma+1}/\gamma) \frac{\gamma+1}{\gamma} c_2 X^\gamma +$$

$$(\gamma c_2^\gamma c_2 X^{\gamma+1}/\gamma)^{\frac{1}{\gamma+1}} f(c_2 X^{\gamma+1}/\gamma) \frac{\gamma+1}{\gamma} c_2 X^\gamma +$$

$$\left(\frac{c_2^\gamma c_2 X^{\gamma+1}/\gamma}{\gamma^\gamma}\right)^{\frac{1}{\gamma+1}} f(c_2 X^{\gamma+1}/\gamma) \frac{\gamma+1}{\gamma} c_2 X^\gamma = 0.$$

This reduces to

$$c - c_1 \int_{c_1 X^{\gamma+1}/\gamma}^\infty dF(g(w)) - \frac{\gamma}{X^{\gamma+1}} \int_{c_2 X^{\gamma+1}/\gamma}^{c_1 X^{\gamma+1}/\gamma} g(w) dF(g(w)) - \qquad (1.3)$$

$$c_2 \int_0^{c_2 X^{\gamma+1}/\gamma} dF(g(w)) = 0.$$

As in Section 1.5 we assume that an approximation ν of $g(\mathbf{w})$ exist that is asymptotically accurate with probability 1: $g(\mathbf{w})/\nu \to 1$ (wp1). We will show that, if moreover asymptotically $E(g(w)/\nu) = 1$, equation (1.3) is asymptotically satisfied by $X^{H_1} = (\gamma\nu/c)^{1/(\gamma+1)}$. Substitution of this value in (1.1) yields

$$c - c_2 \int_0^{\frac{c_2}{c}\nu} dF(g(w)) - \frac{c}{\nu} \int_{\frac{c_2}{c}\nu}^{\frac{c_1}{c}\nu} g(w) dF(g(w)) - c_1 \int_{\frac{c_1}{c}\nu}^\infty dF(g(w)). \qquad (1.4)$$

Since $g(w)/\nu \to 1$ (wp1), the second and the fourth term in (1.4) tend to 0, and the third term tends to $-c$, which makes (1.4) asymptotically equal to 0. That X^{H_1} is indeed a global minimum is shown when we consider the second derivative of $z^{H_1}(X)$:

$$-c_2 f(c_2 X^{\gamma+1}/\gamma) \frac{\gamma+1}{\gamma} c_2 X^\gamma + \frac{\gamma(\gamma+1)}{X^{\gamma+2}} \int_{c_2 X^{\gamma+1}/\gamma}^{c_1 X^{\gamma+1}/\gamma} g(w) dF(g(w)) -$$

$$\frac{\gamma}{X^{\gamma+1}} c_1 X^{\gamma+1}/\gamma f(c_1 X^{\gamma+1}/\gamma) \frac{\gamma+1}{\gamma} c_1 X^\gamma + \frac{\gamma}{X^{\gamma+1}} c_2 X^{\gamma+1}/\gamma f(c_2 X^{\gamma+1}/\gamma)$$

$$\frac{\gamma+1}{\gamma} c_2 X^\gamma +$$

$$c_1 f(c_1 X^{\gamma+1}/\gamma) \frac{\gamma+1}{\gamma} c_1 X^\gamma = \frac{\gamma(\gamma+1)}{X^{\gamma+2}} \int_{c_2 X^{\gamma+1}/\gamma}^{c_1 X^{\gamma+1}/\gamma} g(w) dF(g(w)) > 0$$

We observe that this first stage decision is equal to the one in the first model of Section 1.5. (We omit the integer restriction on X^{H_1} as it does not influence asymptotic results.) For the second stage recourse decision we propose the following heuristic. If $\gamma g(w) > c_1 (X^{H_1})^{1/(\gamma+1)}$ then we acquire an extra number of resources $X_1^{H_2}(w) = (\gamma g(w)/c_1)^{1/(\gamma+1)} - X^{H_1}$. If $\gamma g(w) < c_2 (X^{H_1})^{1/(\gamma+1)}$ we sell

a number of resources $X_2^{H_2}(w) = X^{H_1} - (\gamma g(w)/c_2)^{1/(\gamma+1)}$. If $c_2(X^{H_1})^{1/(\gamma+1)} \leqslant \gamma g(w) \leqslant c_1(X^{H_1})^{1/(\gamma+1)}$ we neither acquire any extra resources nor sell any of the acquired resources.

Again we assume the existence of a heuristic H_3 for the solution of the ultimate combinatorial optimization problem that, given a number X of resources, produces a value $y^{H_3}(X,w)$ the expectation of which is asymptotic to $y^{H_1}(X) = \nu/X^\gamma$ The value produced by the entire hierarchical system (H_1,H_2,H_3) is given by

$$Ez_R^{H_2,H_3}(X^{H_1},\mathbf{w}) = cX^{H_1} + E(c_1 X_1^{H_2}(\mathbf{w}) -$$

$$c_2 X_2^{H_2}(\mathbf{w}) + y^{H_3}(X^{H_1} + X_1(\mathbf{w}) - X_2(\mathbf{w}),\mathbf{w})).$$

If $z^{H_1}(X) = cX + \nu/X^\gamma$ is the lower bound on the objective function of the problem without a recourse possibility, (see (1.)) then the following theorem establishes that there is asymptotically no recourse, and the solution value produced by the hierarchical system (H_1,H_2,H_3) is asymptotic to the optimal value of the problem without recourse.

THEOREM 1.7. *If H_1 and H_3 are such that*
(A) $y^{H_1}(X) = \nu/X^\gamma \leqslant y^*(X,\mathbf{w})$ (wp1) *asymptotically and*
(B) $y^{H_3}(X^{H_1},\mathbf{w})/y^{H_1}(X^{H_1}) \to 1$ (wp1),
then

(i) $X_2^{H_2}(w) \to 0$ (wp1);

(ii) $X_1^{H_2}(w) \to 0$ (wp1);

(iii) $Ez_R^{H_2,H_3}(X^{H_1},\mathbf{w})/Ez^*(X^*,\mathbf{w}) \to 1$.

PROOF. (i) For every $\epsilon > 0$ we have

$$Pr\{\lim_{n \to \infty} X_2^{H_2}(\mathbf{w}) > \epsilon\} = Pr\{\lim_{n \to \infty} \frac{g(\mathbf{w})}{\nu} < (1-\epsilon)^{\gamma+1}\frac{c_2}{c}\}.$$

The latter probability is 0 for every $\epsilon > 0$.
(ii) This is proved analogously to (i).
(iii) Given (i) and (ii), we can prove (iii) in the same way as we proved Theorem 1.4(i). $\square$

2

Hierarchical scheduling problems

In this chapter we consider job shop design and scheduling problems. The aggregate decision concerns the design of the shop while at the detailed level jobs are to be scheduled.

In [Armstrong & Hax 1977] and [Schwimer 1972] hierarchical systems have been described that use an integer programming model and simulation with an embedded heuristic sequencing rule to make the higher level decision. Then, the lower level decision is made with the heuristic sequencing rule.

We will present hierarchical systems and analyses of their performance for some specific job shop design and scheduling problems, following the lines of the previous chapter. Three different types of problems are considered in the three sections that follow. In all the models the cost directly related to the first stage decision is known with certainty, but there is only stochastic information about some job characteristics. Thus, we will assume throughout this chapter that the processing times of the jobs, denoted by w_j for job j, are independent identically distributed random variables with finite expectation μ. The random vector of processing times that corresponds to a problem with n jobs is denoted by $w = (w_1, \ldots, w_n)$. Let $\mathcal{W}$ denote the set of all possible realizations w of $\mathbf{w}$. Throughout this chapter we will use the notation $w_{sum} = \sum_{j=1}^{n} w_j$ and $w_{max} = \max_{j=1,\ldots,n} \{w_j\}$.

In Section 2.1 the aggregate problem is to determine the *number and types of machines* to be acquired, while at the detailed level, given a realization of the job characteristics, we have to schedule the jobs on the available machines so as to minimize the *maximum job completion time*. There are several subsections in which a number of variants of the problem are studied. These variants are obtained by considering different types of machines and different stochastic assumptions on the number of jobs and on their processing times.

In Section 2.2 the first stage decision concerns the determination of a *delivery time* before which all jobs must be completed, while at the second stage we have to minimize the number of machines that allows the construction of a schedule for the jobs that satisfies the delivery time constraint set at the first stage. This problem is also known as the *bin-packing problem*.

The hierarchical scheduling problem presented in Section 2.3 asks again for the determination of a set of machines to be acquired at the higher level. The lower level problem is to schedule the jobs on the acquired machines such as to minimize the *sum of the job completion times*.

For each of the above problems we show how a simple heuristic has many of the performance qualities that were introduced in the previous chapter.

2.1. Maximum job completion time

The models studied in this section are all special cases of the following general stochastic programming model. At the first stage one has to select a subset X of machines to be acquired from a set $\mathfrak{M}$ of available parallel machines, given the number n of jobs to be processed and given the probability distribution of the vector $\mathbf{w} = (\mathbf{w}_1, \ldots, \mathbf{w}_n)$ of their processing times. Let $\mathfrak{X} = 2^{\mathfrak{M}}$ be the power set of $\mathfrak{M}$. The direct acquisition cost is given as a function $f: \mathfrak{X} \to \mathbb{R}$ of X.

The second stage problem is to determine a schedule for the jobs on the set X of machines selected at the first stage, given a realization w of $\mathbf{w}$, such that each machine processes at most one job at a time, each job is processed during an uninterrupted period of length equal to its processing time, and no job is processed prior to time 0. The second stage objective is to minimize the maximum job completion time or 'makespan'. Let $y^*(X,w)$ denote this minimum value. Without loss of generality we may assume that the cost per time unit of the second stage schedule is 1. The total cost of the first stage decision and an optimal second stage decision is denoted by $z^*(X,w) = f(X) + y^*(X,w)$.

The two-stage decision problem is to determine a set of machines $X^* \in \mathfrak{X}$ such that

$$Ez^*(X^*,\mathbf{w}) = \min_{X \in \mathfrak{X}}\{Ez^*(X,\mathbf{w})\}.$$

The distribution problem is to find a function $X^\circ: \mathbb{R}^n \to \mathfrak{X}$ that yields for each realization w of a $\mathbf{w}$ set of machines $X^\circ(w)$ such that

$$z^*(X^\circ(w),w) = \min_{X \in \mathfrak{X}}\{z^*(X,w)\}, \ \forall w \in \mathfrak{W}.$$

Since computing $y^*(X,w)$ is an NP-hard problem [Karp 1972], determining $Ey^*(X,\mathbf{w})$ as a function of X for an arbitrary given probability distribution of $\mathbf{w}$ seems virtually impossible. A heuristic approach is therefore the only practical method for the solution of large instances of this problem.

In Subsection 2.1.1 we investigate the case in which the available machines are *identical* in cost and speed, whereas in Subsection 2.1.2 they are *uniform*;

i.e., each machine has its own cost and speed. In Subsection 2.1.3 we consider two variants in which in addition to stochasticity in the processing times of the jobs there is uncertainty about their number. In the first variant n is still fixed but each job has a fixed probability p of entering the shop while in the second variant n itself is a random variable with a known probability distribution.

2.1.1. Identical machines. In the first variant of the model described above we assume that *identical* parallel machines are available, at cost c each. The first stage problem is to determine a *number* rather than a *set* of machines. If we assume that an unlimited number of machines is available, we can take $\mathcal{X}$ to be the set of positive integers. The direct first stage cost is given by $f(X)=cX$. Note that we are encoutering here the simple model of Section 1.5.

Along the lines of Section 1.5 we design a hierarchical system to solve the two-stage decision problem. The heuristic for the first stage is based on the replacement of $Ey^*(X,\mathbf{w})$ by an approximation $y^{H_1}(X)$. The approximation is derived by applying an idea that is fundamental to many hierarchical planning systems: we suppress the combinatorial fine structure of the second stage problem. The maximum job completion time when all the machines have equal workload, which is equal to w_{sum}/X, is obviously a *lower bound* on $y^*(X,w)$. The required estimate is now taken to be equal to the expectation of this lower bound: $y^{H_1}(X) = E\mathbf{w}_{sum}/X = n\mu/X$. The resulting first stage problem is then to determine the value X^{H_1} that minimizes the lower bound function $z^{H_1}(X) = cX + n\mu/X$. We note that $z^{H_1}(X)$ is a convex function. Its derivative is equal to 0 for $X = \sqrt{n\mu/c}$. Since X^{H_1} must be a positive integer, X^{H_1} is determined as the value $X \in \{\lfloor \sqrt{n\mu/c} \rfloor, \lceil \sqrt{n\mu/c} \rceil\} \cap \mathbb{N}$ that minimizes $z^{H_1}(X)$.

At the second stage we schedule the jobs on the machines according to a *list scheduling rule*: the jobs are placed in an arbitrary fixed order and at each step the next job on the list is assigned to the earliest available machine (see Figure 2.1). Let $y^{H_2}(X,w)$ denote the earliest time when all jobs are completed under this heuristic, for given X and w, and let $z^{H_2}(X,w)=cX+y^{H_2}(X,w)$ be the corresponding total cost. The heuristic solution value provided by the combination of the first and second stage heuristic is then

$$z^{H_2}(X^{H_1},w) = cX^{H_1}+y^{H_2}(X^{H_1},w).$$

We will show that the heuristic (H_1,H_2) is asymptotically clairvoyant by verifying Conditions A and B of Theorem 1.4. Condition A is easily verified by the observation that the strong law of large numbers implies that $\mathbf{w}_{sum}/n\mu \to 1$ (wp1), so that, asymptotically,

$$n\mu/X \leqslant y^*(X,\mathbf{w}) \quad (wp1).$$

To show that the second condition of Theorem 1.4 is satisfied as well we make an additional assumption on the distribution function of the processing times.

ASSUMPTION 2.1. *The processing times* $\mathbf{w}_1, \ldots, \mathbf{w}_n$ *have finite second moment.*
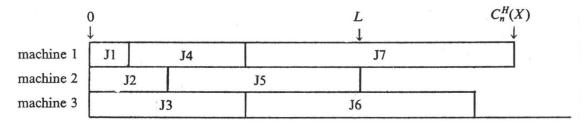
FIGURE 2.1.
Illustration of the list scheduling heuristic.

Under this assumption the following lemma can be proved (cf. [Dempster et al. 1983]).

LEMMA 2.1. *Under Assumption 2.1,*
(i) $\lim_{n \to \infty} \mathbf{w}_{max}/\sqrt{n} = 0$ (wp1);
(ii) $\lim_{n \to \infty} E \mathbf{w}_{max}/\sqrt{n} = 0$. $\square$

The following theorem characterizes the value of an optimal solution of the second stage scheduling problem, given a number X of machines, and shows that, as the number of jobs tends to infinity, the relative error made by the list scheduling rule H_2 almost surely tends to 0.

THEOREM 2.1. *If $X = O(\sqrt{n})$, then, under Assumption 2.1,*
(i) $y^*(X,\mathbf{w})/(n\mu/X) \to 1$ (wp1);
(ii) $y^{H_2}(X,\mathbf{w})/(n\mu/X) \to 1$ (wp1).

PROOF. Consider a schedule produced by the list scheduling rule on X machines for a realization w of $\mathbf{w}$. Let L be the latest time that all machines are occupied and let job k be completed last (cf. Figure 2.1). It follows from the nature of list scheduling that

$$y^{H_2}(X,w) \leqslant L + w_k \leqslant \frac{w_{sum}}{X} + w_k \leqslant \frac{w_{sum}}{X} + w_{max}. \tag{2.1}$$

Therefore, for each realization w of $\mathbf{w}$,

$$\frac{w_{sum}}{X} \leqslant y^*(X,w) \leqslant y^{H_2}(X,w) \leqslant \frac{w_{sum}}{X} + w_{max}.$$

Division by $n\mu/X$ yields

$$\frac{w_{sum}}{n\mu} \leqslant \frac{y^*(X,w)}{n\mu/x} \leqslant \frac{y^{H_2}(X,w)}{n\mu/X}$$

$$\leqslant \frac{w_{\text{sum}}}{n\mu} + \frac{Xw_{\text{max}}}{n\mu}.$$

The strong law of large numbers implies that, since μ is finite, $Pr\{(w_{\text{sum}}/n\mu{\to}1\} = 1$. This observation, Lemma 2.1 (i) and the assumption that $X = O(\sqrt{n})$ imply the theorem. $\square$

In particular, Theorem 2.1 holds for $X^{H_1} \in \{\lceil \sqrt{n\mu/c} \rceil, \lfloor \sqrt{n\mu/c} \rfloor\} \cap \mathbb{N}$, so that Condition B of Theorem 1.4 is also satisfied. Theorem 1.4 now implies asymptotic clairvoyancy of the first stage heuristic H_1 and of the overall hierarchical system (H_1, H_2):

THEOREM 2.2. *Under Assumption 2.1,*
(i) $z^{H_2}(X^{H_1}, \mathbf{w})/z^*(X^\circ(\mathbf{w}), \mathbf{w}) {\to} 1$ (wp1);
(ii) $X^{H_1}/X^\circ(\mathbf{w}) {\to} 1$ (wp1). $\square$

It is easy to verify that, under the reasonable assumption that the distribution of the processing times has bounded support (i.e., constants w^L and w^U exist such that $Pr\{w^L < \mathbf{w} < w^U\} = 1$), Assumptions 1.1 up to 1.5 from Section 1.4 are valid. This implies that the heuristic (H_1, H_2) satisfies a wide range of asymptotic optimality properties (cf. Figure 1.1).

In addition to the probabilistic and asymptotic statements on the behavior of (H_1, H_2), we can give a bound on the relative *worst case* error the heuristic produces when solving the two-stage decision problem.

THEOREM 2.3.

$$\frac{Ez^{H_2}(X^{H_1}, \mathbf{w})}{Ez^*(X^*, \mathbf{w})} \leqslant 1 + \frac{E\mathbf{w}_{\text{max}}}{2\sqrt{cn}\mu}.$$

PROOF. Taking expectations in inequality (2.1) yields

$$Ey^{H_2}(X^{H_1}, \mathbf{w}) \leqslant \frac{E\mathbf{w}_{\text{sum}}}{X^{H_1}} + E\mathbf{w}_{\text{max}}.$$

Hence,

$$Ez^{H_2}(X^{H_1}, \mathbf{w}) \leqslant cX^{H_1} + \frac{n\mu}{X^{H_1}} + E\mathbf{w}_{\text{max}}. \tag{2.2}$$

By the definition of X^{H_1}

$$Ez^*(X^*, \mathbf{w}) \geqslant cX^{H_1} + \frac{n\mu}{X^{H_1}} \geqslant 2\sqrt{cn\mu}. \tag{2.3}$$

(2.2) and (2.3) together imply the theorem. $\square$

From this theorem and Lemma 2.1(ii) asymptotic expectation-optimality of (H_1, H_2) follows immediately.

2.1.2. Uniform machines. The model studied in this section is an extension of the model in Section 2.1.1. The difference is that, at the first stage, one has to select a subset from a set $\mathfrak{M}$ of *uniform machines,* knowing the *cost* c_i and *speed* s_i of each machine $i \in \mathfrak{M}$. When at the second stage, job j is assigned to machine i, it has to be processed during a period of length w_j / s_i. The set of feasible first stage decisions is $\mathfrak{X} = 2^{\mathfrak{M}}$, the power set of $\mathfrak{M}$. In this model $f(X) = \sum_{i \in X} c_i$. Corresponding to each $X \in \mathfrak{X}$ we define $s(X) = \sum_{i \in X} s_i$. We assume that constants c^L, c^U, s^L, s^U exist such that $c^L \leqslant c_i \leqslant c^U$ and $s^L \leqslant s_i \leqslant s^U$ for all $i \in \mathfrak{M}$.

The reader will realize that this problem is even harder than the problem with identical machines, so that a heuristic seems unavoidable for solving problem instances of a large size. With this problem we find ourselves in the situation of the second model sketched in Section 1.5. A hierarchical system will be devised along the lines of that section; by verifying Conditions A, B and C of Theorem 1.6 we will prove its asymptotic clairvoyancy with probability 1.

It is not difficult to verify that $w_{\text{sum}}/s(X)$ is a lower bound on $y^*(X, w)$ for each X and for each realization w of $\mathbf{w}$. As an approximation for $Ey^*(X, \mathbf{w})$ we take the expectation of this lower bound: $y^{H_1}(X) = E\mathbf{w}_{\text{sum}}/s(X) = n\mu/s(X)$. The resulting function to be minimized at the first stage of the heuristic is

$$z^{H_1}(X) = f(X) + n\mu/s(X). \tag{2.4}$$

It is no longer possible here to imitate the procedure developed in the previous subsection. The determination of a set X^{H_1} that minimizes (2.4) is a combinatorial selection problem and it is very unlikely that this problem can be solved in polynomial time.

LEMMA 2.2. *The problem of minimizing $z^{H_1}(X)$ over all $X \in \mathfrak{X}$ is NP-hard.*

PROOF. We will show that the problem of minimizing (2.4) is a generalization of the following known NP-complete problem [Garey & Johnson 1979]:

PARTITION: Given a set $\mathfrak{I} = \{1, \ldots, t\}$ and positive integers $a_t, \ldots, a_1, b$ with $\sum_{i \in \mathfrak{I}} a_i = 2b$, does there exist a subset $T \subseteq \mathfrak{I}$ such that $\sum_{i \in T} a_i = b$?

Given any instance of PARTITION, we construct an instance of our minimization problem defining $\mathfrak{M} = \mathfrak{I}$, $c_i = s_i = a_i$ $(i \in \mathfrak{I})$ and $n\mu = b^2$. It is easily verified that for a subset $T \subseteq \mathfrak{I}$, $\sum_{i \in T} a_i = b$ if and only if $f(T) + n\mu/s(T) \leqslant 2b$. Hence, PARTITION yields a positive answer if and only if the minimum value of $z^{H_1}(X)$ is at most $2b$. $\square$

If we aim to find a hierarchical planning system that solves our problem in polynomial time, then Lemma 2.2 suggests that already the solution to the heuristic first stage problem is to be approximated. We propose a *greedy* heuristic, which derives its name from its appetite for immediate improvement.

We define the ratio's $q_i = c_i/s_i$ $(i \in \mathfrak{M})$ and renumber the machines according to nondecreasing q_i. Let $C_i = \sum_{k=1}^{i}$, $S_i = \sum_{k=1}^{i} s_i$, $Z_i = C_i + n\mu/S_i$ $(i \in \mathfrak{M})$ and $Z_0 = \infty$. The greedy heuristic selects a subset $X^G = \{1,...,g\} \in \mathfrak{X}$ where g is the largest index such that $Z_{g-1} > Z_g$. The greedy decision X^G is an approximation of X^{H_1}. An absolute worst case bound on the quality of this approximation is established by Lemma 2.3.

LEMMA 2.3. *The greedy solution X^G satisfies*
(i) $Z_g = min_{i \in \mathfrak{M}} \{Z_i\}$;
(ii) $z^{H_1}(X^G) \leqslant z^{H_1}(X^{H_1}) + c^U$.

PROOF. (i) The values Z_i define a piecewise linear function $Z(S)$ on the interval $[0, s(\mathfrak{M})]$ as follows: if $S = \lambda S_i + (1-\lambda)S_{i+1}$ for some $i \geqslant 0$ and $\lambda \in [0,1]$, then $Z(S) = \lambda Z_1 + (1-\lambda)Z_{i+1}$. This function is convex, since its slope over $[S_i, S_{i+1}]$ increases with i:

$$\frac{Z_{i+1} - Z_i}{S_{i+1} - S_i} - \frac{Z_i - Z_{i-1}}{S_i - S_{i-1}} = q_{i+1} - q_i + \frac{(s_{i+1} + s_i)n\mu}{S_{i+1} S_i S_{i-1}} > 0.$$

Therefore, S_g is uniquely determined as the smallest value of S for which $Z(S)$ takes on its minimum.
(ii) There exists an $i \geqslant 0$ such that $S_i \leqslant s(X^{H_1}) \leqslant S_{i+1}$. Since C_i is the minimum machine cost at which a total speed S_i can be achieved, we have $C_i \leqslant f(X^{H_1})$. Hence

$$z^{H_1}(X^G) - z^{H_1}(X^{H_1}) \leqslant (C_{i+1} + \frac{n\mu}{S_{i+1}}) - (f(X^{H_1}) + \frac{n\mu}{s(X^{H_1})})$$

$$\leqslant C_{i+1} - C_i = c_{i+1} \leqslant c^u. \quad \square$$

The heuristic for the second stage problem again assigns the jobs to the machines according to a list scheduling rule. Let $y^{H_2}(X,w)$ denote the earliest time by which all jobs are completed under this rule, for given $X \in \mathfrak{X}$ and $w \in \Omega$, and let $z^{H_2}(X,w) = f(X) + y^{H_2}(X,w)$. The overall procedure (H_1, H_2) produces a solution with value

$$z^{H_2}(X^G, w) = f(X^G) + y^{H_2}(X^G, w).$$

For an asymptotic analysis of the heuristic (H_1, H_2), we first make some assumptions on the set $\mathfrak{M}$ of available machines. It is reasonable to assume that the values c^L, c^U, s^L and s^U, defined before, are constants rather than problem instance dependent variables. This will imply that the number of selected machines grows as $\sqrt{n}$, as in Section 2.1.1. It is then also reasonable to assume that the number of available machines grows faster than $\sqrt{n}$, but remains polynomially bounded in n in order to allow an efficient implementation of the greedy heuristic. We therefore require the following.

ASSUMPTION 2.2. *The parameters c^L, c^U, s^L, s^U are fixed constants. Moreover,*

there exist constants $D > 0$, $D' > 0$, $d' \geqslant d > 0$ such that $Dn^{1/2+d} \leqslant |\mathfrak{M}| \leqslant D'n^{1/2+d'}$.

To verify Condition A of Theorem 1.6 we notice that the strong law of large numbers implies that asymptotically

$$\frac{n\mu}{s(X)} \leqslant y^*(X, \mathbf{w}) \quad \text{(wp1)}.$$

From Lemma 2.3 (ii) it follows that

$$1 \leqslant \frac{z^{H_1}(X^G)}{z^{H_1}(X^{H_1})} \leqslant \frac{z^{H_1}(X^{H_1}) + c^U}{z^{H_1}(X^{H_1})}.$$

As the latter term tends to 1 when n tends to infinity, Condition C of Theorem 1.6 is also satisfied. To verify Condition B we need a preliminary lemma that bounds the order of magnitude of the sum of the speeds of the machines selected by the greedy heuristic. We use the notation $h_1(n) = \Theta(h_2(n))$ to indicate that there exist constants $C > 0$ and $C' > 0$ such that $Ch_2(n) \leqslant |h_1(n)| \leqslant C'h_2(n)$ for n sufficiently large.

LEMMA 2.4. *Under Assumptions 2.1 and 2.2, $s(X^G) = \Theta(\sqrt{n})$.*

PROOF. Let $q^L = c^L/s^U$ and $q^U = c^U/s^L$. We observe that $z^{H_1}(X^{H_1})$ is not greater than the minimum lower bound value obtainable under the assumption that $c_i = c^U$ and $s_i = s^L$ for all $i \in \mathfrak{M}$. We have seen in Subsection 2.1.1 that this value tends to $2\sqrt{q^U n\mu}$. Hence we may choose any $q^* > q^U$ to insure that for n sufficiently large

$$q^L s(X^G) + \frac{n\mu}{s(X^G)} \leqslant z^{H_1}(X^G)$$

$$\leqslant z^{H_1}(X^{H_1}) + c^U \quad \text{(by Lemma 2.3(ii))}$$

$$\leqslant 2\sqrt{q^* n\mu} + c^U = \Theta(\sqrt{n}).$$

Thus, a constant C' exists such that

$$q^L s(X^G) + \frac{n\mu}{s(X^G)} \leqslant C'\sqrt{n},$$

which implies $D\sqrt{n} \leqslant s(X^G) \leqslant D'\sqrt{n}$ for

$$D = \frac{C' - \sqrt{C'^2 - 4q^L\mu}}{2q^L}, \quad D' = \frac{C' + \sqrt{C'^2 - 4q^L\mu}}{2q^L}. \quad \square$$

In the following theorem we give a characterization of the optimal value of the second stage scheduling problem, and show that as the number of jobs tends to infinity, the relative error of list scheduling tends to 0 almost surely.

THEOREM 2.4. *If $s(X) = O(\sqrt{n}\,)$, then, under Assumption 2.1,*
(i) $y^*(X,\mathbf{w})/(n\mu/s(X)) \to 1$ (wp1);
(ii) $y^{H_2}(X,\mathbf{w})/(n\mu/s(X)) \to 1$ (wp1).

PROOF. For every realization w of $\mathbf{w}$ we have

$$\frac{w_{\text{sum}}}{s(X)} \leqslant y^*(X,w) \leqslant y^{H_2}(X,w) \leqslant \frac{w_{\text{sum}}}{s(X)} + \frac{w_{\text{max}}}{s^L}. \tag{2.5}$$

Division by $n\mu/s(X)$ yields

$$\frac{w_{\text{sum}}}{n\mu} \leqslant \frac{y^*(X,w)}{n\mu/s(X)} \leqslant \frac{y^{H_2}(X,w)}{n\mu/s(X)}$$

$$\leqslant \frac{w_{\text{sum}}}{n\mu} + \frac{s(X)w_{\text{max}}}{n\mu s^L}.$$

By the strong law of large numbers, $(\mathbf{w}_{\text{sum}}/n\mu \to 1$ (wp1). Due to the finiteness of the second moment of $\mathbf{w}_j$, $\mathbf{w}_{\text{max}}/\sqrt{n} \to 0$ (wp1) (cf. Lemma 2.1(i)). The assumption that $s(X) = O(\sqrt{n}\,)$ implies (i) and (ii). $\square$

The combination of Lemma 2.4 and Theorem 2.4 implies that Condition B of Theorem 1.6 holds for (H_1, H_2). Theorem 1.6 then establishes the asymptotic clairvoyancy of the hierarchical system with probability 1.

THEOREM 2.5. *Under Assumption 2.1,*

$$\frac{z^{H_2}(X^G,\mathbf{w})}{z^*(X^\circ(\mathbf{w}),\mathbf{w})} \to 1 \quad \text{(wp1)}. \quad \square$$

As in the previous subsection, we can derive a bound on the relative difference between the heuristic solution value to the two-stage decision problem and the optimal value.

THEOREM 2.6.

$$\frac{Ez^{H_2}(X^G,\mathbf{w})}{Ez^*(X^*,\mathbf{w})} \leqslant 1 + \frac{c^U + E\mathbf{w}_{\text{max}}/s^L}{2\sqrt{c^L n\mu/s^U}}.$$

PROOF. Taking expectations in (2.5) yields

$$Ey^{H_2}(X^G,\mathbf{w}) \leqslant \frac{E\mathbf{w}_{\text{sum}}}{s(X)} + \frac{E\mathbf{w}_{\text{max}}}{s^L}.$$

Hence,

$$Ez^{H_2}(X^G,\mathbf{w}) \leqslant f(X^G) + \frac{n\mu}{s(X^G)} + \frac{E\mathbf{w}_{\text{max}}}{s^L}.$$

Application of Lemma 2.3 (ii) yields

$$Ez^{H_2}(X^G, \mathbf{w}) \leq f(X^{H_1}) + \frac{n\mu}{s(X^{H_1})} + c^U + \frac{E\mathbf{w}_{\max}}{s^L}. \tag{2.6}$$

By the definition of X^{H_1}

$$Ez^*(X^*, \mathbf{w}) \geq f(X^{H_1}) + \frac{n\mu}{s(X^{H_1})} \tag{2.7}$$

$$\geq \frac{c^L}{s^U} s(X^{H_1}) + \frac{n\mu}{s(X^{H_1})}$$

$$\geq 2\sqrt{c^L n\mu/s^U}.$$

Inequalities (2.6) and (2.7) establish the theorem. $\square$

Lemma 2.1(ii) applied to this worst-case bound again shows directly that (H_1, H_2) is asymptotically expectation-optimal.

2.1.3. Random number of jobs. In this section we consider extensions of the model in Section 2.1.1. In addition to uncertainty about the processing times of the jobs we postulate uncertainty about their number, i.e. the number of jobs becomes a random variable. The following theorem will show that under certain conditions on the distribution of the number of jobs, the results derived for the problem in Section 2.1.1 carry over to the problem situation considered here.

THEOREM. 2.7. *Given a sequence of random variables* $\mathbf{x}_1, \mathbf{x}_2, \ldots$, *a random variable* $\mathbf{x}$, *and a sequence of random numbers* $\mathbf{N}_1, \mathbf{N}_2, \ldots$, *independent of* $\mathbf{x}_1, \mathbf{x}_2, \ldots$. *If* $\mathbf{x}_n \to \mathbf{x}$ (wp1) *and* $N_n \to \infty$ (wp1), *then* $x_{\mathbf{N}_n} \to \mathbf{x}$ (wp1).

PROOF. From the assumptions we have

$$Pr\{\forall_{\epsilon>0} \exists_{n(\epsilon)} \forall_{n \geq n(\epsilon)} : |\mathbf{x}_n - \mathbf{x}| < \epsilon\} = 1$$

and

$$Pr\{\forall_{\epsilon>0} \exists_{m(\epsilon)} \forall_{n > m(\epsilon)} : \mathbf{N}_n \geq n_\epsilon\} = 1.$$

These two probabilities combined yield

$$Pr\{\forall_{\epsilon>0} \exists_{k(\epsilon) = \max\{n(\epsilon), m(\epsilon)\}} \forall_{n \geq k(\epsilon)} : |\mathbf{x}_{\mathbf{N}_n} - \mathbf{x}| < \epsilon\} = 1. \quad \square$$

A special case is obtained when we consider the number of jobs fixed but each job enters the shop with fixed probability p. In this case the number of jobs that enter the shop is a random variable that has a binomial distribution with expectation np. For this situation we define the independent random variables $\delta_j, j = 1, \ldots, n$ as follows: $\delta_j = 1$ if job j enters the shop and $\delta_j = 0$ otherwise. Their common distribution is given by $Pr\{\delta_1 = 1\} = p$ and $Pr\{\delta_1 = 0\} = 1 - p$. The random variables $\tau_j = \delta_j \mathbf{w}_j$ $(j = 1, \ldots, n)$ model the stochastic assumptions

of this problem. They are independent and identically distributed and have common expectation

$$E\tau_1 = p\mu.$$

Any reader can verify that the design and the analysis of the hierarchical system for the problem of Section 2.1.1 carries over to this problem by simply substituting τ for w and $p\mu$ for μ.

In the second extension the number of jobs is a random variable $\mathbf{n}$. We assume that $\mathbf{n}$ has mean η. The first stage heuristic of the hierarchical system that we devise for this problem is again based on the estimate $E\mathbf{w}_{\text{sum}}/X$ of the optimal second stage cost. In this case

$$
\begin{aligned}
E\mathbf{w}_{\text{sum}} &= \sum_{n=1}^{\infty} E(\mathbf{w}_{\text{sum}}|\mathbf{n}=n)Pr\{\mathbf{n}=n\} \\
&= \mu\sum_{n=1}^{\infty} nPr\{\mathbf{n}=n\} \\
&= n\mu.
\end{aligned}
$$

For the solution of the second stage scheduling problem we use again the list scheduling rule. Analogously to the analysis in Subsection 2.1 it can be proved that the relative error produced by the hierarchical system (H_1, H_2), if compared to the optimal solution value of the two-stage decision problem, tends to 0 when μ tends to infinity.

2.2. Delivery time

The model studied in this section is the symmetric counterpart of the model described in Section 2.1. While in the latter model a set of machines was chosen first and a schedule of the jobs constructed next so as to complete all jobs as early as possible, here the decisions are reversed. At the first stage a decision is asked with respect to a *delivery time* Y, within which all jobs must be completed, under the same stochastic assumptions on the jobs as in Section 2.1. The cost of extending the delivery time by one unit is known with certainty and given by a function $f: \mathbb{R} \rightarrow \mathbb{R}$. At the second stage, given a realization w of the job characteristics, one has to determine the minimum number of *identical* parallel machines, denoted by $x^*(Y,w)$, that allows a nonpreemptive schedule satisfying the delivery time constraint Y set at the first stage. We assume that the cost of a machine is 1. The total cost of the first stage decision Y and an optimal second stage decision will be denoted by $z^*(Y,w)=f(Y)+x^*(Y,w)$. We will assume that $f(Y)=dY$, i.e., the cost of extending the delivery time by one unit is a fixed amount d.

The overall objective of the two-stage decision problem is to determine a value Y^* for Y for which

$$Ez^*(Y^*,\mathbf{w}) = \min_{Y \in \mathbf{R}}\{Ez^*(Y,\mathbf{w})\}.$$

The distribution problem is to find a function $Y^\circ: \mathbb{R}^n \rightarrow \mathbb{R}$ such that for every realization w of $\mathbf{w}$,

$$z^*(Y^\circ(w),w) = \min_{Y \in \mathbf{R}}\{z^*(Y,w)\}, \forall w \in \mathcal{W}.$$

We can interpret this hierarchical scheduling problem as a *hierarchical bin packing problem*. In a bin packing problem we are given n items with weights $w_1,...,w_n$ and an unlimited supply of bins, each with the same capacity. The problem is to pack the items into a minimum number of bins. If we identify jobs with items, processing times with weights, delivery times with capacities and machines with bins, then the above hierarchical scheduling problem is the problem of choosing a capacity for the bins at the first decision level with only probabilistic information about the item weights, and packing the items into as few bins as possible at the second level, when a realization of the item weights is given. We have chosen for the presentation of this problem as a scheduling problem, however, because the cost structure under consideration is more natural in the scheduling context than in the bin packing context.

As in the previous section, the second stage problem is NP-hard [Garey & Johnson 1979], which in conjunction with the stochastic formulation of the problem forces us again to look for heuristics for its solution. We will design a hierarchical system proceeding along the lines of the first model of Section 1.5. For each realization w of $\mathbf{w}$, w_{sum}/Y is a lower bound on $x^*(Y,w)$. The expectation of this lower bound provides an estimate of $Ex^*(Y,\mathbf{w})$: $x^{H_1}(Y)=E\mathbf{w}_{sum}/Y=n\mu/Y$. The resulting first stage heuristic problem is to find a value Y^{H_1} that minimizes the function

$$z^{H_1}(Y) = dY + \frac{n\mu}{Y}.$$

From the first order condition for this convex function it follows that $Y^{H_1} = \sqrt{n\mu/d}$.

For the solution of the second stage problem we use a heuristic that assigns jobs to machines according to the following rule. The jobs are placed in some fixed order, and the machines are indexed. The jobs are assigned to the machines in the given order, starting with job 1 on machine 1. Suppose job j is the next one to be scheduled, and let machine i be the highest indexed employed machine. Job j is assigned to machine i if this is feasible, i.e. if the addition of its processing time to the sum of the processing times of the jobs already assigned to machine i does not exceed the delivery time. Otherwise, job j is placed at the end of the list of jobs that are still to be scheduled, and job $j+1$ is the first job assigned to machine $i+1$ (see Figure 2.2). The number of machines required under this rule will be denoted by $x^{H_2}(Y,w)$, for a given Y and a given realization w of $\mathbf{w}$. The overall cost of the heuristic solution is

$$z^{H_2}(Y^{H_1},w) = dY^{H_1}+x^{H_2}(Y^{H_1},w).$$

To prove asymptotic clairvoyancy with probability 1 of both the first stage heuristic H_1 and the overall heuristic (H_1,H_2) we will verify Conditions A and B of Theorem 1.4 for this model. The strong law of large numbers allows us to verify the first condition. It implies that $\mathbf{w}_{sum}/n\mu \to 1$ (wp1) so that

asymptotically $z^{H_1}(Y)$ is a lower bound on $z^*(Y,w)$ with probability 1.
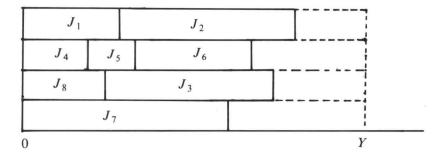

FIGURE 2.2. Heuristic solution

To verify the second condition we need Assumption 2.1 (see Section 2.1.1) which states finiteness of the second moment of the random processing times, so that Lemma 2.1 can be used as a preliminary. The following theorem gives asymptotic characterizations of the optimal and the heuristic number of machines in the second stage scheduling problem, given Y. It shows that both are asymptotic to the lower bound $n\mu/Y$.

THEOREM 2.8. *If there exists a constant $L \geqslant 0$ such that $\sqrt{n}/Y \to L$ and $Y/n \to 0$, then*
(i) $x^*(Y,\mathbf{w})/(n\mu/Y) \to 1$ (wp1);
(ii) $x^{H_2}(Y,\mathbf{w})/(n\mu/Y) \to 1$ (wp1).

PROOF. To bound $x^{H_2}(Y,w)$ from above we observe that after having employed $\lceil w_{\text{sum}}/Y \rceil$ machines we have tried to fit all jobs at least once and a set of at most $\lceil w_{\text{sum}}/Y \rceil - 1$ jobs remains to be scheduled, each of which has a processing time of at most w_{max}. For this set of jobs we need at most another $(\lceil w_{\text{sum}}/Y \rceil - 1)/ \lfloor Y/w_{\text{max}} \rfloor$ machines. Thus

$$\frac{w_{\text{sum}}}{Y} \leqslant x^*(Y,w) \leqslant x^{H_2}(Y,w) \leqslant \lceil \frac{w_{\text{sum}}}{Y} \rceil + (\lceil \frac{w_{\text{sum}}}{Y} \rceil - 1)/ \lfloor \frac{Y}{w_{\text{max}}} \rfloor$$

$$\leqslant \frac{w_{\text{sum}}}{Y} + 1 + \frac{w_{\text{sum}}}{Y}/(\frac{Y}{w_{\text{max}}} - 1)$$

$$= \frac{w_{\text{sum}}}{Y} + 1 + \frac{w_{\text{sum}}}{Y}(\frac{w_{\text{max}}/Y}{1-(w_{\text{max}}/Y)}).$$

Division by $n\mu/Y$ yields

$$\frac{\mathbf{w}_{sum}}{n\mu} \leqslant \frac{x^{\star}(Y,w)}{n\mu/Y} \leqslant \frac{x^{H_2}(Y,w)}{n\mu/Y}$$

$$\leqslant \frac{\mathbf{w}_{sum}}{n\mu} + \frac{Y}{n\mu} + \frac{\mathbf{w}_{sum}}{n\mu}\left(\frac{w_{max}/Y}{1-(w_{max}/Y)}\right)$$

The strong law of large numbers implies that $\mathbf{w}_{sum}/n\mu \to 1$ (wp1). Under the given conditions on Y and under Assumption 2.1, Lemma 2.1(ii) implies that $\mathbf{w}_{max}/Y \to 0$ (wp1). These two observations together with the condition that $Y/n \to 0$ imply the theorem. $\square$

In particular, the conditions on Y are satisfied by $Y^{H_1} = \sqrt{n\mu/d}$, so that Theorem 2.8 implies that Condition B of Theorem 1.4 is satisfied. Asymptotic clairvoyancy with probability 1 has now been established for both the first stage heuristic H_1 and the hierarchical system (H_1, H_2).

THEOREM 2.9. *Under Assumption 2.1,*
(i) $Y^{H_1}/Y^{\circ}(\mathbf{w}) \to 1$ (wp1);
(ii) $z^{H_2}(Y^{H_1},\mathbf{w})/z^{\star}(Y^{\circ}(\mathbf{w}),\mathbf{w}) \to 1$ (wp1). $\square$

Also here we can derive a bound on the relative difference between the heuristic and the optimal solution value to the two-stage decision problem.

THEOREM 2.10.

$$\frac{Ez^{H_2}(Y^{H_1},\mathbf{w})}{Ez^{\star}(Y^{\star},\mathbf{w})} \leqslant 1 + \frac{(E\mathbf{w}^2_{sum})^{1/2}}{2n\mu}(E((\sqrt{d}\,\mathbf{w}_{max}/\sqrt{n\mu})/(1-\sqrt{d}\,\mathbf{w}_{max}/\sqrt{n\mu}))^2)^{1/2}.$$

PROOF. Taking expectations in (2.8) yields

$$Ex^{H_2}(Y,\mathbf{w}) \leqslant \frac{E\mathbf{w}_{sum}}{Y} + 1 + \frac{E(\mathbf{w}_{sum}(\mathbf{w}_{max}/Y)/(1-\mathbf{w}_{max}/Y))}{Y}.$$

Hence,

$$Ez^{H_2}(Y^{H_1},\mathbf{w}) \leqslant dY^{H_1} + \frac{n\mu}{Y^{H_1}} + 1 + \frac{E(\mathbf{w}_{sum}(\mathbf{w}_{max}/Y^{H_1})/(1-\mathbf{w}_{max}/Y^{H_1}))}{Y^{H_1}}.$$

Application of Hölder's inequality (see [Serfling 1980, p. 352]) yields

$$Ez^{H_2}(Y^{H_1},\mathbf{w}) \leqslant dY^{H_1} + \frac{n\mu}{Y^{H_1}} + 1 \qquad (2.9)$$

$$+ \frac{(E\mathbf{w}^2_{sum})^{1/2}}{Y^{H_1}}(E((\mathbf{w}_{max}/Y^{H_1})/(1-\mathbf{w}_{max}/Y^{H_1}))^2)^{1/2}$$

$$= 2\sqrt{dn\mu} + 1$$

$$+ \frac{(E\mathbf{w}^2_{sum})^{1/2}}{\sqrt{n\mu/d}}(E((\sqrt{d}\,\mathbf{w}_{max}/\sqrt{n\mu})/(1-\sqrt{d}\,\mathbf{w}_{max}/\sqrt{n\mu}))^2)^{1/2}$$

By the definition of Y^{H_1}

$$Ez^*(Y^*,\mathbf{w}) \geqslant 2\sqrt{dn}\,\mu. \tag{2.10}$$

Inequalities (2.9) and (2.10) establish the theorem. $\square$

To derive asymptotic expectation-optimality of (H_1,H_2) from the above theorem we observe that from Lemma 2.1 it follows immediately that $(E((\sqrt{d}\,\mathbf{w}_{max}/\sqrt{n}\mu)/(1-\sqrt{d}\,\mathbf{w}_{max}/\sqrt{n}\mu))^2)^{1/2}\rightarrow 0$. Moreover we can rewrite $(E\mathbf{w}_{sum}^2)^{1/2}/n\mu$ as

$$\frac{1}{\mu}\left(\frac{E\sum\mathbf{w}_i^2}{n^2} + \frac{n(n-1)}{n^2}\mu^2\right)^{1/2}$$

Since $E\sum\mathbf{w}_i^2/n^2\rightarrow 0$ and $n(n-1)\mu^2/n^2\rightarrow\mu^2$, we have that $(E\mathbf{w}_{sum}^2)^{1/2}/n\mu\rightarrow 1$. The above two observations establish the desired result.

2.3. Average job completion time.

The last model studied in this chapter is similar to the model studied in Section 2.1.1. The difference occurs at the second stage. Given a realization of the processing times of the jobs at the second stage of the model described here, we have to decide on a schedule of the jobs on the machines acquired, so as to achieve a minimum value of the *average job completion time*. This difference is important because, contrary to the minimization of the maximum job completion time, the scheduling problem in this model is not NP-hard but solvable in polynomial time by a simple priority rule.

The first stages of both models are identical: a number X of identical machines that are to be acquired has to be decided on, given the cost c of a machine and the distribution of the job processing times. We again have $\mathfrak{X}=\mathbb{N}$. Let $F(w_1)$ be the common distribution function of $\mathbf{w}_1,\ldots,\mathbf{w}_n$ and assume that it is continuous, with finite expectation μ. Let $y^*(X,w)$ denote the optimal value of the second stage scheduling problems. Let us again assume that the cost per time unit is 1. The total cost will be denoted by $z^*(X,w) = cX + y^*(X,w)$.

The objective for the two-stage decision problem is to determine a value $X^*\in\mathfrak{X}$ such that

$$Ez^*(X^*,\mathbf{w}) = \min_{X\in\mathfrak{X}}\{Ez^*(X,\mathbf{w})\}.$$

The objective of the distribution problem is to find a function $X^\circ:\mathbb{R}^n\rightarrow\mathfrak{X}$ that yields for each realization w of $\mathbf{w}$ machines $X^\circ(w)$ for which

$$z^*(X^\circ(w),w) = \min_{X\in\mathfrak{X}}\{z^*(X,w)\}, \forall w\in\mathfrak{W}.$$

As mentioned above, this scheduling problem has the feature that the second level problem can be solved in polynomial time for each realization of the processing times of the jobs. An optimal schedule can be constructed by assigning the jobs in order of increasing processing times to the first available

machine [Conway et al. 1967]. Let $w^{(1)} \leqslant w^{(2)} \leqslant \cdots \leqslant w^{(n)}$ be the *order statistics* of $w_1, \ldots, w_n$. The optimality of the above *shortest processing time first rule* (SPT-rule) implies that

$$y^\star(X,w) = \frac{1}{n}\sum_{j=1}^n \lfloor \frac{n-j+1}{X} \rfloor w^{(j)}.$$

The analysis of the expected value of this term as a function of X is not a trivial task. To find a suitable value of X at the aggregate level, we will still have to rely on a heuristic approach. As in the previous models, this first stage heuristic will again be based on a lower bound on the second level objective that is asymptotically accurate. Obvious lower and upper bounds on $y^\star(X,w)$ are given by

$$\frac{1}{n}\sum_{j=1}^n \frac{n-j+1}{X} w^{(j)} \leqslant y^\star(X,w) \leqslant \frac{1}{n}\sum_{j=1}^n \frac{n-j+X+1}{X} w^{(j)}. \quad (2.12)$$

As in the previous models, we will use the expected value of the lower bound as an approximation of the expectation of the optimal second stage cost. To calculate the bound, we first rewrite the lower bound as

$$\frac{1}{X}\sum_{j=1}^n w_j - \frac{1}{nX}\sum_{j=1}^n (j-1)w^{(j)}. \quad (2.13)$$

The expected value of the first term in (2.13) is equal to $n\mu/X$. The expected value of the second term is calculated as follows. Let $F(x)$ be the common distribution function of the processing times.

$$\sum_{j=1}^n (j-1)E\,\mathbf{w}^{(j)}$$

$$= n\int_0^\infty \sum_{j=1}^n (j-1)\binom{n-1}{j-1}F(x)^{j-1}(1-F(x))^{n-j}x\,dF(x)$$

$$= n(n-1)\int_0^\infty \sum_{k=0}^{n-2}\binom{n-2}{k}F(x)^k(1-F(x))^{n-2-k}xF(x)\,dF(x)$$

$$= n(n-1)\int_0^\infty xF(x)\,dF(x).$$

The resulting lower bound on $Ey^\star(X,\mathbf{w})$ is given by

$$y^{H_1}(X) = \frac{1}{X}(n\mu-(n-1)\int_0^\infty xF(x)\,dF(x)).$$

Minimization of the lower bound

$$z^{H_1}(X) = cX + y^{H_1}(X)$$

yields the heuristic choice X^{H_1} for X at the aggregate level: X^{H_1} is the most favorable round-off of $((n\mu-(n-1)\nu)/c)^{1/2}$ with $\nu = \int_0^\infty xF(x)\,dF(x)$. We observe that ν can be readily calculated for some special cases of practical

importance. For example, if the processing times are uniformly distributed on an interval $[a,b]$, then $\nu=(b^3-a^3)/(3(b-a)^2)$, and if they come from a negative exponential distribution with parameter λ, then $\nu=3/(4\lambda)$.

At the detailed level, we schedule the jobs on the X^{H_1} machines acquired using the SPT-rule. The overall heuristic solution value is given by

$$z^*(X^{H_1},w) = cX^{H_1}+y^*(X^{H_1},w).$$

As a preliminary for the analysis of the performance of the heuristic we will analyze the asymptotic behavior of the bounds in (2.12). We rewrite these inequalities for each $w \in \mathcal{W}$ as

$$\frac{1}{n^2X}\sum_{j=1}^n w_j + \frac{1}{nX}\sum_{j=1}^n (1-\frac{j}{n})w^{(j)} \leqslant \frac{1}{n}y^*(X,w) \tag{2.14}$$

$$\leqslant \frac{X+1}{n^2X}\sum_{j=1}^n w_j + \frac{1}{nX}\sum_{j=1}^n (1-\frac{j}{n})w^{(j)}.$$

We observe that

$$T_n = \frac{1}{n}\sum_{j=1}^n (1-\frac{j}{n})w^{(j)}$$

is an example of a so-called *L-statistic,* a linear combination of order statistics, which has the general form

$$\frac{1}{n}\sum_{j=1}^n J(\frac{j}{n})w^{(j)},$$

where in our case $J(t)=1-t$.

We establish the following almost sure convergence result for such statistics. The following theorem was proved in a more general form in [Serfling 1980]. The proof in our special case is however straightforward enough to merit presentation on its own.

THEOREM 2.11. *If $J:[0,1]\rightarrow\mathbb{R}$ is a continuous function, then*

$$\frac{1}{n}\sum_{j=1}^n J(\frac{j}{n})\mathbf{w}^{(j)}\rightarrow \int_0^\infty xJ(F(x))dF(x) \quad (\text{wp1}).$$

PROOF. If we define the empirical distribution function by

$$\mathbf{F}_n(x) = \frac{1}{n}|\{j \mid \mathbf{w}_j \leqslant x\}|,$$

then

$$\mathbf{U}_n = \frac{1}{n}\sum_{j=1}^n J(\frac{j}{n})\mathbf{w}^{(j)} = \int_0^\infty xJ(\mathbf{F}_n(x))d\mathbf{F}_n(x). \tag{2.15}$$

We consider the inverse function $F^{-1}(y) = \inf_x\{x \mid F(x)>y\}$ of $F(x)$ and observe that

$$\mathbf{F}_n(F^{-1}(y)) = \frac{1}{n} | \{ j \, | \, F(\mathbf{w}_j) \leqslant y \} |.$$

However, $\mathbf{v}_j = F(\mathbf{w}_j)$ is uniformly distributed on [0,1] [Feller 1968] and hence

$$\mathbf{F}_n(F^{-1}(y)) = \mathbf{V}_n(y),$$

where $\mathbf{V}_n(y)$ is the empirical distribution function of n uniformly, independently distributed random variables. Thus, if we substitute $x = F^{-1}(y)$ in (2.15), we obtain

$$\mathbf{U}_n = \int\limits_0^1 F^{-1}(y) J(\mathbf{V}_n(y)) d\mathbf{V}_n(y) \tag{2.16}$$

$$= \int\limits_0^1 F^{-1}(y)(J(\mathbf{V}_n(y)) - J(y)) d\mathbf{V}_n(y) + \int\limits_0^1 F^{-1}(y) J(y) d\mathbf{V}_n(y).$$

Since $J(t)$ is continuous on [0,1] and hence uniformly continuous, we may use the fact that

$$\lim_{n \to \infty} \sup_{y \in [0,1]} | \mathbf{V}_n(y) - y | = 0 \quad \text{(wp1)}$$

(the Glivenko-Cantelli Lemma [Billingsley 1979, p. 232]) to conclude that, for any $\epsilon > 0$, with probability 1,

$$\limsup_{n \to \infty} \left| \frac{\int_0^1 F^{-1}(y) J(\mathbf{V}_n(y)) d\mathbf{V}_n(y) - \int_0^1 F^{-1} J(y) d\mathbf{V}_n(y)}{\int_0^1 F^{-1}(y) d\mathbf{V}_n(y)} \right| \leqslant \epsilon.$$

$$\tag{2.17}$$

Because of the strong law of large numbers, the denominator in (2.17) converges to μ (wp1), and hence

$$\limsup_{n \to \infty} \left| \int\limits_0^1 F^{-1}(y)(J(\mathbf{V}_n(y)) - J(y)) d\mathbf{V}_n(y) \right| = 0 \quad \text{(wp1)}. \tag{2.18}$$

We again invoke the strong law of large numbers to analyze the second term in the right hand side of (2.16)

$$\lim_{n \to \infty} \int\limits_0^1 F^{-1}(y) J(y) d\mathbf{V}_n(y) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n F^{-1}(\mathbf{v}_i) J(\mathbf{v}_i) \tag{2.19}$$

$$= E(F^{-1}(\mathbf{v}_i) J(\mathbf{v}_i))$$

$$= \int\limits_0^1 F^{-1}(y) J(y) dy \quad \text{(wp1)}.$$

Together (2.18) and (2.19) imply the theorem. $\square$

As a special case, we obtain

$$\mathbf{T}_n \to \mu - \nu \quad \text{(wp1)}. \tag{2.20}$$

We will establish asymptotic clairvoyancy of the heuristic by verifying the conditions of Theorem 1.4. Condition A is not satisfied in this case because $z^{H_1}(X)$ is not an almost sure lower bound on the optimal value of the distribution problem. Therefore we compare $z^{H_1}(X^{H_1})$ to a value that is such a lower bound. If we define, for each $X \in \mathbb{N}$ and for given w,

$$y^{LB}(X,w) = \frac{n}{X} T_n \tag{2.21}$$

and

$$z^{LB}(X,w) = cX + y^{LB}(X,w), \tag{2.22}$$

then we can derive from (2.12) that,

$$z^*(X^\circ(w),w) \geqslant \min_{X \in \mathbb{N}} \{ z^{LB}(X,w) \}. \tag{2.23}$$

The value $X^{LB}(w)$ that minimizes $z^{LB}(X,w)$ is given by $X^{LB}(w) = \sqrt{nT_n/c}$, where $z^{LB}(X^{LB}(w),w) = 2\sqrt{cnT_n}$. We have that

$$\frac{z^{H_1}(X^{H_1})}{z^{LB}(X^{LB}(\mathbf{w}),\mathbf{w})} = \frac{2\sqrt{cn(\mu - \frac{n-1}{n}\mu)}}{2\sqrt{cn\mathbf{T}_n}}.$$

Application of (2.20) to this ratio yields

$$\frac{z^{H_1}(X^{H_1})}{z^{LB}(X^{LB}(\mathbf{w}),\mathbf{w})} \to 1 \quad \text{(wp1)}. \tag{2.24}$$

We can again characterize the value of an optimal solution of the second stage problem asymptotically. We will see that it is asymptotically equal to $y^{H_1}(X)$.

THEOREM 2.12. *If $X = o(n)$, then*

$$\frac{y^*(X,\mathbf{w})}{n(\mu - \nu)/X} \to 1 \quad \text{(wp1)}.$$

PROOF. Using (2.12) we can write

$$\frac{n}{X} T_n \leqslant y^*(X,w) \leqslant \frac{n}{X} T_n + \frac{1}{n} \frac{X+1}{X} w_{\text{sum}}. \tag{2.25}$$

Division of (2.25) by $\frac{n(\mu - \nu)}{X}$ yields

$$\frac{T_n}{\mu - \nu} \leqslant \frac{y^*(X,w)}{n(\mu - \nu)/X} \leqslant \frac{T_n}{\mu - \nu} + \frac{w_{\text{sum}}(X+1)}{n^2(\mu - \nu)}. \tag{2.26}$$

As $X = o(n)$, the strong law of large numbers implies that $w_{\text{sum}}(X+1)/(n^2(\mu - \nu)) \to 0$ (wp1). This together with (2.20) implies the theorem. $\square$

Obviously,

$$\frac{y^{H_1}(X)}{n(\mu-\nu)/X} = \frac{n(\mu-\frac{n-1}{n}\mu)/X}{n(\mu-\nu)/X} \to 1. \tag{2.27}$$

We have deterministically that

$$z^{LB}(X^{LB}(w),w) \leqslant z^*(X^\circ(w),w) \leqslant cX^{H_1}+y^*(X^{H_1},w).$$

Division by $z^{H_1}(X^{H_1})$ yields

$$\frac{z^{LB}(X^{LB}(\mathbf{w}),\mathbf{w})}{z^{H_1}(X^{H_1})} \leqslant \frac{z^*(X^\circ(\mathbf{w}),\mathbf{w})}{z^{H_1}(X^{H_1})} \leqslant \frac{cX^{H_1}+y^*(X^{H_1},\mathbf{w})}{cX^{H_1}+y^{H_1}(X^{H_1})}.$$

As $X^{H_1} \in \{ \lfloor((n\mu-(n-1)\nu)/c)^1/2\rfloor, \lceil((n\mu-(n-1)\nu)/c)^1/2\rceil \} \cap \mathbb{N}$, Theorem 2.12 together with (2.27) implies that the latter term of the above inequalities tends to 1 with probability 1. This observation together with (2.24) implies asymptotic clairvoyance of the hierarchical system. From the definition of $z^{LB}(X^{LB}(\mathbf{w}),\mathbf{w})$ it is easy to prove (cf. Theorem 1.4 (ii)) that for every $\epsilon>0$,

$$Pr\{\lim_{n\to\infty}\frac{X^{LB}(\mathbf{w})}{X^\circ(\mathbf{w})} < \frac{1}{1+\epsilon}\} = 0$$

and

$$Pr\{\lim_{n\to\infty}\frac{X^{LB}(\mathbf{w})}{X^\circ(\mathbf{w})} > \frac{1}{1-\epsilon}\} = 0.$$

This together with the fact that

$$Pr\{\lim_{n\to\infty}\frac{X^{H_1}}{X^{LB}(\mathbf{w})}=1\} = 1$$

implies the second part of the following theorem.

THEOREM 2.13.
(i) $z^*(X^{H_1},\mathbf{w})/z^*(X^\circ(\mathbf{w}),\mathbf{w})\to 1$ (wp1);
(ii) $X^{H_1}/X^\circ(\mathbf{w})\to 1$ (wp1). $\square$

If the second moment $E\mathbf{w}_1^2$ is finite we can even establish the rate at which $z^*(X^{H_1},\mathbf{w})/z^*(X^\circ(\mathbf{w}),\mathbf{w})$ converges to 1, something that was not done in previous cases. For this purpose, we make use of a result that is established in the following theorem. As in the case of Theorem 2.11 a generalization of this theorem was proved in [Serfling 1980].

THEOREM 2.14. *If $J:[0,1]\to\mathbb{R}$ is a continuously differentiable function, then with probability 1*

$$\limsup_{n\to\infty}\frac{\sqrt{n}}{\sqrt{\log\log n}}\left|\frac{1}{n}\sum_{j=1}^n J(\frac{j}{n})\mathbf{w}^{(j)} - \int_0^\infty xJ(F(x))dF(x)\right| < \infty.$$

PROOF. In this proof we will use the notation that was introduced in the proof of Theorem 2.11:

$$\mathbf{U}_n - \int_0^1 F^{-1}(y)J(y)dy = \int_0^1 F^{-1}(y)(J(\mathbf{V}_n(y)) - J(y))d\mathbf{V}_n(y) \qquad (2.28)$$

$$+ \int_0^1 F^{-1}(y)J(y)d\mathbf{V}_n(y) - \int_0^1 F^{-1}(y)J(y)dy$$

and analyze the right hand side of (2.28) in parts.

Since $J(t)$ is continuously differentiable on $[0,1]$, we may apply the mean value theorem to conclude that there exists a $\theta \in (0,1)$ such that

$$J(\mathbf{V}_n(y)) - J(y) = J'(\mathbf{W}_n(y))(\mathbf{V}_n(y) - y) \qquad (2.29)$$

with

$$\mathbf{W}_n(y) = \Theta\mathbf{V}_n(y) + (1-\Theta)y.$$

Since $\mathbf{V}_n(y)$ is an increasing function and $F^{-1}(y) \geqslant 0$, we may conclude, after substitution of (2.29) in the first term of the right hand side of (2.28) that

$$\left| \int_0^1 F^{-1}(y)(J(\mathbf{V}_n(y)) - J(y))d\mathbf{V}_n(y) \right| \leqslant$$

$$\sup_{y \in [0,1]} |\mathbf{V}_n(y) - y| \int_0^1 F^{-1}(y)|J'(\mathbf{W}_n(y))|d\mathbf{V}_n(y).$$

Now, since F is continuous [Chung 1949], we have

$$\limsup_{n \to \infty} \frac{\sqrt{n} \, \limsup_{n \to \infty} |\mathbf{V}_n(y) - y|}{\sqrt{2\log\log n}} = \tfrac{1}{2} \quad (\text{wp1}).$$

Furthermore, there exists a constant M such that

$$\int_0^1 F^{-1}(y)|J'(\mathbf{W}_n(y))|d\mathbf{V}_n(y) \leqslant M\int_0^1 F^{-1}(y)d\mathbf{V}_n(y),$$

because $J'(y)$ is continuous on $[0,1]$. Now,

$$\lim_{n \to \infty} \frac{1}{\mu} \int_0^1 F^{-1}(y)d\mathbf{V}_n(y) = 1.$$

Hence, with probability 1,

$$\limsup_{n \to \infty} \frac{\sqrt{n}}{\sqrt{\log\log n}} \left| \int_0^1 F^{-1}(y)J(\mathbf{V}_n(y))d\mathbf{V}_n(y) - \int_0^1 F^{-1}(y)J(y)d\mathbf{V}_n(y) \right| < \infty.$$

The last two terms of the right hand side of (2.28) can be rewritten as

$$\frac{1}{n}\sum_{j=1}^n F^{-1}(\mathbf{v}_i)J(\mathbf{v}_i) - \int_0^1 F^{-1}(y)J(y)dy$$

48

If $E\mathbf{w}_1^2 < \infty$, we may apply the law of the iterated logarithm [Hartman &

$$\limsup_{n\to\infty} \frac{\sqrt{n}}{\sqrt{\log\log n}} \left| \int_0^1 F^{-1}(y)J(y)d\mathbf{V}_n(y) - \int_0^1 F^{-1}(y)J(y)dy \right| < \infty.$$

(2.31)

Together (2.30) and (2.31) imply the theorem. $\square$

We use Theorem 2.14 to analyze the convergence rate of $z^*(X^{H_1},\mathbf{w})/z^*(X^\circ(\mathbf{w}),\mathbf{w})$. From (2.29) and (2.31) we obtain

$$1 \leq \frac{z^*(X^{H_1},\mathbf{w})}{z^*(X^\circ(w),\mathbf{w})} \leq \frac{cX^{H_1} + \dfrac{n\mathbf{T}_n}{x^{H_1}} + \dfrac{1}{n}\dfrac{X^{H_1}+1}{X^{H_1}}\mathbf{w}_{\text{sum}}}{2\sqrt{cn\mathbf{T}_n}}.$$

(2.32)

From the definition of X^{H_1} we have

$$\frac{cX^{H_1}}{(cn\mathbf{T}_n)^{1/2}} - 1 = \frac{(c(n\mu-(n-1)\nu))^{1/2} - (cn\mathbf{T}_n)^{1/2}}{(cn\mathbf{T}_n)^{1/2}}$$

$$= \frac{\dfrac{1}{n}\mu + (\mu-\nu) - \mathbf{T}_n}{\mathbf{T}_n^{1/2}((\mu - \dfrac{(n-1)}{n}\nu)^{1/2} + \mathbf{T}_n^{1/2})}.$$

Hence, Theorem 2.14 applied once again to the special case that $J(t)=1-t$ yields that

$$\limsup_{n\to\infty} \left| \frac{cX^{H_1}}{(cn\mathbf{T}_n)^{1/2}} - 1 \right| \left(\frac{n}{\log\log n}\right)^{1/2} < \infty \quad (\text{wp1}).$$

(2.33)

Furthermore,

$$\lim_{n\to\infty} \frac{\dfrac{n\mathbf{T}_n}{X^{H_1}}}{(cn\mathbf{T}_n)^{1/2}} = \left(\frac{c(n\mu-(n-1)\nu)}{cn(\mu-\nu)}\right)^{1/2} = 1 \quad (\text{wp1}).$$

(2.34)

We observe that

$$\frac{c}{(cn\mathbf{T}_n)^{1/2}} = \frac{(cn\mathbf{T}_n)^{1/2}}{\dfrac{n\mathbf{T}_n}{X^{H_1}}}.$$

(2.35)

Together, (2.33), (2.34) and (2.35) imply that with probability 1

$$\limsup_{n\to\infty} \left| \frac{\dfrac{n\mathbf{T}_n}{X^{H_1}}}{(cn\mathbf{T}_n)^{1/2}} - 1 \right| \left(\frac{n}{\log\log n}\right)^{1/2}$$

(2.36)

$$= \limsup_{n \to \infty} \left| \frac{n\mathbf{T}_n}{(cn\mathbf{T}_n)^{1/2}} \right| \left| 1 - \frac{cX^{H_1}}{(cn\mathbf{T}_n)^{1/2}} \right| (\frac{n}{\log\log n})^{1/2} < \infty.$$

Finally

$$\limsup_{n \to \infty} \frac{\frac{2}{n}\mathbf{w}_{\text{sum}}}{(cn\mathbf{T}_n)^{1/2}}(\frac{n}{\log\log n})^{1/2} = 0 \quad (\text{wp1}). \tag{2.37}$$

We apply (2.33), (2.36) and (2.37) to (2.32) to arrive at the following strong extension of Theorem 2.13.

THEOREM 2.15.

$$\limsup_{n \to \infty} \left| \frac{z^*(X^{H_1},\mathbf{w})}{z^*(X^\circ(\mathbf{w}),\mathbf{w})} - 1 \right| (\frac{n}{\log_2\log_2 n})^{1/2} < \infty \quad (\text{wp1}). \quad \square$$

We finally establish the rate of convergence of the heuristic solution at the aggregate level to the optimal one in the following theorem.

THEOREM 2.16.

$$\limsup_{n \to \infty} \left| \frac{X^\circ(\mathbf{w})}{X^{H_1}} - 1 \right| n^{1/4} < \infty \quad (\text{wp1}).$$

PROOF. Recall $z^{LB}(X^{LB}(w),w)$. We have

$$z^{LB}(X^{LB}(\mathbf{w}),\mathbf{w}) \leqslant z^*(X^\circ(\mathbf{w}),\mathbf{w}) \leqslant z^{LB}(X^{LB}(\mathbf{w}),\mathbf{w}) + \frac{2}{n}\sum_{j=1}^n \mathbf{w}_j. \tag{2.38}$$

We now compute $X_1(\mathbf{w})$ and $X_2(\mathbf{w})$ such that

$$z^{LB}(X_1(\mathbf{w}),\mathbf{w}) = z^{LB}(X_2(\mathbf{w}),\mathbf{w}) = z^{LB}(X^{LB}(\mathbf{w}),\mathbf{w}) + \frac{2}{n}\sum_{j=1}^n \mathbf{w}_j.$$

To do so, we solve the equality

$$cX + \frac{n\mathbf{T}_n}{X} = 2(cn\mathbf{T}_n)^{1/2} + \frac{2}{n}\sum_{j=1}^n \mathbf{w}_j$$

rewritten as

$$cX^2 - (2(cn\mathbf{T}_n)^{1/2} + \frac{2}{n}\sum_{j=1}^n \mathbf{w}_j)X + n\mathbf{T}_n = 0,$$

to find two roots

$$X_1(\mathbf{w}) = (\frac{n\mathbf{T}_n}{c})^{1/2} + \frac{1}{cn}\sum_{j=1}^n \mathbf{w}_j - (\frac{2}{cn}\sum_{j=1}^n \mathbf{w}_j(cn\mathbf{T}_n)^{1/2} + (\frac{1}{cn}\sum_{j=1}^n \mathbf{w}_j)^2)^{1/2}$$

and

$$X_2(\mathbf{w}) = (\frac{n\mathbf{T}_n}{c})^{1/2} + \frac{1}{cn}\sum_{j=1}^n \mathbf{w}_j + (\frac{2}{cn}\sum_{j=1}^n \mathbf{w}_j(cn\mathbf{T}_n)^{1/2} + (\frac{1}{cn}\sum_{j=1}^n \mathbf{w}_j)^2)^{1/2}.$$

The definitions of $X_1(\mathbf{w})$, $X_2(\mathbf{w})$ and $X^\circ(\mathbf{w})$ imply that $X_1(\mathbf{w}) \leqslant X^\circ(\mathbf{w}) \leqslant X_2(\mathbf{w})$ and hence

$$\frac{X_1(\mathbf{w})}{X^{H_1}} \leqslant \frac{X^\circ(\mathbf{w})}{X^{H_1}} \leqslant \frac{X_1(\mathbf{w})}{X^{H_1}} + \frac{X_2(\mathbf{w}) - X_1(\mathbf{w})}{X^{H_1}}. \tag{2.39}$$

Now

$$\frac{(cn\,\mathbf{T}_n)^{1/2}}{cX^{H_1}} + \frac{(\frac{2}{cn}\sum_{j=1}^n \mathbf{w}_j(cn\,\mathbf{T}_n)^{1/2})^{1/2}}{X^{H_1}} \leqslant \frac{X_1(\mathbf{w})}{X^{H_1}} \tag{2.40}$$

$$\leqslant \frac{(cn\,\mathbf{T}_n)^{1/2}}{cX^{H_1}} + \frac{(\frac{1}{cn}\sum_{j=1}^n \mathbf{w}_j)}{X^{H_1}}$$

and

$$\frac{X_2(\mathbf{w}) - X_1(\mathbf{w})}{X^{H_1}} \leqslant \frac{2}{X^{H_1}}(\frac{2}{cn}\sum_{j=1}^n \mathbf{w}_j(cn\,\mathbf{T}_n)^{1/2})^{1/2} + \frac{2}{X^{H_1}}(\frac{1}{cn}\sum_{j=1}^n \mathbf{w}_j).$$

$$\tag{2.41}$$

As

$$\lim_{n\to\infty}(\frac{2}{cn}\sum_{j=1}^n \mathbf{w}_j)^{1/2}\frac{(cn\,\mathbf{T}_n)^{1/4}}{n^{1/4}} < \infty \quad (\text{wp1}),$$

(2.39), (2.40), (2.41) together with (2.34) and the strong law of large numbers imply the theorem. $\square$

For this problem we can also establish a worst case bound on the relative error the heuristic produces with respect to the optimal solution of the two-stage decision problem.

THEOREM 2.17.

$$\frac{Ez^*(X^{H_1},\mathbf{w})}{Ez^*(X^*,\mathbf{w})} \leqslant 1 + (\frac{X^{H_1}+1}{X^{H_1}})\frac{\frac{1}{n}E\mathbf{w}_{\text{sum}}}{2\sqrt{c(n\mu-(n-1)\nu)}}.$$

PROOF. Using the bounds in (2.12) we have

$$\frac{Ez^*(X^{H_1},\mathbf{w})}{Ez^*(X^*,\mathbf{w})} \leqslant \frac{cX^{H_1} + \frac{n}{X^{H_1}}\mathbf{T}_n + \frac{1}{n}\frac{X^{H_1}+1}{X^{H_1}}E\mathbf{w}_{\text{sum}}}{z^{H_1}(X^{H_1})}.$$

The observation that $cX^{H_1} + \frac{n}{X^{H_1}}E\mathbf{T}_n = z^{H_1}(X^{H_1})$ and $z^{H_1}(X^{H_1}) \geqslant 2\sqrt{c(n\mu-(n-1)\nu)}$ yields the theorem. $\square$

Asymptotic expectation-optimality of the heuristic follows easily from the

definition of X^{H_1}.

3

Hierarchical vehicle routing and
location problems

Vehicle routing problems, in which *customers* have to be served from a central *depot* by one or more *vehicles,* have acquired the reputation of being notoriously difficult on one hand and of being of great practical importance on the other hand. These problems are usually formulated under the assumption that perfect information about the customers is available. In actual practice, this assumption is not always justified. In particular, the medium or long term planning problem of acquiring a suitable fleet of vehicles has to be solved with vague and at best probabilistic information about what will ultimately be required of them.

Hierarchical routing problems, involving the trade-off between the cost of acquiring vehicles now and the expected benefits of having them available at a later stage, form the first class of problems studied in this chapter.

Proceeding along the lines of Chapter 1, we will present hierarchical systems for some special cases of the vehicle routing problem and analyze their performance. In the problems under consideration we assume that the costs directly related to a first stage decision are known with certainty, whereas only stochastic information is available about some of the customer characteristics. Thus, we will assume that the locations of the customers are independent identically distributed random variables with known probability distributions.

In Section 3.1 the aggregate problem is to determine the number and types of vehicles to be acquired, while at the detailed level, given a realization of the customer characteristics, we have to route the vehicles available through the customers so as to minimize the length of the longest route assigned to any of them. This detailed level criterion is convenient since it works in favor of a

reasonable division of labor among the vehicles.

Location problems, that are the subject of Section 3.2, deal with the location of depots from which customers must be served. The hierarchical location problem that we study in this chapter asks for a decision of the number of depots to be established at the aggregate level, when only stochastic information is available about the locations of customers that are to be served from these depots. At the detailed level, given a realization of the customer locations, the depots must be located so as to minimize the sum of the distances of the customers to the depots, where the distance of a customer to the depots is defined as the Euclidean distance of that customer to the nearest depot. The set of possible locations for the depots will be restricted to the customers locations. The detailed level problem is known as as a median-location problem. It is NP-hard in itself.

The hierarchical location and routing problem that we study in Section 3.3 can be viewed as a generalization of the hierarchical location problem above. Whereas in that problem each customer is served individually, here customers are served in groups by vehicles, that have a capacity of serving only a limited number of customers. The aggregate level decision still concerns the number of depots to be established, under the same stochastic assumptions as in the hierarchical location problem. At the detailed level the depots are to be located and vehicles are to be routed from the depots through the customers so as to minimize the total distance to be traveled. Also, in this case the detailed level problem is NP-hard.

Each of the above problems is defined in the plane. The locations of the customers are independent and have common uniform distribution on a region that is for each of the problems defined in the respective section. The vector that represents the random locations of the n customers will be denoted by w. Let the set of all possible realizations of w be denoted by $\mathcal{W}$.

3.1. Hierarchical vehicle routing

In this section we study variants of the following stochastic programming model. At the first stage one has to acquire a fleet X of vehicles with unlimited capacity, to be selected from a set $\mathcal{M}$ of available vehicles. Let $\mathcal{X} = 2^{\mathcal{M}}$ be the power set of $\mathcal{M}$. The direct acquisition cost is given as a function $f : \mathcal{X} \to \mathbb{R}$ of X. The vehicles are used to serve n customers from a single depot. We assume that the customers are located randomly in a circular area R with radius r, with the depot at its center. The distribution is uniform.

Subsequently, at the second stage, the vehicles that have been acquired have to be routed from the depot through the n customers, a realization of whose locations is now given. For each $X \in \mathcal{X}$ and each realization w of $\mathbf{w}$ let $v_i^*(X, w)$ be the length of the route assigned to the ith vehicle in an optimal solution of the second stage problem. We assume that the detailed level cost is proportional to the length of the longest route. Hence the second stage optimal cost is defined by

$$y^*(X,w) = \max_{i \in X} \{v_i^*(X,w)\}.$$

The sum of the first stage cost $f(X)$ and the cost of an optimal second stage decision is denoted by $z^*(X,w) = f(X)+y^*(X,w)$. The two-stage decision objective is to find a set of vehicles $X^* \subseteq \mathcal{X}$ such that

$$Ez^*(X^*,\mathbf{w}) = \min_{X \subseteq \mathcal{X}} \{Ez^*(X,\mathbf{w})\}.$$

The distribution problem is to find a function $X^\circ: \mathcal{W} \to \mathcal{X}$ that yields for each realization w of $\mathbf{w}$ a set of vehicles $X^\circ(w)$ for which

$$z^*(X(w),w) = \min_{X \subseteq \mathcal{X}} \{z^*(X,w)\}, \ \forall w \in \mathcal{W}.$$

Determining $Ez^*(X,\mathbf{w})$ as a function of X requires integration of the function $y^*(X,w)$, a single evaluation of which requires the solution of an NP-hard problem [Garey & Johnson 1979].

The section is divided into three subsections. In Subsection 3.1.1 we assume that the available vehicles are *identical* in cost and speed, whereas in Subsection 3.1.2 we assume that they are *uniform,* i.e., each vehicle has its own cost and speed. In Subsection 3.1.3 we study two variants of the problem in which in addition to stochasticity in the location of the customers, there is uncertainty about the number of customers to be served at the detailed level.

3.1.1. Identical vehicles. In this subsection we assume that the available vehicles are identical. Each of them has a cost c, so that at the aggregate level one has to decide on a *number* X of vehicles to be acquired. In this case we thus have $\mathcal{X}=\mathbb{N}$ and $f(X)=cX$.

This is a special case of the simple model outlined in Section 1.5. Along the lines given there we will design a hierarchical system to solve the two-stage decision problem and show that the system is asymptotically clairvoyant with probability 1 by verification of Conditions A and B of Theorem 1.4.

A lower bound on $y^*(X,\mathbf{w})$ can be derived as follows. For each realization w of the customer locations, let $t^*(w)$ be the length of an optimal *traveling salesman tour* through all customers, i.e. a tour that visits each customer exactly once. Since the sum of the lengths of the optimal routes of X vehicles, $\sum_{i=1}^{X} v_i^*(X,w)$ is greater than or equal to $t^*(w)$, we have

$$y^*(X,w) = \max_{i=1,\dots,X} \{v_i^*(X,w)\} \geqslant \frac{1}{X}\sum_{i=1}^{X} v_i^*(X,w) \tag{3.1}$$

$$\geqslant \frac{1}{X} t^*(w).$$

To estimate $t^*(\mathbf{w})/X$ we apply a theorem due to Steele [Steele 1981] which extends earlier work by Beardwood et al. [Beardwood et al. 1959], and which characterizes asymptotically the optimal value of a traveling salesman tour.

THEOREM 3.1. *If n customers are distributed uniformly over a compact region R with area $v(R)$, then there exists a constant $\beta>0$ such that*

$$\frac{t^*(w)}{\sqrt{n\,\nu(R)}} \to \beta \quad (\text{wp1}). \quad \square \tag{3.2}$$

In our problem R is a circular region with radius r, so that $\nu(R)=\pi r^2$. Substitution of this in (3.2) yields, in view of (3.1), that the function

$$z^{H_1}(X) = cX + \frac{\beta}{X}\sqrt{n\,\pi r^2} \tag{3.3}$$

is asymptotically an almost sure lower bound on $z^*(X,w)$. As a heuristic decision at the aggregate level, we now choose the number of vehicles equal to the integer value X^{H_1} that minimizes $z^{H_1}(X)$. Since $z^{H_1}(X)$ is a convex function of X, the first order condition, given by

$$c - \frac{\beta}{X^2}\sqrt{n\,\pi r^2} = 0$$

implies that X^{H_1} is equal to the most favorable round off of

$$\left[\frac{\beta\sqrt{\pi r^2}}{c}\right]^{\frac{1}{2}} n^{\frac{1}{4}}.$$

At the detailed level we have to route X^{H_1} vehicles through n customers with locations given by w, so as to minimize the maximum length of any route assigned to a vehicle. We propose to solve this problem by means of a *partitioning heuristic* that is similar to Karp's heuristic [Karp 1977] for the Euclidean traveling salesman problem. Roughly, the heuristic consists of three steps. At first, R is partitioned into smaller subregions, each of which contains no more than d customers for some parameter d that is yet to be determined. In the second step, an optimal traveling salesman tour is constructed through the customers in each of these subregions. In the third and final step, the tours are combined in a suitable manner to form the routes for the vehicles.

The partitioning of the circular area in the first step is carried out by means of *cuts,* of which we distinguish two types. We represent the location of each customer by its polar coordinates. A *radial cut* of a region splits up the region by means of the radius through the customer in the region with median angular coordinate (see Figure 3.1a). Similarly, a *circular cut* splits up a region by means of the circle arc (with the depot as center) through the customer in the region with median radial coordinate (see Figure 3.1b).

In a *round* of cutting, each subregion existing at the beginning of the round is split up exactly once. We carry out K of these rounds, with

$$K = \lceil \log_2 \frac{(n-1)}{(d-1)} \rceil.$$

We start our cutting procedure from the circular area with one radius arbitrarily fixed. The first $K/2$ rounds involve only radial cuts, thus creating $2^{K/2}$ *sectors;* the last $K/2$ rounds involve only circular cuts. We number the 2^K subregions obtained by starting with an arbitrary sector, numbering the
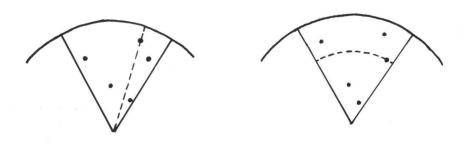
FIGURE 3.1a.
Radial cut

FIGURE 3.1b.
Circular cut

subregions according to increasing distance from the depot, and continuity on the adjacent sector in, say, clockwise direction until all subregions have been numbered. The jth region will be denoted by $R_j(w)$ ($j = 1,...,2^K$) (see Figure 3.2). It is easy to show that the above cutting procedure (which is simpler than the one proposed in [Karp 1977]) results in subregions containing no more than d customers each.
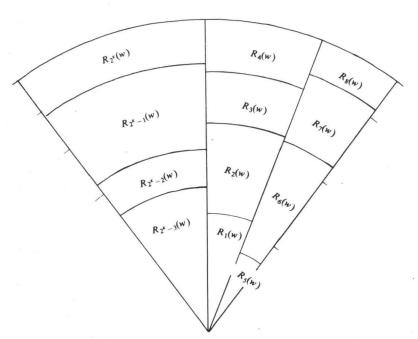
FIGURE 3.2. The subregions

LEMMA 3.1. $R_j(w)$ $(j = 1,...,2^K)$ *contains at most d customers.*

PROOF. By induction we will show that the number of customers in each of the subregions existing after k rounds of cutting is no more than

$$\frac{n}{2^k} + \frac{1}{2^{k-1}} + \cdots + \frac{1}{2} + 1 = \frac{n-2}{2^k} + 2.$$

After one round of cutting one of the subregions contains $n/2$ customers and the other $(n/2) - 1$ or $n/2$ depending on the parity of n. The customer on the splitting segment is added to both subregions so that for $k = 1$ the hypothesis is obviously satisfied. Suppose that it is also satisfied for $k - 1$ rounds. If in the kth round we split a region with $(n-2)/2^{k-1} + 2$ customers the number of customers in each of the two newly created subregions is at most

$$(\frac{n}{2^{k-1}} + \frac{1}{2^{k-2}} + \cdots + \frac{1}{2} + 1)\frac{1}{2} + 1 = \frac{n}{2^k} + \frac{1}{2^{k-1}} + \cdots + \frac{1}{2} + 1,$$

which verifies the induction hypothesis. Therefore, after K rounds of cutting the above upper bound is smaller than or equal to

$$(\frac{n-2}{n-1})(d-1) + 2,$$

which is no more than d if and only if $d \leqslant n$. $\square$

In the second step of the heuristic, an optimal traveling salesman tour of length $t^*(R_j(w),w)$ is formed through the customers (including those on the boundary) in each region $R_j(w)$ by means of a suitable optimization method. Let the graphical configuration corresponding to such a tour be indicated by $T^*(R_j(w),w)$. It is not difficult to see that the union of these tours defines an *Euler graph*, i.e., a connected graph on the set of all customers in which each customer has even degree. Euler's theorem implies that there exists a *spanning walk* that passes through each edge exactly once. The length of this spanning walk is $W(w) = \sum_{i=1}^{2^K} t^*(R_j(w),w)$.

In the final step we assign each customer to a specific vehicle, in such a way that the route for each vehicle is approximately equal to $W(w)/X$. We do so in the obvious manner, by considering $R_1(w), R_2(w), R_3(w),...$ until we find the greatest l such that

$$\delta = W(w)/X - \sum_{j=1}^{l} t^*(R_j(w),w) \geqslant 0.$$

If $\delta > 0$, we divide $T^*(R_{l+1}(w),w)$ into two parts. The customers on the part with length δ, together with the customers on the tours $T^*(R_1(w),w),...,$ $T^*(R_l(w),w)$ are assigned to the first vehicle. The customers on the other part are the first ones to be assigned to the second vehicle. We continue this procedure until each vehicle has a set of tours (including at most two partial tours) assigned to it whose lengths sum exactly to $W(w)/X$.

The union of these tours does not necessarily define a spanning walk. In general, it will not be connected and have the form depicted in Figure 3.3 in

heavy lines. As indicated in the figure at most eight additional dotted segments may be required to create a spanning walk.
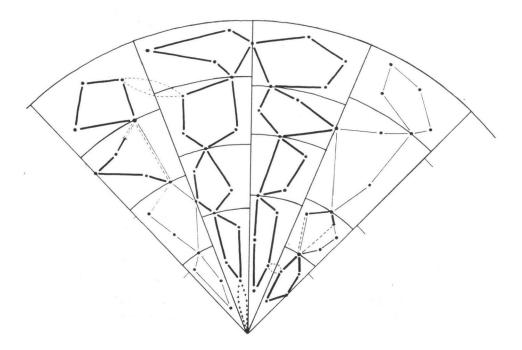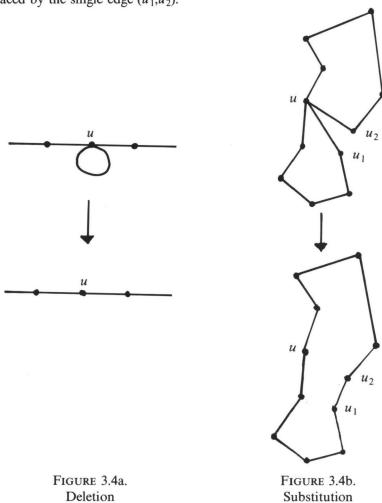
FIGURE 3.3. From a set of tours to a spanning walk

Two additional segments (indicated by $+ + +$) are needed to connect the depot to the customer that is closest to it. It is easy to see that the total length of the additional segments is a constant $\gamma(w)$ depending only on r. The resulting spanning walk can be transformed into a tour by subsequent application of two operations. The first one deletes a loop (see Figure 3.4a).

The second operation is applied whenever a node u has degree greater than two and has no loop (see Figure 3.4b). Then a pair of edges (u_1,u) and (u_2,u) exists that does not form a cut set of the spanning walk; i.e., if we remove this pair from the walk, the walk remains connected. These edges are removed and replaced by the single edge (u_1,u_2).

FIGURE 3.4a.
Deletion

FIGURE 3.4b.
Substitution

The triangle inequality ensures that the resulting tour has no greater length than the original spanning walk. The lengths of the resulting routes are denoted by $v_i^{H_2}(X,w)$ $(i=1,...,X)$. The longest of them has length denoted by $y^{H_2}(X,w)$, which is the value produced by the second stage heuristic, for each realization w of the customer locations. The value produced by the hierarchical system formed by the first stage and the second stage heuristics H_1 and H_2 for a given realization w of $\mathbf{w}$ is given by

$$z^{H_2}(X^{H_1}, w) = cX^{H_1} + y^{H_2}(X^{H_1}, w).$$

It is not difficult to see that, subject to the usual assumption that each elementary operation on real numbers requires unit time, the above second stage heuristic can be implemented so as to require a running time that is polynomial in the number of customers, provided that d is chosen to depend appropriately on n.

In the first step, all customers have to be sorted with respect to their angular as well as to their radial coordinates. In addition, each round of cutting takes linear time. Altogether, this step requires $O(n \log n)$ time.

The second step, calculation of an optimal traveling salesman tour in each subregion, can be carried out in $O(\theta^d)$ time per region for some constant $\theta > 2$ (e.g., by dynamic programming [Held & Karp 1962]) and hence in $O(n\theta^d/d)$ time overall.

In the third step, the assignment of each customer to a vehicle takes time that is linear in the number of subregions and in the number of vehicles. This includes the time needed to create the extra segments, which is proportional to $d^2 X^{H_1}$.

Since the first stage heuristic requires a constant amount of time and results in $X^{H_1} = O(n^{1/4})$, the overall running time of the hierarchical system is $O(n \log n + n\theta^d/d + d^2 n^{1/4})$. If we take d equal to $\log_2 n$, this running time is $O(n^2/\log n)$.

To prove asymptotic clairvoyancy of both the first stage heuristic H_1 and the hierarchical system (H_1, H_2) we verify Conditions A and B of Theorem 1.4 in Section 1.5. Verification of A is easy from (3.1) and Theorem 3.1. The error analysis of the detailed level partitioning heuristic, which leads to the verification of Condition B, is much harder.

As a preliminary we will prove two lemmas that together yield an upper bound on the length of the spanning walk constructed in the first two steps of the heuristic.

Consider a subregion $R_j(w)$ and let $T^*(w) \cap R_j(w)$ denote the intersection of the optimal tour through n customers with $R_j(w)$. Let $per(R)$ be the perimeter of the region R. The proof of the following lemma is from [Karp 1977].

LEMMA 3.2.

$$t^*(R_j(w), w) - t^*(w) \cap R_j(w) \leq \tfrac{3}{2} per(R_j(w)).$$

PROOF. Let $T^*(w) \cap R_j(w)$ consist of k continuous curves $C_1, C_2, \ldots, C_k$. (see Figure 3.5).

Let the $2k$ end points of these curves, in clockwise order around the boundary of $R_j(w)$, be $r_1, r_2, \ldots, r_{2k}$. Assume without loss of generality that $|r_1 r_2| + |r_3 r_4| + \cdots + |r_{2k-1} r_{2k}| \leq |r_2 r_3| + \cdots + |r_{2k} r_1|$, where $|r_{i_1} r_{i_2}|$ denotes the distance from r_{i_1} to r_{i_2} along the boundary of $R_j(w)$. Consider the walk consisting of the following three parts: the curves $C_1, \ldots, C_k$; two copies of each of the segments $r_1 r_2, r_3 r_4, \ldots, r_{2k-1} r_{2k}$; and one copy of each of the
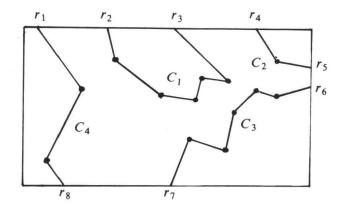
FIGURE 3.5. Converting $T \star (w) \cap R_j (w)$ into a walk

segments $r_2 r_3$, $r_4, r_5, ..., r_{2k} r_1$. Then the length of the first part is $t^\ast (w) \cap R_j (w)$, and the sum of the lengths of the second and third part is less than or equal to $\frac{3}{2}$ the perimeter of $R_j (w)$. As the length of an optimal traveling salesman tour through the customers in $R_j (w)$ is smaller than the length of this spanning walk the lemma follows. $\square$

From Lemma 3.2 it follows that

$$W(w) = \sum_{j=1}^{2^K} t^\ast (R_j(w), w) \leqslant \sum_{j=1}^{2^K} t^\ast (w) \cap R_j (w) + \frac{3}{2} \sum_{j=1}^{2^K} per(R_j(w))$$

$$= t^\ast (w) + \frac{3}{2} \sum_{j=1}^{2^K} per(R_j(w)).$$

Our cutting procedure, which is different from the one in [Karp 1977] is not only simpler to implement, but also leads to a more exact evaluation of $\sum_{j=1}^{2^K} per(R_j(w))$.

LEMMA 3.3.

$$\sum_{j=1}^{2^K} per(R_j(w)) = O(\sqrt{n/d}).$$

PROOF. After $K/2$ radial cuts, the sum of the perimeters of the sectors is clearly equal to

$$2^K/2(2r) + 2\pi r. \tag{3.4}$$

In the first round of circular cuts, all sectors are split by circle arcs, the sum of which is certainly smaller than $2\pi r$, so that (3.4) is increased by no more than $4\pi r$. In the second round, the increase is bounded in a similar manner by $8\pi r$. Hence, the overall increase is bounded by

$$(2^K/2 - 1)4\pi r. \tag{3.5}$$

Since $K = \lceil \log_2((n-1)/(d-1)) \rceil$, (3.4) and (3.5) together imply the lemma. $\square$

From the above two lemmas an upper bound can be derived on the route length for each vehicle, and hence, on $y^{H_2}(X,w)$:

$$y^{H_2}(X,w) \leqslant \frac{W(w)}{X} + \gamma(w) \leqslant \frac{t^*(w)}{X} + \frac{O(\sqrt{n/d})}{X}. \tag{3.6}$$

This upper bound is used in the following theorem, which establishes an asymptotic characterization of the optimal solution and the heuristic solution of the second stage vehicle routing problem and shows that they are asymptotic to the same form.

THEOREM 3.2.
(i) $y^*(X,\mathbf{w})/(\beta\sqrt{n\pi r^2}/X) \to 1$ (wp1);
(ii) $y^{H_2}(X,\mathbf{w})/(\beta\sqrt{n\pi r^2}/X \to 1$ (wp1).

PROOF. (3.1) and (3.6) together yield

$$\frac{t^*(w)}{X} \leqslant y^*(X,w) \leqslant y^{H_2}(X,w) \leqslant \frac{t^*(w)}{X} + \frac{O(\sqrt{n/d})}{X}. \tag{3.7}$$

Division by $\frac{\beta}{X}\sqrt{n\pi r^2}$ yields,

$$\frac{t^*(w)}{\beta\sqrt{n\pi r^2}} \leqslant \frac{y^*(X,w)}{\beta\sqrt{n\pi r^2}/X} \leqslant \frac{y^{H_2}(X,w)}{\beta\sqrt{n\pi r^2}/X} \leqslant \frac{t^*(w)}{\beta\sqrt{n\pi r^2}} + \frac{O(\sqrt{n/d})}{\beta\sqrt{n\pi r^2}}. \tag{3.8}$$

For $d = \log_2 n$, we have

$$\frac{O(\sqrt{n/d})}{\beta\sqrt{n\pi r^2}} \to 0.$$

This observation, Theorem 3.1 and (3.8) together imply the theorem. $\square$

Theorem 3.2 shows that Condition B of Theorem 1.4 is also satisfied so that asymptotic clairvoyancy with probability 1 of the heuristic H_1 and (H_1,H_2) is established.

THEOREM 3.3.
(i) $z^{H_2}(X^{H_1},\mathbf{w})/z^*(X^\circ(\mathbf{w}),\mathbf{w}) \to$ (wp1);
(ii) $X^{H_1}/X^\circ(\mathbf{w}) \to 1$ (wp1). $\square$

For the verification of Assumptions 1.1-5 it is sufficient to verify that for sufficiently large n there exist constants C_1 and C_2 such that with probability 1,

$$C_1\sqrt{n} \leqslant t^*(w) \leqslant C_2\sqrt{n}.$$

The upper bound is established deterministically in [Few 1955]. The lower bound is derived in [Karp & Steele 1985, Exercise 5]. Therefore Theorem 3.3 implies a whole range of other convergence properties as well (cf. Figure 1.1).

3.1.2. Uniform vehicles. This subsection is an extension of the previous subsection in the same way that Subsection 2.1.2 was one of Subsection 2.1.1. At the first stage we have to select a subset from a set $\mathfrak{M}$ of *uniform* vehicles, knowing the *cost* c_i and the *speed* s_i of each vehicle $i \in \mathfrak{M}$. The set of first stage feasible solutions is $\mathfrak{X} = 2^{\mathfrak{M}}$, the power set of $\mathfrak{M}$. In this model the first stage cost are defined as $f(X) = \sum_{i \in X} c_i$. Corresponding to each $X \in \mathfrak{X}$ we define $s(X) = \sum_{i \in X} s_i$. We assume that constants c^L, c^U, s^L, s^U exist such that $c^L \leqslant c_i \leqslant c^U$ and $s^L \leqslant s_i \leqslant s^U$ ($i \in \mathfrak{M}$).

The second stage objective is to determine routes for the vehicles selected so as to minimize the maximum *time* required for any of the vehicles to traverse its route. Using the notation of the previous subsection we write the optimal second stage cost now as

$$y^*(X,w) = \max_{i \in X}\{v_i^*(X,w)/s_i\}.$$

Similar to relation (3.1), here

$$y^*(X,w) \geqslant \frac{t^*(w)}{s(X)}$$

for each X and each realization w of $\mathbf{w}$. And, analogously to (3.3) the function

$$z^{H_1}(X) = c(X) + \frac{\beta}{s(X)}\sqrt{n\pi r^2} \tag{3.9}$$

tends to be a lower bound on $z^*(X,w)$, with probability 1. Minimization of this function is NP-hard, so that we need an approximation of the minimum X^{H_1}. As in Subsection 2.1.2, a greedy heuristic will be used. We renumber the vehicles according to nondecreasing ratios c_i/s_i ($i \in \mathfrak{M}$). Let $C_i = \sum_{k=1}^{i} c_k$, $S_i = \sum_{k=1}^{i} s_k$, $Z_i = C_i + n\mu/S_i$ ($i \in \mathfrak{M}$) and $Z_0 = \infty$. The greedy heuristic selects a subset $X^G = \{1,\ldots,I\} \subset \mathfrak{M}$, where I is the largest index such that $Z_{I-1} > Z_I$. The quality of the greedy solution X^G is established in the following lemma, the proof of which is similar to the proof of Lemma 2.3 and therefore omitted here.

LEMMA 3.4. *The greedy solution X^G satisfies*
(i) $Z_I = \min_{i \in \mathfrak{M}}\{Z_i\}$;
(ii) $z^{H_1}(X^G) \leqslant z^{H_1}(X^{H_1}) + c^U$. $\square$

The heuristic for the solution of the second stage problem first constructs a spanning walk through all the customers, in the same way as was done in the previous subsection. Subsequently, rather than cutting the walk into pieces of equal length, we allocate a part of length $W(w)s_i/s(X^G)$ to the ith vehicle and transform this part into a route also in the manner described in the previous subsection. Let the length of the resulting route be denoted by $v_i^{H_2}(X^G,w)$, and let the heuristic value of the second stage routing problem be defined by

$$y^{H_2}(X^G,w) = \max_{i \in X^G}\{v_i^{H_2}(X^G,w)/s_i\}.$$

The value of the solution produced by the entire hierarchical system is denoted by

$$z^{H_2}(X^G,w) = f(X^G)+y^{H_2}(X^G,w).$$

For an asymptotic analysis of the heuristic (H_1,H_2), we first have to restrict the set $\mathfrak{M}$ of available vehicles, just as we had to restrict the set of available machines in Subsection 2.1.2.

ASSUMPTION 3.1. *The parameters c^L,c^U,s^L,s^U are fixed constants. Moreover, there exist constants $D>0$, $D'>0$, $\delta'\geqslant\delta>0$ such that $Dn^{1/4+\delta}\leqslant|\mathfrak{M}|\leqslant D'n^{1/4+\delta'}$.*

Verification of Condition A of Theorem 1.6 is simple from the construction of $z^{H_1}(X)$.

To verify Condition C of Theorem 1.6 we notice that Lemma 3.4 (ii) implies that

$$1 \leqslant \frac{z^{H_1}(X^G)}{z^{H_1}(X^{H_1})} \leqslant \frac{z^{H_1}(X^{H_1})+c^U}{z^{H_1}(X^{H_1})}.$$

The latter term tends to 1 when the number n of customers goes to infinity. This, together with inequality (3.3) implies Condition C. For the verification of Condition B we prove the following theorem, which shows that the heuristic value and the optimal value of the second stage problem are asymptotic to the same function, with probability 1.

THEOREM 3.4.

(i) $\dfrac{y^*(X,\mathbf{w})}{\beta\sqrt{n\pi r^2}/s(X)}\to 1$ (wp1);

(ii) $\dfrac{y^{H_2}(X,\mathbf{w})}{\beta\sqrt{n\pi r^2}/s(X)}\to 1$ (wp1).

PROOF. It is easy to see that analogous to (3.7) we have for this problem

$$\frac{t^*(w)}{s(X)} \leqslant y^*(X,w)\leqslant y^{H_2}(X,w)\leqslant\frac{t^*(w)}{s(X)} + \frac{O(\sqrt{n/d})}{s(X)}. \qquad (3.10)$$

Division by $\beta\sqrt{n\pi r^2}/s(X)$ yields

$$\frac{t^*(w)}{\beta\sqrt{n\pi r^2}} \leqslant \frac{y^*(X,w)}{\beta\sqrt{n\pi r^2}/s(X)} \leqslant \frac{y^{H_2}(X,w)}{\beta\sqrt{n\pi r^2}/s(X)} \qquad (3.11)$$

$$\leqslant \frac{t^*(w)}{\beta\sqrt{n\pi r^2}} + \frac{O(\sqrt{n/d})}{\beta\sqrt{n\pi r^2}}.$$

Since we have chosen $d=\log n$,

$$\frac{O(\sqrt{n/d})}{\beta\sqrt{n\pi r^2}} \to 0.$$

This observation, Theorem 3.1 and (3.11) imply the theorem. $\square$

Theorem 3.4 implies directly that Condition B of Theorem 1.6 holds for (H_1, H_2). Theorem 1.6 then establishes asymptotic clairvoyancy of the hierarchical system, with probability 1.

THEOREM 3.5.

$$\frac{z^{H_2}(X^G, \mathbf{w})}{z^*(X^\circ(\mathbf{w}), \mathbf{w})} \to 1 \quad (\text{wp1}). \quad \square$$

3.1.3. Random number of customers. In this section we consider extensions of the model in Subsection 3.1.1. In addition to uncertainty about the location of the customers we postulate uncertainty about their number at the aggregate level. Theorem 2.7 implies that the results of Section 3.1 carry over to this situation. A special case is obtained when the number of customers is fixed but each of the customers places an order with some fixed probability p. In this case the number N of customers to be visited is a random variable that has a binomial distribution and expectation np. Then

$$\frac{t^*(\mathbf{w})}{\sqrt{n}} = \frac{t^*(\mathbf{w})}{\sqrt{N}} \cdot \sqrt{\frac{N}{n}}.$$

Since $N/n \to p$ (wp1), Theorem 2.7 implies that

$$\frac{t^*(\mathbf{w})}{\sqrt{n}} \to \beta \sqrt{p \pi r^2} \quad (\text{wp1}).$$

Any reader can verify that the results of Subsection 3.1.1 carry over to this situation by substituting $p \pi r^2$ for πr^2.

3.2. A hierarchical location problem

In this section we consider a two-level decision problem with a depot location problem at the detailed level. With only probabilistic information about the locations of n customers one has to decide upon the number X of depots to be established. The set-up cost c is fixed for each depot. Hence $f(X) = cX$. We assume that the customers are located in a compact region R with area $v(R)$, according to the model described in the introduction to this chapter.

Let the cost to serve a customer be proportional to its distance to the depot from which it is served. Then, at the second stage, given a realization of the customer locations, the depots must be located so as to minimize the sum of the distances of each of the customers to its nearest depot. This problem is known in the literature as a *median location problem*. We assume that the cost per distance unit is 1. For the stochastic programming objective we use the notation that we have used so far throughout this book. In [Papadimitriou 1981] it is proved that determination of the optimal second stage cost $y^*(X, w)$

is an NP-hard combinatorial optimization problem. A heuristic to solve the two-stage decision problem is devised and analyzed along the lines of Section 1.5. Actually, this is no more than an application of results for median location problems available in the literature [Papadimitriou 1981], [Zemel 1984]. Specifically, in [Zemel, 1984] an asymptotic characterization of the optimal solution value of a median location problem is established.

THEOREM 3.6. *If* $X = o(\dfrac{n}{\log n})$, *then for* $\alpha = 0.377196...$

$$y^*(X,\mathbf{w})/n \;\rightarrow\; \alpha(\nu(R)/X)^{1/2} \quad (\text{wp1}). \quad \square$$

We use this result to formulate our heuristic first stage problem as determination of the integer value X^{H_1} that minimizes the function

$$z^{H_1}(X) = cX + \alpha n (\nu(R)/X)^{1/2}.$$

$z^{H_1}(X)$ is a convex function and it is easy to compute that X^{H_1} is equal to the most favorable integer round-off of $(\frac{1}{2}\alpha n \sqrt{\nu(R)/c}\,)^{2/3}$. The heuristic that we use for the solution of the second stage median location problem was proposed for the first time in [Papadimitriou 1981]. It received the telling name of 'honeycomb' heuristic. The region R is tiled with hexagons each with area $\nu(R)/X^{H_1}$ and in each of them a depot is established. In each of the hexagons that are proper subsets of R a depot is located. Obviously their number is less than X^{H_1}, and it may be completed to X^{H_1} by the location of depots in hexagons that are selected arbitrarily from the hexagons that intersect the boundary of R. In each hexagon the depot is located in the customer location that has minimal total distance to the other customers in the hexagon in question. In [Zemel 1984] it is shown that this heuristic is asymptotically optimal with probability 1 if $X = o(n/\log n)$. Since in our heuristic $X^{H_1} \approx (\frac{1}{2}\alpha n \sqrt{\nu(R)/c}\,)^{2/3}$, the latter condition is satisfied and Zemel's result implies Condition B of Theorem 1.4. We note that the estimate used in the first stage problem is not so much a lower bound on the optimal value but rather an almost sure asymptotically accurate approximation of the optimal value. Hence a stronger condition than Condition A of Theorem 1.4 is satisfied. As a result the hierarchical planning system proposed is asymptotically clairvoyant with probability 1.

3.3. A hierarchical location and routing problem

As has been argued in the introduction to this chapter, the problem that we study in this section can be regarded as an extension of the hierarchical location problem studied in the previous section.

At the aggregate level one has to decide upon a number X of depots to be established at a cost c each. At the detailed level the depots are to be located and the n customers must be served by vehicles, each of which is located at a depot and has a capacity of serving q customers. The customers are located randomly in a compact region R with area $\nu(R)$ according to the model

described in the introduction to this chapter. The second stage objective is to locate the depots and route the vehicles so as to minimize the total distance to be traveled to serve all customers, given a realization of their locations. The optimal cost is indicated by $u^*(X,w)$ for any X and any realization w of $\mathbf{w}$. The sum of this cost and the first stage cost $f(X)=cX$ is denoted by $z^*(X,w)$.

The above second stage problem is NP-hard, since it generalizes the median location problem (see Section 3.2), in which $q=1$. Along the lines of Section 1.5 we devise a hierarchical planning system to solve the two-stage decision problem and we prove that this system is asymptotically clairvoyant with probability 1. For the formulation of a heuristic first stage problem we first define $y^*(X,w)$ as the optimal solution value of the median location problem with X depots and customer locations represented by w. Then the following lemma establishes a lower bound on $u^*(X,w)$.

LEMMA 3.5.

$$u^*(X,w) \geqslant \frac{2y^*(X,w)}{q}.$$

PROOF. Consider a group of q customers to be served by one vehicle. In one or another way the vehicle has to reach the customer in this group that is farthest from a depot, and return from there to the depot. This implies that the length of any route for this vehicle through this group of q customers is at least twice the distance from this customer to its nearest depot. This is certainly more than or equal to twice the average distance of the q customers to their nearest depots, given the depot location. Summation over disjunct groups of at most q customers, the union of which is the group of all n customers, yields that the total distance to be traveled is at least the sum of the distances of the n customers to their nearest depot divided by q, which is the optimal median location cost divided by q. $\square$

We combine this lemma with the asymptotic characterization of the optimal value of the median location problem given in Theorem 3.6 so as to arrive at an approximation $u^{H_1}(X)$ of $u^*(X,\mathbf{w})$ that satisfies Condition A of Theorem 1.4 for the values X that satisfy $X = o(n/\log n)$:

$$u^{H_1}(X) = \frac{2\alpha n}{q}(v(R)/X)^{1/2}.$$

The heuristic first stage decision X^{H_1} is then the integer value of X that minimizes

$$z^{H_1}(X) = cX + \frac{2\alpha n}{q}(v(R)/X)^{1/2}.$$

The function $z^{H_1}(X)$ is convex. It turns out that X^{H_1} is equal to the most favorable integer round-off of $(\alpha n \sqrt{v(R)}/cq)^{2/3}$. We observe that given this value Condition A of Theorem 1.4 is indeed satisfied.
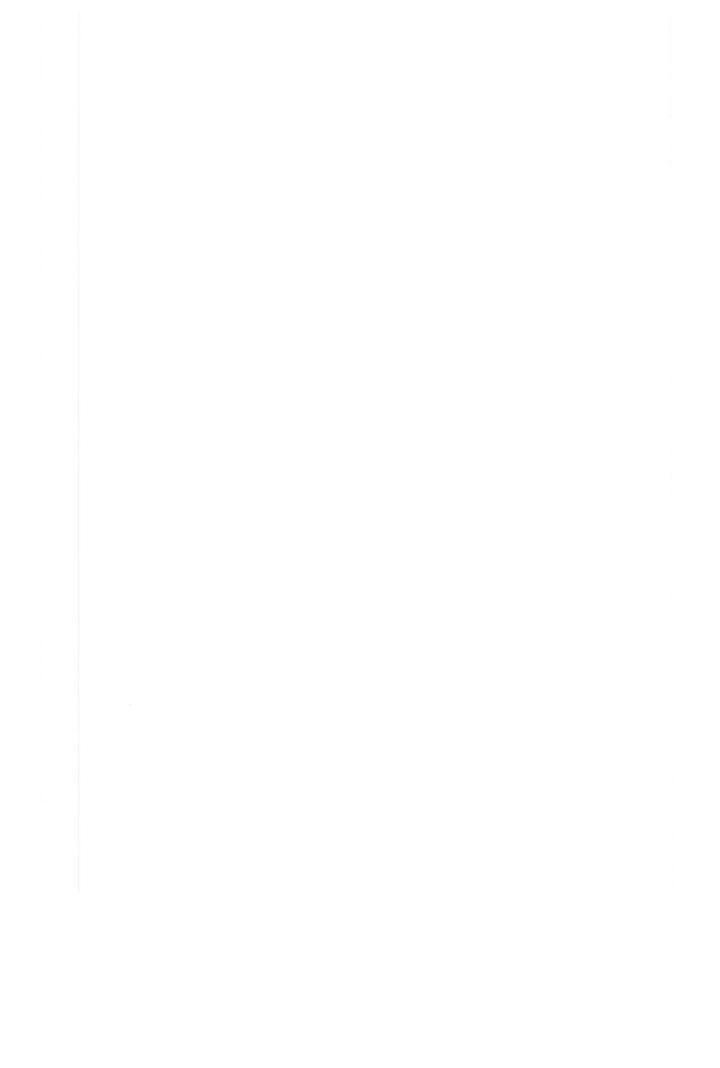
The heuristic that we devise for the second stage distribution problem is a composition of the honeycomb heuristic described in Section 3.2 for the location of the depots and a tour partitioning heuristic proposed in [Haimovich & Rinnooy Kan 1983] for the routing of the vehicles. In the first step of the tour partitioning heuristic a traveling salesman tour through all n customers is constructed using a space partitioning heuristic similar to the one described in Section 3.1.1. At the second step this tour is cut into pieces that contain q customers each. Each of these q-chains is connected to a depot as follows. We select the customer in the chain that is closest to any of the depots, and we add the corresponding segment to the chain. In this way we obtain a spanning tree through the q customers and a depot. This is transformed into a spanning walk by duplicating the tree. The spanning walk is then converted into a tour for a vehicle by means of the operations described in Section 3.1.1. The cost of the heuristic second stage configuration, denoted as $u^{H_2}(X,w)$, is bounded from above by twice the length of the traveling salesman tour produced plus twice the sum of the lengths of the additional segments mentioned above. An upper bound on the traveling salesman tour has been established already in Section 3.1.1, and is equal to the length of an optimal traveling salesman tour denoted by $t^*(w)$ plus a term that is $O(\sqrt{n/\log n})$. An upper bound on the sum of the lengths of the segments added is $y^{H_2}(X^{H_1},w)/q$, since from each group of q customers we consider only the one closest to a depot established by the honeycomb heuristic. It is trivial that the corresponding distance is less than the average distance of the q customers to their nearest depots. The above observations together yield

$$u^{H_2}(X^{H_1},w) \leqslant 2t^*(w) + O(\sqrt{n/\log n}) + 2y^{H_2}(X^{H_1},w)/q.$$

Therefore

$$\frac{u^{H_2}(X^{H_1},w)}{u^{H_1}(X^{H_1})} \leqslant \frac{t^*(w)}{(c\nu(R))^{1/3}(\alpha n/q)^{2/3}} + \frac{O(\sqrt{n/\log n})}{(c\nu(R))^{1/3}(\alpha n/q)^{2/3}} + \frac{y^{H_2}(X^{H_1},w)}{(c\nu(R))^{1/3}(\alpha n/q)^{2/3}}.$$

Since $t^*(w)$ is asymptotic to $\beta\sqrt{n\nu(R)}$ (cf. Theorem 3.1) the first term of the right hand side of the above inequality tends to 0 with probability 1. Obviously, the second term tends to 0. As $X^{H_1} = o(n/\log n)$, the almost sure asymptotic optimality of the honeycomb heuristic together with Theorem 3.6 implies that the third term tends to 1 with probability 1. The above implies that the second stage cost is asymptotically determined only by the lengths of the segments, which can be interpreted as the median-location cost of n/q groups of customers, where each group is considered as one entity. Together, these three observations imply that $u^{H_2}(X^{H_1},w)/u^{H_1}(X^{H_1})$ tends to 1 with probability 1. This implies in its turn that the hierarchical system (H_1, H_2) satisfies Condition B of Theorem 1.4. Now, this theorem establishes asymptotic clairvoyancy of the hierarchical system as a whole.

4

Stochastic integer programming by

dynamic programming

Whereas the preceding three chapters dealt with *heuristics* for stochastic integer programming problems, this chapter is devoted to true *optimization methods*.

As has been argued in the introduction, the only hope to achieve anything useful in this direction is through the exploitation of *special structure*. The special structure of the stochastic integer programming problems in this chapter is derived first from the special structure of the second stage combinatorial problems and secondly from the assumption that the stochastic parameters have discrete distributions with a fixed (small) number of points in which the probability mass is concentrated. For example, in the hierarchical scheduling problem that we consider in Section 4.1 we assume that the processing times of the jobs can take on only a limited number of possible values.

We will show how this structure leads to recurrence relations between the various feasible solutions of the problems, which can be efficiently exploited by *dynamic programming* routines. Analysis of the routines will show that their running times are polynomially bounded in the problem size (such as the number of jobs in the scheduling problem) but exponential in the number of possible realizations of the stochastic parameters.

In Sections 4.1 and 4.2 we consider the two hierarchical *scheduling* problems described in Sections 2.1 and 2.2 respectively. We will refer to the hierarchical scheduling problem of Section 2.2 as the hierarchical *bin packing* problem. In Section 4.3 we present a dynamic programming algorithm for solving a hierarchical *multi-knapsack* problem. This problem can be viewed as a capital

72

budgeting problem. At the aggregate level one has to decide on the sizes of budgets that are to be reserved for financing a number of projects at the detailed level. Each of the budgets has a given unit price and each of the projects requires a certain amount of the budgets and has a profit, which is initially known only stochastically. A remarkable feature of this problem is that the solution of the distribution version can be read directly from the problem parameters and does not require any recursive calculations (see Section 4.3.1).

In each of the sections there is a subsection in which the basic dynamic programming algorithm for the problem in question is presented and a subsection in which we report on computational results. In the latter subsections we study the shape of the objective function of the two-stage decision problem as a function of the first stage decision variables. In Sections 4.1.2 and 4.2.2 we compare heuristic and optimal values. In Sections 4.1 and 4.2 we use the notation of Sections 2.1 and 2.2 respectively.

4.1. Dynamic programming for hierarchical scheduling

In this section we restrict ourselves to instances of the problem described in Section 2.1 in which the processing times $\mathbf{w}_1, \ldots, \mathbf{w}_n$ have a common discrete distribution with a fixed number k of distinct values $a_1,...,a_k$ in its support. We will show in Subsection 4.1.1 that this restriction allows for the construction of a dynamic programming algorithm that calculates the values of the optimal solutions of the problem for all possible realizations of the processing times and for each relevant number of machines, and hence solves the problem in a time that is polynomial in the number n of jobs but exponential in k. Refinements of this algorithm are presented subsequently. In Subsection 4.1.2 we report on computational experience obtained by implementation of the dynamic programming routine.

4.1.1. Dynamic programming. As mentioned in Section 2.1 determining the optimal second stage scheduling cost $y^*(X,w)$ in NP-hard. Let us denote by $w =[n_1,...,n_k]$ the vector of processing times in which the value a_j occurs n_j times, for $j = 1,...,k$.

One can obtain an optimal schedule on X machines by assigning a certain subset of jobs optimally to $X-1$ machines and putting the remaining jobs on another machine. This observation leads to the following recurrence relations:

$$y^*(X,[n_1,...,n_k])=\min\{\max\{y^*(X-1,[n_1-\ell_1,...,n_k-\ell_k]),y^*(1,[\ell_1,...,\ell_k])\}$$

$$|0\leqslant\ell_j\leqslant n_j(j=1,...,k)\} \ \ (X>1),$$

$$y^*(1,[n_1,...,n_k])=\sum_{j=1}^k n_j a_j.$$

Computation of $y^*(X,w)$ by a dynamic programming algorithm based on this recursion requires $O(X\Pi_{j=1}^k n_j)$ time, which is exponential in k but

polynomial for fixed k.

In the more general context of the two-stage scheduling problem, we assume that the processing times have a common discrete distribution with k integral values $a_1,...,a_k$ in its support. The independence of the processing times implies that $\mathbf{w}=[\mathbf{n}_1, \ldots, \mathbf{n}_k]$ has a multinomial distribution. The idea is now to go through the entire recursion once in order to compute $y^*(X,w)$ for all values $X \in \{1,...,n\}$ and for all realizations $w \in \mathcal{W}$, where $\mathcal{W}$ is given by

$$\mathcal{W} = \{[n_1,...,n_k] \,|\, 0 \leqslant n_j \leqslant n \,(j=1,...,k),\ n_1 + ... + n_k = n\}.$$

The distribution model is then solved by the selection, for each $w \in \mathcal{W}$, of a value of X that minimizes $z^*(X,w)=cX+y^*(X,w)$. The two-stage decision model is solved by the determination of a value of X that minimizes $Ez^*(X,\mathbf{w})$ $=cX+\sum_{w \in \mathcal{W}} Pr\{\mathbf{w}=w\}y^*(X,w)$.

A straightforward application of the above dynamic programming algorithm requires $O(n^k)$ comparisons for each of the $O(n^{k+1})$ pairs (X,w), and hence $O(n^{2k+1})$ time altogether. The multinomial probabilities are easily computed within this time bound.

A more efficient implementation of the algorithm is obtained as follows. Let $a_1 = \max\{a_1,...,a_k\}$. It is not hard to see that, for any X and $w=[n_1, \ldots, n_k]$

$$\left\lceil \textstyle\sum_{j=1}^k n_j a_j / X \right\rceil \leqslant y^*(X,[n_1, \ldots, n_k]) \leqslant \left\lceil \textstyle\sum_{j=1}^k n_j a_j / X \right\rceil + a_1 - 1.$$

The lower bound is trivial, and the upper bound follows from the observation that any list scheduling algorithm will start every job strictly before the lower bound. Further, we assume without loss of generality that in the above recurrence relations the second maximand attains the maximum:

$$y^*(X,[n_1,...,n_k])=y^*(1,[\ell_1, \ldots, \ell_k]) \text{ for some } \ell_1, \ldots, \ell_k.$$

We can therefore restrict our attention to vectors $[\ell_1, \ldots, \ell_k]$ that yield a value $y^*(1,[\ell_1, \ldots, \ell_k])$ within a given range of a_1 integers. This implies that only a single value of ℓ_1 has to be considered for given $\ell_2, \ldots, \ell_k$ and that $O(n^{k-1})$ comparisons suffice for each pair (X,w). The overall running time is thereby reduced to $O(n^{2k})$.

Other, more intricate, refinements lead to a running time of $O(n^{2k-1} a_1^{2k-3} \log na_1)$. Although that implementation is more efficient for small values $a_1, \ldots, a_k$, it is of little avail in view of the results that will be presented in Section 4.2.1.

4.1.2. Computational results. The dynamic programming algorithm was coded in PASCAL and run on a CD Cyber 170-750 to solve several instances of the two-stage scheduling problem. The solution of instances with 100 jobs and two possible processing time values or with 50 jobs and three processing time values required about 30 seconds. The values of k considered are admittedly small, but the values of n are realistic and the running times are such that the brute force approach of this chapter should not be dismissed on grounds of

manifest inefficiency.

We illustrate the numerical results on a set of representative instances given by

$$c = 1,$$

$$n = 1, \ldots, 100,$$

$$k = 2, a_1 = 18, a_2 = 14,$$

$$Pr\{\mathbf{w}_j = a_1\} = Pr\{\mathbf{w}_j = a_2\} = \tfrac{1}{2} \ (j = 1, \ldots, n).$$

Figure 4.1 shows four functions of the number of jobs:
- the minimal lower bound $\min_X \{z^{H_1}(X)\}$ mentioned in Section 2.1.1;
- the minimal expected total cost $Ez^*(X^*, \mathbf{w})$ (the optimum for the two-stage decision model);
- the expected minimal total cost $Ez^*(X^\circ(\mathbf{w}), \mathbf{w})$ (the optimum for the distribution model, averaged over all realizations);
- the expected approximate total cost obtained by the heuristic designed in Section 2.1.1.

Note that the last three functions are defined only for integral n; linear interpolation has been applied to improve the presentation. The distribution model yields slightly better results than the two-stage decision model on average, as expected. A comparison between the optima and the lower and upper bounds confirms that the absolute differences are significant while the relative differences disappear with increasing problem size.

For the case that $n = 100$, Figure 4.2 shows three functions of the first stage decision variable, the number X of machines:
- the lower bound $z^{H_1}(X)$;
- the expected total cost $Ez^*(X, \mathbf{w})$ in case of an optimal second stage decision;
- the expected total cost in case of an approximate second stage decision.

Note that we have interpreted X as a continuous variable: acquisition of a fractional machine costs a fraction of c but yields no benefit at the second stage; the vertical line segments correspond to discontinuities. In spite of the smoothing effect due to averaging over all realizations, both the optimal and the approximate cost functions are highly nonconvex and multimodal. The functions consist of a first stage component, which is linear and increasing, and a second stage component, which is nonconvex and nonincreasing. Addition of the two components can turn the nonconvexities into local minima, and small values of c appear to be most effective in this respect.

4.2. Dynamic programming for hierarchical bin packing

As in the previous section we will show that, when we restrict ourselves to k distinct values $a_1,...,a_k$ as possible realizations of each of the item sizes, for fixed k, we can use a dynamic programming algorithm to solve the hierarchical bin packing problem within polynomial time. This algorithm and a refinement of it are presented in Subsection 4.2.1. In Subsection 4.2.2 we report on computational experience obtained by implementation of the algorithm.

4.2.1. Dynamic programming. As mentioned in Section 2.2 determining the optimal second stage bin packing cost $x^*(Y,w)$ for given Y and w is NP-hard. We write $w =[n_1, \ldots , n_k]$ to denote the vector in which the value a_j occurs n_j times, for $j =1,...,k$.

The following dynamic programming algorithm is due to [Held, Karp & Shareshian 1963]. Let $C(Y,w)$ be the total amount of capacity needed to pack items with weights specified by w into bins of capacity Y. It is assumed that $C(Y,w)$ includes the slack capacity of each bin (which is equal to Y minus the total weight of the items assigned to that bin) except for the slack capacity of the last bin. Thus, if $C(Y,w)=xY-\Lambda$ with $x \in \mathbb{Z}_+$ and $0 \leq \Lambda < Y$, then an optimal packing requires x bins and the last bin has a slack capacity of Λ. Let $\Delta(Y,w,a)$ be the extra capacity needed when an item with weight a is added to this packing:

$$\Delta(Y,w,a) = \begin{cases} a & \text{if } \Lambda \geq a, \\ \Lambda+a & \text{if } \Lambda < a. \end{cases}$$

It is not hard to see that

$$C(Y,[n_1, \ldots , n_k])=\min_{1\leq j \leq k :n_j >0}\{C(Y,[n_1, \ldots , n_{j-1},n_j -1,n_{j+1},...,n_k])$$
$$+\Delta(Y,[n_1, \ldots , n_{j-1},n_j -1,n_{j+1}, \ldots , n_k],a_j)\}$$
$$(n_1+ \cdots +n_k >0),$$
$$C(Y,[0,...,0])=0.$$

We finally have that $x^*(Y,w)= \lceil C(Y,w)/Y \rceil$.

For the two-stage bin packing problem, we make the same assumptions concerning the distribution of the stochastic parameters as in Section 4.1.1 and apply the same strategy to obtain solutions to both stochastic optimization models. Since the values $a_1, \ldots , a_k$ are integral, there is no loss of generality in considering only integral capacities Y. Let $a_{max}=\max\{a_1, \ldots , a_k\}$ and note that $1 \leq Y \leq na_{max}$. The algorithm requires a fixed number of comparisons for each of the $O(n^{k+1}a_{max})$ pairs (Y,w), and hence $O(n^{k+1}a_{max})$ time altogether.

A more efficient implementation of the algorithm is obtained as follows. Let $a_{sum}=\sum_{j=1}^{k}n_j a_j$. It is not hard to see that, for any Y and $w =[n_1,...,n_k]$

$$\lceil a_{sum}/Y\rceil \leqslant x^*(Y,[n_1,\ldots,n_k]) \leqslant 2\lceil a_{sum}/Y\rceil -1.$$

The lower bound is trivial. The upper bound is a performance guarantee of the following simple heuristic: deal with the items in a fixed order and fill each of $\lceil a_{sum}/Y\rceil$ bins successively, thereby splitting an item if necessary; next, reassign each of the split items to a separate bin, of which no more than $\lceil a_{sum}/Y\rceil -1$ will be needed. Addition of the first stage cost yields

$$dY + a_{sum}/Y \leqslant z^*(Y,[n_1,...,n_k]) \leqslant dY + 2a_{sum}/Y + 1.$$

These lower and upper bound functions are both convex in Y. The function $z^*(Y,w)$ therefore attains its minimum for a value of Y that is bounded by the two values of the argument for which the lower bound is equal to the minimum of the upper bound. A straightforward calculation shows that the latter values are given by $(1/2+(2a_{sum}d)^{1/2}\pm (a_{sum}d+(2a_{sum}d)^{1/2}+1/4)^{1/2})/d$. This implies that for all n^k realizations w only $O((na_{max}/d)^{1/2})$ values of Y have to be considered. The overall running time is thereby reduced to $O(n^{k+1/2}a_{max}^{1/2}d^{-1/2})$.

Due to the relation between the two-stage scheduling and bin packing problems that was observed above, the $Y^*(x,w)$ values from Section 4.1.1 could be used to derive the $x^*(Y,w)$ values needed here and vice versa, as long as the set $\{a_1,\ldots,a_k\}$ is the same in both cases. The former recursion has the advantage of requiring strictly polynomial time; the latter one is pseudopolynomial but much faster for small values $a_1,\ldots,a_k$.

4.2.2. Computational results.

For the typical problem instance given by

$$d = 1,$$

$$n = 100,$$

$$k = 2, a_1 = 18, a_2 = 14,$$

$$Pr\{\mathbf{w}_j = a_1\} = Pr\{\mathbf{w}_j = a_2\} = \tfrac{1}{2} \ (j = 1,...,n),$$

Figure 4.3 shows three functions of the first stage decision variable, the capacity Y:
- the lower bound $z^{H_1}(Y)$;
- the expected total cost $Ez^*(Y,\mathbf{w})$ in case of an optimal second stage decision;
- the expected total cost in case of an approximate second stage decision.

An investigation of these and other results leads to the same conclusions concerning running time, quality of lower and upper bounds, and the occurrence of multiple local minima as in Section 4.1.2.

4.3. Dynamic programming for hierarchical multiknapsack

At the aggregate level of the capital budgeting problem that we consider here, one has to decide on the sizes $X_1,...,X_m$ of m budgets that are to be reserved for financing a number of projects, while knowing the cost c_i of reserving one unit of budget i ($i = 1,...,m$), the requirement r_{ij} of project j out of budget i ($i = 1,...,m$, $j = 1,...,n$), and the probability distribution of the vector $\mathbf{w} = (\mathbf{w}_1, \ldots, \mathbf{w}_n)$ of revenues that the projects will yield. It is assumed that all c_i, r_{ij} and $\mathbf{w}_j$ are nonnegative and that the r_{ij} are integral. At the detailed level, after $X = (X_1,...,X_m)$ has been determined, a realization w of $\mathbf{w}$ becomes known, and one has to decide on a selection S of the projects that maximizes the total revenue $y^*(X,w)$ within the budget constraints:

$$y^*(X,w) = \max_{S \subseteq \{1,...,n\}} \left\{ \sum_{j \in S} w_j \mid \sum_{j \in S} r_{ij} \leqslant X_i \ (i = 1,...,m) \right\}.$$

The total profit of the budgeting decision X and the optimal selection decision is denoted by $z^*(X,w) = -\sum_{i=1}^m c_i X_i + y^*(X,w)$.

In the two-stage decision model, the objective is to determine a vector $X^* \in \mathbb{R}_+^m$ such that

$$Ez^*(X^*,\mathbf{w}) = \max_{X \in \mathbb{R}_+^m} \{Ez^*(X,\mathbf{w})\}.$$

Let $\mathcal{W}$ be the set of all possible realizations of $\mathbf{w}$. In the distribution model, the objective is to determine a function $X^\circ : \mathcal{W} \to \mathbb{R}_+^m$ such that

$$z^*(X^\circ(w),w) = \max_{X \in \mathbb{R}_+^m} \{z^*(X,w)\}, \forall w \in \mathcal{W}.$$

4.3.1. The distribution model.

The knapsack problem, i.e., the second stage problem with $m = 1$, is already NP-hard [Garey & Johnson 1979]. Surprisingly, the distribution model is easily solved to optimality. For each $w \in \mathcal{W}$, the selection $S(w)$ of profitable projects is given by $S(w) = \{j \mid w_j - \sum_{i=1}^m c_i r_{ij} > 0\}$. The minimum budgets needed to finance these projects are equal to $X_i^\circ(w) = \sum_{j \in S(w)} r_{ij} (i = 1,...,m)$, and the corresponding total profit is

$$z^*(X^\circ(w),w) = \sum_{j \in S(w)} \left(w_j - \sum_{i=1}^m c_i r_{ij} \right), \ \forall w \in \mathcal{W}.$$

In the situation that each revenue w_j can assume only k distinct values, the determination of X° requires $O(mn)$ computations for each of k^n realizations w.

4.3.2. Dynamic programming.

The second stage multiknapsack problem is solvable by a classical dynamic programming algorithm from [Bellman 1957]. Let $F_j(X,w)$ be the maximum revenue if only the first j projects can be selected, for given budgets $X = (X_1,...,X_m)$ and revenues $w = (w_1, \ldots, w_n)$. An optimal selection is either restricted to the first $j - 1$ projects or includes project j:

$$F_j((X_1, \ldots, X_m), w) = \max\{F_{j-1}((X_1, \ldots, X_m), w),$$

$$F_{j-1}((X_1 - r_{1j}, \ldots, X_m - r_{mj}), w) + w_j\} (j = 1, \ldots, n),$$

$$F_0((X_1, \ldots, X_m), w) = \begin{cases} 0 & \text{if } X_1 = \cdots = X_m = 0, \\ -\infty & \text{otherwise.} \end{cases}$$

Since the requirements r_{ij} are integral, also the budgets X_i can be assumed to be integral. Computation of $y^*(X, w) = F_n(X, w)$ requires a single comparison for each of $\prod_{i=1}^{m} X_i$ vectors $X' \leqslant X$ at each of n successive stages, and hence $O(n \prod_{i=1}^{m} X_i)$ time altogether.

For the two-stage multiknapsack problem, we again consider the situation in which each revenue w_j can assume only k distinct values, for a fixed k. Let $R_i = \sum_{j=1}^{n} r_{ij}$ and note that $0 \leqslant X_i \leqslant R_i$ $(i = 1, \ldots, m)$. At stage j, only the k^j different realizations of $(w_1, \ldots, w_j)$ need to be distinguished $(j = 1, \ldots, n)$. The algorithm therefore has to consider $O(k^j \prod_{i=1}^{m} R_i)$ pairs (X, w) at stage j. Summation over all j yields an $O(k^n \prod_{i=1}^{m} R_i)$ time bound for the computation of all $y^*(X, w)$ and also for the determination of a budget vector X^* that is optimal in expectation.

4.3.3. Computational results

The dynamic programming algorithm was coded in PASCAL and run on a CD Cyber 170-750 to solve several instances of the two-stage knapsack problem. We set $m = 1$ at the outset and did not attempt to solve proper multiknapsack problems, for which $m \geqslant 2$. We assumed independence of the revenues w_j and tried to make the second stage knapsack problem nontrivial by specifying a high correlation between the expected revenue $E w_j$ of project j and its budget requirement r_{1j}. The solution of instances with twelve projects and two possible revenue values for each of them required about ten seconds.

For the problem instance given by

$$m = 1, c = 1,$$

$$n = 12, Pr\{w_j = a_{1j}\} = Pr\{w_j = a_{2j}\} = \tfrac{1}{2} \ (j = 1, \ldots, n),$$

with the values of $r_{1j}, a_{1j}, a_{2j} (j = 1, \ldots, n)$ given in Table 4.1, Figure 4.4 shows the expected total profit $Ez^*((X_1), w)$ as a function of the budget size X_1. Note that the profit is shown only for integral X_1; the line segments that start from the points shown with a slope $-c_1$ and that indicate the profit for fractional X_1 have been deleted. Even if we restrict our attention to integral values of X_1, the profit function has many local maxima.

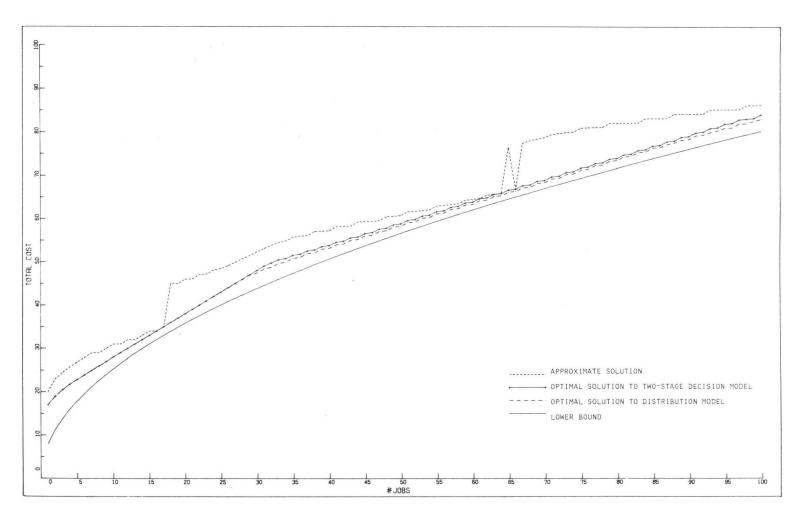| j	1	2	3	4	5	6	7	8	9	10	11	12
r_{1j}	5	2	9	13	10	8	4	7	10	6	4	9
a_{1j}	7	4	12	17	15	12	5	9	14	9	6	1
a_{2j}	3	1	6	11	8	7	1	4	7	7	2	8

TABLE 4.1. Knapsack: numerical data

FIGURE 4.1. Scheduling: the total cost as a function of the number of jobs
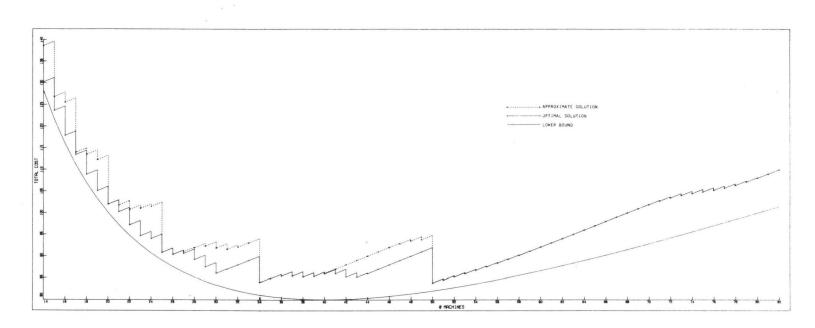
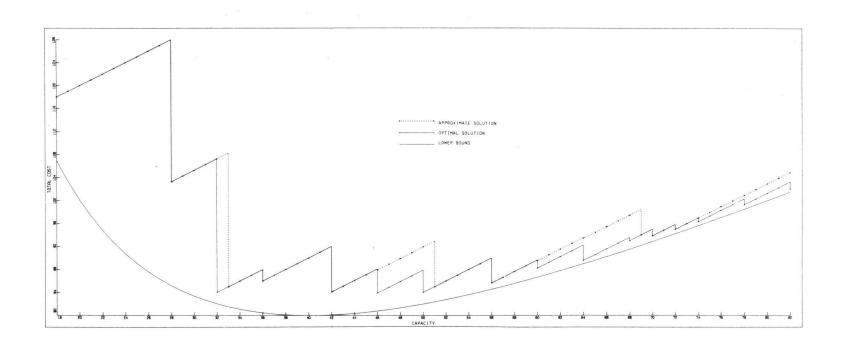FIGURE 4.2. Scheduling: the total cost as a function of the number of machines

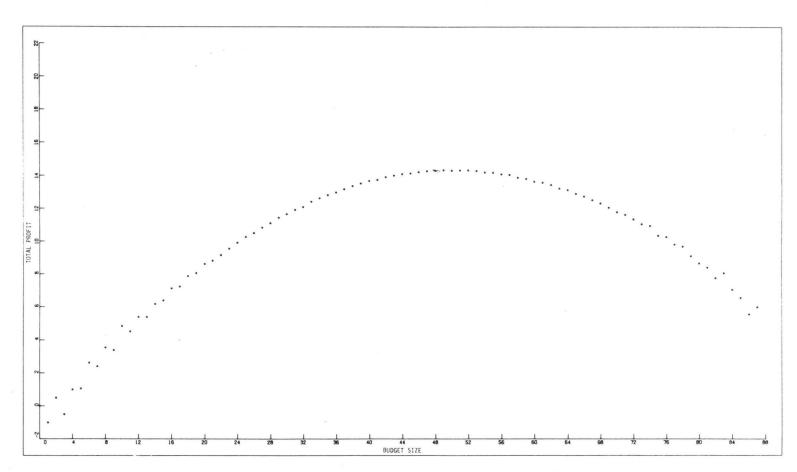FIGURE 4.3. Bin packing: the total cost as a function of the bin capacity

FIGURE 4.4. Knapsack: the total profit as a function of the budget size

5

The structure of stochastic integer

programming problems

In this short chapter we investigate the structure of objective functions of stochastic integer programming problems as a first step towards the development of optimization methods for their solution. We have seen in the previous chapter that these functions can be discontinuous and non convex. These are properties that do not work in favor of the design of a successful optimization routine. However, the problems studied in Chapter 4 were of a restricted nature. In this chapter a much broader class of stochastic integer programming problems is considered. In Section 5.1 we define a general stochastic integer programming problem and we analyze its objective function. The main results, that we derive are that, when the parameters have a continuous distribution, the objective function is continuous but not necessarily convex and if the parameters have a discrete distribution, it is in general discontinuous.

These results represent no more than an initial theoretical insight into the structure of stochastic integer programs. We are still far from a general algorithm for such problems. In Section 5.2 we review some directions for future research that may lead to such an algorithm.

5.1. The stochastic integer programming objective function
Let us consider the general *linear programming* model

$$\min \ cx \tag{5.1}$$

$$\text{s.t. } Ax = b, \tag{5.2}$$

$$x \geqslant 0 \tag{5.3}$$

with $c, x \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^m \times \mathbb{R}^n$. The fundamental extension that is obtained by the additional constraint

$$x \in \mathbb{Z}^n \tag{5.4}$$

yields the general *integer programming* problem. We have already discussed the computational difficulties that this extension gives rise to. Many of these difficulties carry over to stochastic integer programming.

Let us define the random vectors $\mathbf{q} \in \mathbb{R}^k$, $\mathbf{p} \in \mathbb{R}^l$ and the random matrices $\mathbf{W} \in \mathbb{R}^l \times \mathbb{R}^k$, $\mathbf{T} \in \mathbb{R}^l \times \mathbb{R}^n$. From the general two-stage stochastic linear programming problem

$$\min \ cx + E \ \min\{\mathbf{q}y \mid \mathbf{W}y \geqslant \mathbf{T}x + \mathbf{p}, y \geqslant 0, y \in \mathbb{R}^k\} \tag{5.5}$$

$$\text{s.t. } Ax = b, \tag{5.6}$$

$$x \geqslant 0 \tag{5.7}$$

we derive the general two-stage stochastic integer programming problem

$$\min \ cx + E \min\{\mathbf{q}y \mid \mathbf{W}y \geqslant \mathbf{T}x + \mathbf{p}, y \geqslant 0, y \in \mathbb{Z}^k\} \tag{5.8}$$

$$\text{s.t. } Ax = b, \tag{5.9}$$

$$x \geqslant 0, \tag{5.10}$$

$$x \in \mathbb{Z}^n. \tag{5.11}$$

To study the shape of the stochastic integer programming objective function we introduce the concept of a *value function*. If we consider c and A in (5.1) and (5.2) as being fixed, then the optimal solution to the linear program (5.1-3) and the integer program (5.1-4) are functions of the right hand side b. These functions, denoted by $Q^{LP}(b)$ and $Q^I(b)$ respectively, are called value functions.

In [Blair & Jeroslow 1982] the classes of value functions of linear and integer programs are characterized. They showed that these can be constructed iteratively by simple operations. Each value function of a linear program is obtained by starting with linear functions of the form λb ($\lambda \in \mathbb{R}^m$), and by repeating finitely often the operations of taking sums, taking maxima and taking nonnegative multiples of the functions already obtained. And reversely, each function that can be constructed in this way is the value function of some linear program. Thus, for example, the function $Q^{LP}((b_1, b_2)) = \max\{2b_1 + \frac{3}{2}b_2, b_1 + 5b_2\}$ is the value function of some two-constraint linear program.
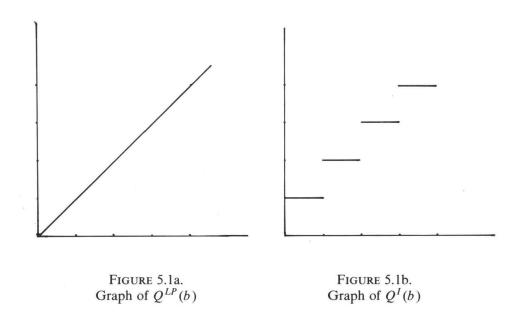
The value function of an integer program is obtained in the same way except that the operation of taking integer round-ups is added to the set of operations. Also here each function that can be constructed in this way is the value function of some integer program. Thus, for example, the function $Q^I((b_1,b_2)) = \max\{2b_1 + \frac{3}{2}b_2, b_1 + 3b_2 + \lceil \frac{5}{4}b_2 \rceil\}$ is the value function of some two-constraint integer program. The addition of the round-up operation induces the irregular behavior of integer programming value functions. Whereas, by the way they are constructed, linear programming value functions are continuous, piecewise linear and convex, integer programming value functions are in general discontinuous. Consider for example the simple value functions

$$Q^{LP}(b) = \min\{y \mid y \geqslant b, y \geqslant 0\}$$

and

$$Q^I(b) = \min\{y \mid y \geqslant b, y \geqslant 0, y \in \mathbb{Z}\}.$$

Their graphs, depicted in Figure 5.1, show the peculiar discontinuities that integer programming can generate.

FIGURE 5.1a.
Graph of $Q^{LP}(b)$

FIGURE 5.1b.
Graph of $Q^I(b)$

The above features of the value functions carry over to the objective functions of stochastic programming problems. For a given realization

(q, P, W, T) of $(\mathbf{q}, \mathbf{P}, \mathbf{W}, \mathbf{T})$,

$$Q^{LP}(Tx + p) = \min\{qy \mid Wy \geq Tx + p, y \geq 0\}$$

is the value function of a realization of the second stage problem of the stochastic linear program (5.1-3). We notice that it is a random function. The expected optimal second stage cost is given by the function

$$\mathcal{Q}^{LP}(x) = E_{q,p,W,T} Q^{LP}(Tx + \mathbf{p})$$

that depends only on the first stage decision variable x. In a similar way we define the functions

$$Q^{I}(Tx + p) = \min\{qy \mid Wy \geq Tx + p, y \geq 0, y \in \mathbb{Z}^{k}\}$$

and

$$\mathcal{Q}^{I}(x) = E_{q,p,W,T} Q^{I}(\mathbf{T}x + \mathbf{p}).$$

When the random parameters have discrete distributions with a finite number of points with positive density, then the functions $\mathcal{Q}^{LP}(x)$ and $\mathcal{Q}^{I}(x)$ are convex combinations of a finite number of value functions. Therefore the objective functions of the corresponding stochastic linear programs are piecewise linear and convex, whereas the objective functions of the stochastic integer programs may be discontinuous. To illustrate the above we consider the following simple functions:

$$z^{LP}(x) = x + E \min\{y \mid y \geq \mathbf{p} - x, y \geq 0\} \tag{5.12}$$

and

$$z^{I}(x) = x + E \min\{y \mid y \geq \mathbf{p} - x, y \geq 0, y \in \mathbb{Z}\}, \tag{5.13}$$

where $\mathbf{p}$ is a random variable with distribution $Pr\{\mathbf{p} = 2\} = Pr\{\mathbf{p} = 2.5\} = \frac{1}{2}$. Their graphs are given in Figure 5.2. As in the deterministic programming examples, the objective function of the stochastic linear program is piecewise linear and convex and the objective of the stochastic integer program is discontinuous.

For stochastic programming models with a continuous distribution on the random parameters we may expect smoother objective functions. This is illustrated by an analysis of the functions $z^{LP}(x)$ and $z^{I}(x)$ given above, where we now assume that $\mathbf{p}$ is uniformly distributed on the interval $[2, 2+\gamma]$ with $\gamma < 1$. Their graphs are given in Figure 5.3. The stochastic linear programming objective function is no longer piecewise linear but still convex.

The stochastic integer programming objective function is no longer discontinuous. The latter observation will be generalized in the following theorem.

Consider the stochastic program (5.8-11) with only $\mathbf{p}$ and $\mathbf{T}$ random. We define for each realization (p, T) of $(\mathbf{p}, \mathbf{T})$
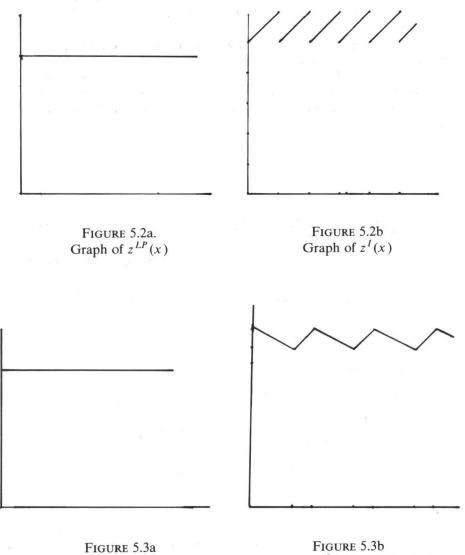
FIGURE 5.2a.
Graph of $z^{LP}(x)$

FIGURE 5.2b
Graph of $z^{I}(x)$

FIGURE 5.3a
Graph of $z^{LP}(x)(\mathbf{p}\sim U[2,2+\gamma])$

FIGURE 5.3b
Graph of $z^{I}(x)(\mathbf{p}\sim U[2,2+\gamma])$

$$Q^{I}(Tx+p)=\min\{qy\,|\,Wy\geqslant Tx+p,y\geqslant 0,y\in \mathbf{Z}^{k}\}.$$

Its expected value is given by

$$\mathscr{Q}^{I}(x)=EQ^{I}(\mathbf{T}x+\mathbf{p}).$$

Let $f(p,T)$ be the joint density function of $\mathbf{p}$ and $\mathbf{T}$. Let $g(T)$ be the marginal density function of $\mathbf{T}$ and let $h(p|T)$ be the conditional density function of $\mathbf{p}$

given T.

THEOREM 5.1. *If $f(p,T)$ is continuous with support (Π,Ξ) and $Q^I(Tx+p)$ is bounded with probability 1, then $\mathcal{Q}^I(x)$ is continuous.*

PROOF. We will prove that for every $\epsilon>0$ there exists a δ such that for any pair of vectors x and x' satisfying $\|x-x'\|_\infty \leq \delta$, we have that $|\mathcal{Q}^I(x)-\mathcal{Q}^I(x')|\leq\epsilon$. Let ι be the n-dimensional vector consisting of all onces. For a fixed value $\epsilon>0$ we choose any pair x and x' such that $x'=x+\delta\iota$, for a value of δ that is to be determined later. For given p and T it is easy to see that

$$Q^I(Tx'+p)=Q^I(Tx+p+\delta T\iota).$$

Therefore

$$|\mathcal{Q}^I(x')-\mathcal{Q}^I(x)| =$$

$$|\int_\Xi \int_\Pi Q^I(Tx+p+\delta T\iota)f(p,T)dpdT - \int_\Xi \int_\Pi Q^I(Tx+p)f(p,T)dpdT|$$

We substitute $p'-\delta T\iota$ for p in the first term of the right hand side of the above equality. This yields

$$|\mathcal{Q}^I(x')-\mathcal{Q}^I(x)|=|\int_\Xi \int_{\Pi'} Q^I(Tx+p')f(p'-\delta T\iota,T)dp'dT -$$

$$\int_\Xi \int_\Pi Q^I(Tx+p)f(p,T)dpdT|,$$

where $\Pi'=\Pi+\delta T\iota$. We write p instead of p' again:

$$|\mathcal{Q}^I(x)-\mathcal{Q}^I(x)|=|\int_\Xi \int_{\Pi'\setminus\Pi} Q^I(Tx+p)f(p-\delta T\iota,T)dpdT +$$

$$\int_\Xi \int_{\Pi'\cap\Pi} Q^I(Tx+p)f(p-\delta T\iota,T)dpdT -$$

$$\int_\Xi \int_{\Pi\cap\Pi'} Q^I(Tx+p)f(p,T)dpdT -$$

$$\int_\Xi \int_{\Pi\setminus\Pi'} Q^I(Tx+p)f(p,T)dpdT|$$

$$\leq|\int_\Xi \int_{\Pi'\setminus\Pi} Q^I(Tx+p)f(p-\delta T\iota,T)dpdT|+$$

$$|\int_\Xi \int_{\Pi\setminus\Pi'} Q^I(Tx+p)f(p,T)dpdT|+$$

$$|\int_\Xi \int_{\Pi\cap\Pi'} Q^I(Tx+p)(f(p-\delta T\iota,T)-f(p,T))dpdT|$$

By assumption, for each x there exists a finite L such that $Q^I(Tx+p)\leq L$ for almost every realization p,T of $(\mathbf{p},\mathbf{T})$ Therefore the above absolute difference is bounded from above by

$$L\int\limits_{\Xi}\int\limits_{\Pi'\setminus\Pi} f(p-\delta T\iota,T)dpdT + L\int\limits_{\Xi}\int\limits_{\Pi\setminus\Pi'} f(p,T)dpdT \qquad (5.14)$$

$$+ L|\int\limits_{\Xi}\int\limits_{\Pi\cap\Pi'} (f(p-\delta T\iota,T)-f(p,T))dpdT|$$

We consider each of the above three terms separately. We rewrite the first term as

$$L\int\limits_{\Xi} g(t)\int\limits_{\Pi'\setminus\Pi} h(p-\delta T\iota|T)dpdT.$$

Because under our assumption that $f(p,t)$ is continuous, $g(p|T)$ is continuous and we have that for each $\epsilon>0$ we can choose $\delta_1(\epsilon)$ small enough such that

$$\int\limits_{\Pi'\setminus\Pi} h(p-\delta_1(\epsilon)T\iota|T)dp \leqslant \frac{\epsilon}{3L}$$

and hence

$$L\int\limits_{\Xi} g(T)\int\limits_{\Pi'\setminus\Pi} h(p-\delta_1(\epsilon)T\iota|T)dpdT \leqslant \frac{\epsilon}{3}. \qquad (5.15)$$

In the same way we can for each $\epsilon>0$ choose $\delta_2(\epsilon)$ such that for $\Pi'=\Pi+\delta_2(\epsilon)T\iota$.

$$L\int\limits_{\Xi}\int\limits_{\Pi\setminus\Pi'} f(p,T)dpdT \leqslant \frac{\epsilon}{3}. \qquad (5.16)$$

As f is a continuous function of p, for each $\epsilon>0$ there exists a $\delta_3(\epsilon)$ such that

$$|f(p-\delta_3(\epsilon)T\iota,T)-f(p,T)| \leqslant \frac{\epsilon}{3L}.$$

Hence

$$L|\int\limits_{\Xi}\int\limits_{\Pi\cap\Pi'} (f(p-\delta_3(\epsilon)T\iota,T)-f(p,T))dpdT| < L\int\limits_{\Xi}\int\limits_{\Pi\cap\Pi'} \frac{\epsilon}{3L}dpdT \leqslant \frac{\epsilon}{3}. \qquad (5.17)$$

If we choose $\delta=\min\{\delta_1(\epsilon),\delta_2(\epsilon),\delta_3(\epsilon)\}$ then (5.14), (5.15), (5.16) and (5.17) together imply the theorem. $\square$

We note that the boundedness condition on $Q^I(Tx+p)$ is a rather strong requirement in integer programming. However, if the density function $f(p,T)$ has bounded support, then the condition is met for problems with all coefficients positive.

As we have seen in Figure 5.3b continuous objective functions of stochastic integer programming problems are not necessarily convex. The success of algorithms for stochastic linear programs is partially due to the convexity of their objective functions. If integrality constraints appear only at the first stage of a stochastic program, then the expected optimal second stage cost is still convex in the first stage decision variables and the problem can be dealt with by fairly conventional adaptations of stochastic linear programming

methods (see e.g. [Wollmer 1981]). The nonconvexities induced by integrality constraints at the second stage cause more fundamental problems. It is not al all obvious how these difficulties can be dealt with. In the following section, we give some directions for future investigations that might lead to practically useful stochastic integer programming methods.

5.2. Directions for future research

In the previous section we examined the shape of objective functions of stochastic integer programming problems without making attempts to use the insights acquired for the construction of an algorithm. This is mainly due to the fact that the nonconvexity of the functions does not lead to a natural proposal for such an algorithm. To imitate solution procedures that are successful for stochastic linear programs requires more than minor modifications because of the integrality constraints on the variables.

One possibility may be to imitate the *L-shaped method* for stochastic linear programming, which is based on Benders' decomposition [Wets 1983], by substitution of the stochastic integer programming objective function by a convex approximation. More specifically, if we consider a stochastic minimization problem and define the *epigraph* of the objective function as the set of all points lying above or on the function, then the minimum of the function is the lowest point of the epigraph (see Figure 5.4). This is also the lowest point of the convex hull of this epigraph. Therefore, minimization of the function can be replaced by minimization of the convex hull of its epigraph. Determination of this convex hull is not surprisingly, a nontrivial task. But a useful insight can be derived from [Blair & Jeroslow 1979]. They define the *carrier function* of an integer programming value function as the function that is obtained by deleting the round-up operators from its description. The epigraph of a carrier function is convex and contains the convex hull of the epigraph of the corresponding value function. It would be interesting to see if and how carrier functions can be used in a decomposition method to solve the stochastic integer programming problem.

Another possible approach for the optimization of multimodal objective functions is to use a *global optimization algorithm.* In the literature on global optimization, a variety of algorithms is presented (see [Timmer 1984] for a review). For reasons indicated many times before, we are interested in methods that require only a small number of function evaluations. For example, we may consider global optimization methods that represent the function to be optimized as a realization of a stochastic process (see [Boender 1984]).

A drawback of the latter type of algorithm is that it ignores most prior information about the shape of the function. For example, in some cases it may be possible to compute a Lipschitz constant for the objective function of a stochastic integer program, given the probability distribution of the random parameters and the values of the non-random parameters. In such cases one
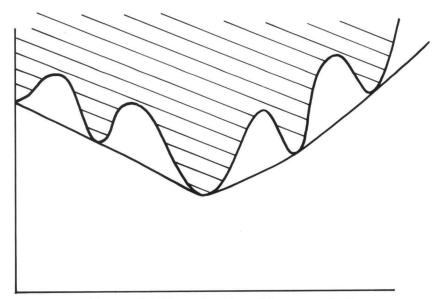
FIGURE 5.4. The epigraph and its convex hull

could use this information in a global optimization algorithm similar to the one presented in [Evtushenko 1971].

The construction of optimization methods for stochastic integer programming represents a formidable challenge. Notwithstanding the high asymptotic quality of the heuristics discussed in previous chapters, we hope that this challenge will be properly responded to by further research.

REFERENCES

R.J. ARMSTRONG, A.C. HAX (1977). A hierarchical approach for a naval-tender job-shop design. S.P. BRADLEY, A.C. HAX, T.L. MAGNANTI (eds.). *Applied mathematical programming*, Addison Wesley, Reading, Massachusetts, Ch. 10.

J. BEARDWOOD, H.J. HALTON, J.M. HAMMERSLEY (1959). The shortest path through many points. *Proc. Cambridge Phil. Soc. 55*, 299-327.

R.E. BELLMAN (1957). *Dynamic programming*, Princeton University Press, Princeton, New York.

P. BILLINGSLEY (1979). *Probability and measure*, John Wiley & Sons, New York.

C. BLAIR, R.G. JEROSLOW (1979). On the value function of an integer program. *Math. Programming 23*, 237-273.

C.G.E. BOENDER (1984). *The generalized multinomial distribution: a Bayesian analysis and applications*, Ph.D. thesis, Erasmus University, Rotterdam.

K.L. CHUNG (1949). An estimate concerning the Kolmogorov limit distribution. *Trans. Amer. Math. Soc. 67*, 36.

R.W. CONWAY, W.L. MAXWELL, L.W. MILLER (1967). *Theory of scheduling*, Addison Wesley, Reading, MA.

M.A.H. DEMPSTER, M.L. FISHER, L. JANSEN B.J. LAGEWEG, J.K. LENSTRA, A.H.G. RINNOOY KAN (1983). Analysis of heuristics for stochastic programming: results for hierarchical scheduling problems. *Math. Oper. Res. 8*, 525-537.

Y.P. EVTUSHENKO (1971). *Zh. Vychisl. Mat. i Mat. Fiz. 11*, 1390-1403.

W. FELLER (1968). *An introduction to probability theory and its applications, vol. 1, 3rd ed.*, John Wiley & Sons, New York.

L. FEW (1955). The shortest path and the shortest road through n points. *Mathematika 2*, 141-144.

M.L. FISHER, D.S. HOCHBAUM (1980). Probabilistic analysis of the planar K-median problem. *Math. Oper. Res. 5*, 27-34.

M.R. GAREY, D.S. JOHNSON (1979), *Computers and intractability: a guide to the theory of NP-completeness*, Freeman, San Francisco.

M. HAIMOVICH, A.H.G. RINNOOY KAN (to appear). Bounds and heuristics for capacitated routing problems. *Math. Oper. Res.*

P. HARTMAN, A. WINTNER (1941). On the law of the iterated logarithm. *Amer. J. Math. 63*, 169.

M. HELD, R.M. KARP (1962). A dynamic programming approach to sequencing problems. *SIAM J. 10*, 156-210.

M. HELD, R.M. KARP, R. SHARESHIAN (1963). Assembly-line balancing: dynamic programming with precedence constraints. *Oper. Res. 11*, 442-459.

D.S. HOCHBAUM, J.M. STEELE (1981). Steinhaus' geometric location problem for random samples in the plane. *Adv. in Appl. Probab. 14*, 56-67.

R.M. KARP (1972). Reducibility among combinatorial problems. R.E.

MILLER, J.W. THATCHER (eds.). *Complexity of computer computations,* Plenum, New York, 85-103.

R.M. KARP (1977). Probabilistic analysis of partitioning algorithms for the traveling-salesman problem in the plane. *Math. Oper. Res. 2,* 209-224.

R.M. KARP, J.M. STEELE (1985). Probabilistic analysis. E.L. LAWLER, J.K. LENSTRA, A.H.G. RINNOOY KAN, D.B. SHMOYS (eds.). *The traveling salesman problem: a guided tour of combinatorial optimization.* John Wiley & Sons, Chichester, England.

C.H. PAPADIMITRIOU (1981). Worst case and probabilistic analysis of a geometric location problem. *SIAM J. Comput. 10,* 542-557.

R.J. SERFLING (1980). *Approximation theorems of mathematical statistics,* John Wiley & Sons, New York.

J. SHWIMER (1972). *Interactions between aggregate and detailed scheduling in a job shop,* Ph. D. thesis, Sloan School of Management, Massachusetts Institute of Technology, MA.

J.M. STEELE (1981). Subadditive Euclidean functionals and nonlinear growth in geometric probability. *Ann. Probab. 9,* 365-376.

G.T. TIMMER (1984). *Global optimization: a stochastic approach,* Ph.D. thesis, Erasmus University, Rotterdam.

R.J.B. WETS (1983). Stochastic programming: solution techniques and approximation schemes. A. BACHEM, M. GRÖTSCHEL, B. KORTE (eds.). *Mathematical programming: the state of the art - Bonn 1982,* Springer, Berlin, 566-603.

R.M. WOLLMER (1980). Two-stage linear programming under uncertainty with 0-1 integer first stage variables. *Math. Programming 19,* 279-288.

E. ZEMEL (1984). Probabilistic analysis of geometric location problems. *Ann. Oper. Res. 1.*

Samenvatting

Stochastische programmeringsproblemen zijn mathematische programmeringsproblemen waarvan sommige parameters als kansvariabelen gemodelleerd worden om de onzekerheid over hun waarde weer te geven. Voor de onbekende parameters is gewoonlijk een kansverdeling gespecificeerd. Er zijn twee soorten optimaliseringsmodellen in de stochastische programmering.

Het eerste is het distributiemodel, waarbij een optimale beslissing voor elke realisatie van de stochastische parameters wordt gezocht. Dit geeft als resultaat de kansverdeling van de optimale waarde van het stochastische programmeringsprobleem. Deze benadering, waarbij de beslissing genomen wordt zodra perfecte informatie beschikbaar is, is voornamelijk van theoretisch belang.

Het tweede model is het twee-fasen beslissingsmodel, waarbij we een beslissing zoeken die in verwachting optimaal is. Bij de waardering van een beslissing houden we rekening met de kosten van een latere beslissing die genomen kan worden om de eerste bij te stellen zodra een realisatie van de kansvariabelen bekend wordt. In dit model wordt de eerste beslissing genomen gegeven imperfecte informatie. We kunnen de bijstelling van de beslissing ruim zien. Het betreft niet alleen een noodzakelijke actie als de realisatie ongelukkig uitvalt, maar elke mogelijke actie die onder de gegeven omstandigheden tot een in totaal betere beslissing leidt.

Het onderzoek op dit gebied heeft zich tot nu toe bijna volledig geconcentreerd op stochastische lineaire programmering. In het distributiemodel leidt elke realisatie van de stochastische parameters in dit geval tot een gewoon lineair programmeringsprobleem. In het twee-fasen beslissingsmodel, is het bijstellingsprobleem in de tweede fase een lineair programmeringsprobleem, waarbij de rechterzijde van de restricties gewoonlijk bepaald wordt door de beslissing uit de eerste fase. Dit impliceert dat de verwachte kosten van de tweede fase een convexe functie zijn van de beslissingsvariabelen uit de eerste fase. Succesvolle algoritmen voor het lineaire twee-fasen probleem maken uitvoerig gebruik van deze convexiteit.

Door geheeltalligheidseisen aan sommige van de beslissingsvariabelen op te leggen betreden we het gebied van de stochastische geheeltallige programmering. De complexiteit van de te bestuderen problemen wordt hierbij drastisch verhoogd. Eén reden hiervoor is dat er voor de oplossing van lineaire programmeringsproblemen methoden bestaan die op theoretische gronden efficiënt genoemd kunnen worden, terwijl dergelijke methoden voor

geheeltallige programmering ontbreken. Een andere reden is dat, in het twee-fasen beslissingsmodel, geheeltalligheid van de beslissingsvariabelen in de tweede fase niet-convexiteiten en zelfs discontinuïteiten in de verwachte kosten in de tweede fase kan veroorzaken, zodat de de algoritmen voor het lineaire geval niet eenvoudig aangepast kunnen worden.

Stochastische geheeltallige programmering biedt niet alleen veel theoretische uitdagingen. Het is ook een practisch hulpmiddel voor het modelleren van hierarchische planningsituaties die voorkomen bij operationele beheersactiviteiten. Dergelijke situaties vragen om een reeks beslissingen over de tijd, die in toenemende mate gedetailleerd zijn en waarbij steeds meer informatie beschikbaar komt. Er kunnen tenminste twee beslissingsvineaus onderscheiden worden: een geaggregeerd niveau, waarop beslissingen genomen moeten worden over de aanschaf van hulpbronnen, zoals machines, vrachtwagens of mankracht, zonder dat er exacte informatie over het toekomstige gebruik ervan aanwezig is, en een gedetailleerd niveau, waarop, gegeven deze exacte informatie, taken aan de hulpbronnen toegewezen worden. Geheeltalligheidseisen kunnen op het eerste niveau voorkomen als de hulpbronnen alleen in gehele aantallen aangeschaft kunnen worden, en op het tweede niveau als het toewijzingsprobleem combinatorisch van aard is.

Vanwege de inherente moeilijkheid van stochastische geheeltallige programmering lijkt het redelijk om niet optimaliteit van de beslissingen te eisen maar ons tevreden te stellen met benaderingen hiervan. Daarom is een groot deel van het onderzoek in dit proefschrift gericht op het ontwerpen en analyseren van benaderingsmethoden of 'heuristieken'. Zo worden eenvoudige heuristieken voorgesteld voor de oplossing van verscheidene twee-fasen productie- en distributieplanningproblemen. Probabilistische analyses van deze heuristieken leiden dan tot exacte uitspraken over de kwaliteit van de benaderingen. Bijvoorbeeld kunnen we afleiden dat voor voldoend grote problemen de relatieve fout die de heuristiek maakt in een bepaalde stochastische zin heel klein wordt. Voor het ontwerpen en analyseren van de heuristieken wordt een algemeen kader gepresenteerd.

Voor het 'optimaal' oplossen van stochastische geheeltallige programmeringsproblemen moeten we ons beperken tot problemen met een dusdanig speciale structuur dat exploitatie hiervan kan leiden tot een efficiënte optimaliseringsalgoritme. Zo beperken we ons voor een aantal van de bovengenoemde hierarchische planningproblemen tot situaties vaarbij de stochastische parameters een discrete verdeling hebben met de kansmassa geconcentreerd op een eindig aantal punten. Elke realisatie van deze problemen leidt tot een combinatorisch optimaliseringsprobleem, dat oplosbaar is met behulp van een dynamische programmeringsalgoritme. Het gehele stochastische programmeringsprobleem wordt dan opgelost in één grote recursie, waarin de aparte dynamische programmeringsberekeningen voor alle individuele realisaties gecombineerd worden. Computerexperimenten hiermee geven enig empirisch inzicht in de doelstellingsfuncties van stochastische geheeltallige programmeringsproblemen.

Verdere onderzoekingen naar de vorm van de doestellingsfunctie tonen aan dat een onderscheid tussen discrete en continue kansverdelingen van de stochastische parameters gemaakt dient te worden. In het laatste geval komen er, onder bepaalde voorwaarden, geen discontinuïteiten voor. De resultaten die we hier afleiden voor de vorm van de doelstellingsfuncties moeten beschouwd worden als niet meer dan een eerste stap in de richting van een theorie voor stochastische geheeltallige programmering.

Definieer de optimale waarde van een meer-dimensionaal knapzakprobleem als

$$z_n = \max\{\textstyle\sum_{j=1}^n c_j x_j \mid \sum_{j=1}^n a_{ij} x_{ij} \leqslant n b_i \, (i=1,...,m), x_j \in \{0,1\} \, (j=1,...,n)\}$$

waarbij $0 \leqslant b_i \leqslant 1$, $c_j \geqslant 0$, $a_{ij} \geqslant 0$ $(i=1,...,m; j=1,...,n)$. Beschouw een stochastisch model van dit probleem waarbij de coëfficiënten $\mathbf{c}_j$ en $\mathbf{a}_{ij}$ onafhankelijke identiek verdeelde kansvariabelen zijn. Dan bestaat er een constante L zodat met kans 1

$$\lim_{n \to \infty} \frac{1}{n} \mathbf{z}_n = L.$$

Deze constante kan berekend worden als de minimumwaarde van de volgende convexe functie:

$$L(\lambda_1,...,\lambda_m) = \textstyle\sum_{i=1}^m \lambda_i b_i + E \mathbf{c}_1 \mathbf{x}_1(\lambda) - \sum_{i=1}^m \lambda_i E \mathbf{a}_{i1} \mathbf{x}_1(\lambda),$$

waarbij

$$\mathbf{x}_1(\lambda) = \begin{cases} 1 & \text{als } \mathbf{c}_1 - \sum_{i=1}^m \lambda_i \mathbf{a}_{i1} \geqslant 0, \\ 0 & \text{anders.} \end{cases}$$

M. MEANTI, A.H.G. RINNOOY KAN, L. STOUGIE, C. VERCELLIS (1984). *A probabilistic analysis of the multi-knapsack value function,* Working paper #1611-84, Massachusetts Institute of Technology, Cambridge, MA.

We kunnen de optimale opllossing z_n van het meer-dimensionale knapzakprobleem uit de vorige stelling benaderen met behulp van een 'greedy' heuristiek die als volgt beschreven kan worden. Gegeven m gewichten $w_1,...,w_m$, rangschikken we de objecten $j=1,...,n$ naar afnemende waarden van de verhoudingen

$$\rho_j = \frac{c_j}{\sum_{i=1}^m w_i a_{ij}}.$$

Objecten worden dan in deze volgorde geselecteerd totdat het volgende te selecteren object één van de restricties zou schenden als het aan de knapzak toegevoegd zou worden. Wanneer we de gewichten gelijkstellen aan de waarden $\lambda_1,...,\lambda_m$ die de functie $L(\lambda)$ uit de vorige stelling minimaliseren, dan gaat de relatieve fout die de heuristiek maakt bijna zeker naar nul.

M. Meanti, A.H.G. Rinnooy Kan, L. Stougie, C. Vercellis (1984). *A probabilistic analysis of the multi-knapsack value function*, Working paper #1611-84, Massachusetts Institute of Technology, Cambridge, MA.

III

Gegeven een aantal aardappelrassen, een aantal aardappelziektes en, voor elk van de rassen, de verzameling ziektes waarvoor het vatbaar is, selecteer een zo klein mogelijk aantal aardappelrassen zodanig dat elk paar ziektes door ten minste één aardappelras uit de selectie onderscheiden wordt.

Als elk aardappelras voor precies twee ziektes vatbaar is, dan kan het probleem als volgt gemodelleerd worden. Elke aardappelziekte correspondeert met een punt in een graaf en elk aardappelras met een kant in die graaf. De doelstelling is dan om in de graaf zoveel mogelijk disjuncte paden van lengte 2 te vinden. Dit probleem is NP-moeilijk [Lageweg et al. 1980].

B.J. Lageweg, J.K. Lenstra, A.H.G. Rinnooy Kan (1980). Uit de practijk van de besliskunde. A.K. Lenstra, H.W. Lenstra, J.K. Lenstra (eds.). *Tamelijk briljant: opstellen aangeboden aan Dr. T.J. Wansbeek,* Amsterdam.

IV

Voor de oplossing van het probleem uit de vorige stelling formuleren we de volgende heuristiek. Kies eerst zoveel mogelijk disjuncte paden van lengte 2, en verbind vervolgens de resterende losse punten en paden van lengte 1 aan de gekozen paden van lengte 2. Op deze manier wordt een verzameling aardappelrassen geselecteerd die in het slechtste geval $\frac{4}{3}$ keer zo groot is als een optimale verzameling aardappelrassen.

Deze oplossing kan verbeterd worden door onderling disjuncte groepen van k samenhangende configuraties uit de oplossing te beschouwen en binnen de verkregen deelgrafen de optimale oplossingen te bepalen. In het geval $k=2$ is de verhouding tussen de heuristische en de optimale oplossing in het slechtste geval $\frac{5}{4}$. Voor $k=3$ is deze verhouding $\frac{6}{5}$ [Lageweg et al.].

B.J. Lageweg, J.K. Lenstra, J.B. Orlin, L. Stougie (te verschijnen). *The minimum test set problem.*

Aan de verzameling testfuncties in [Dixon & Szegö 1978], die gebruikt worden voor het vergelijken van globale optimaliseringsprocedures (cf. [Boender et al. 1982]), dient een functie te worden toegevoegd waarvoor de kans zeer klein is dat een lokale zoekprocedure gestart in een willekeurig getrokken punt uit het domein van de functie naar het globale optimum convergeert.

C.G.E. BOENDER, A.H.G. RINNOOY KAN, L. STOUGIE, G.T. TIMMER (1982). A stochastic method for global optimization. *Math. Programming 22, 125-140.*
L.C.W. DIXON, G.P. SZEGÖ (1978). *Towards global optimisation 2,* North-Holland, Amsterdam.

De elfstedentocht beschrijft de kortste gesloten route door de elf stempelsteden.

E.L. LAWLER, J.K. LENSTRA, A.H.G. RINNOOY KAN, D.B. SHMOYS (1985). *The traveling saleman problem: a guided tour of combinatorial optimization,* John Wiley & Sons, Chichester, England.

De schade die wilde ganzen in Nederland aan landbouwgewassen en grasland aanrichten zal verminderen als de jacht op deze vogels wordt afgeschaft.

Het is verwerpelijk om de bezinkingsput van de voormalige suikerfabriek in Oud-Beijerland met recreatieve plannen te bedreigen. De put heeft een volkomen miskende ornithologisch educatieve waarde. Bijna alle vogelsoorten van het noordelijke Deltagebied komen er voor. Bovendien wordt de observatie van de vogels vergemakkelijkt doordat de put dichtbij de bebouwde kom ligt en omringd wordt door een dijk.

L. STOUGIE, M. STOUGIE. *Vogelwaarnemingen,* Admiraal de Ruyterstraat 143, Oud-Beijerland.

Bij de consumptie van stroop wordt er helaas meestal geen weloverwogen keuze gemaakt tussen de verschillende merken. Op spekpannekoeken verdienen de variëteiten met een zachte smaak de voorkeur, terwijl de soorten met een branderige smaak juist tot hun recht komen op een boterham met boter. Voor wie zijn hele huishouden met één pot stroop wenst te doen is de echte Zeeuwse stroop de gulden middenweg.