stichting

mathematisch

centrum

$\sum$
MC

J.K. LENSTRA, A.H.G. RINNOOY KAN, L. STOUGIE

A FRAMEWORK FOR THE PROBABILISTIC ANALYSIS
OF HIERARCHICAL PLANNING SYSTEMS

Prepublication

A FRAMEWORK FOR THE PROBABILISTIC ANALYSIS
OF HIERARCHICAL PLANNING SYSTEMS

J.K. LENSTRA

Mathematisch Centrum, Amsterdam

A.H.G. RINNOOY KAN

Erasmus University, Rotterdam

L. STOUGIE

Mathematisch Centrum, Amsterdam

ABSTRACT

As we have argued in previous papers, multi-level decision problems can often be modeled as multi-stage stochastic programs, and hierarchical planning systems designed for their solution, when viewed as stochastic programming heuristics, can be subjected to analytical performance evaluation. The present paper gives a general formulation of such stochastic programs and provides a framework for the design and analysis of heuristics for their solution. The various ways to measure the performance of such heuristics are reviewed, and some relations between these measures are derived. Our concepts are illustrated on a simple two-level planning problem of a general nature and on a more complicated two-level scheduling problem.

1. INTRODUCTION

Many practical situations in operations management planning and control require a series of decisions over time at an increasing level of detail and with an increasing amount of information being available. At least two decision stages can usually be recognized: an *aggregate* level, at which one has to decide upon the acquisition of resources, given vague information about what certain tasks will require of them, and a *detailed* level, at which one has to decide upon the allocation of resources to tasks, given precise information about the requirements. In *job shop* design and scheduling, for example, the resources are machines and the tasks are jobs, whose processing times can only be roughly estimated at the outset. In *distribution system* design and control, the resources are vehicles and the tasks are deliveries to clients, whereby the locations of the clients demanding service are subject to stochastic fluctuations. The challenge of such *hierarchical planning problems* is to incorporate the initially imperfect detailed 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 multi-level decision problems is to design a *hierarchical planning system* [Bitran & Hax 1977, Bradley et al. 1977]. In such a system, each level is modeled as a separate deterministic optimization problem. The resulting series of linked mathematical programming models is then evaluated by means of simulation techniques.

In [Dempster et al. 1981], it was argued that hierarchical planning problems can be naturally formulated as *multi-stage stochastic programs*. Corresponding to each decision level, there is a stage that incorporates probabilistic information about the later stages and that aims at setting the decision variables in such a way that the overall result is, in some sense, optimal. In the examples quoted above, the scheduling and routing problems appearing at the second level are NP-hard combinatorial optimization problems. Apart from that, the stochastic optimization problem at the aggregate level generally represents an even more formidable computational challenge. Thus, one should design approximation algorithms or *heuristics*

for multi-stage stochastic programming, and hierarchical planning systems are essentially nothing but that. This observation may seem obvious, but the stochastic programming formulation of hierarchical planning problems provides a proper framework for an analytical rather than empirical evaluation of the performance of heuristics designed for their solution. Indeed, in [Dempster et al. 1983, Marchetti Spaccamela et al. 1982, Frenk et al. 1983] exact statements about the behavior of hierarchical scheduling and routing problems have been 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 have been derived are similar. Also, the hierarchical planning systems constructed have many features in common. The purpose of this paper is to outline a general approach to the design and analysis of hierarchical planning systems.

In Section 2 we will formulate stochastic programming models for a hierarchical planning problem with two decision levels and indicate how to construct heuristics for its solution. We will also review the various ways to measure the performance of such heuristics and exhibit some relations between these measures. In Section 3 we will illustrate our concepts, first on a simple two-level planning problem of a general nature and finally on a more specific and more complicated two-level scheduling problem.

2. MODELS, HEURISTICS, AND PERFORMANCE MEASURES

2.1. Stochastic programming models

Consider the typical two-stage decision situation outlined in the first paragraph of Section 1.

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 X, and the direct cost associated with X by $f(X)$, where $f: X \to \mathbb{R}$ is a real function. Probabilistic information about future resource requirements is represented by a vector ω; it comes from a sample space Ω and has distribution function F. (We write a tilde under a variable

to indicate that it is a random variable.)

The input to the detailed level consists of the first stage decision X and a realization ω of the random vector $\underset{\sim}{\omega}$. The objective at the second stage is to decide upon a certain allocation of the resources acquired so as to minimize a cost $g(X,\omega)$, where $g: X \times \Omega \to \mathbb{R}$ is a real function. The optimal value of $g(X,\omega)$ will be denoted by $g^*(X,\omega)$, and the total cost of the acquisition decision X and the optimal allocation decision by $z^*(X,\omega) = f(X)+g^*(X,\omega)$.

To complete the formulation of the *two-stage decision* model, we define the objective at the first stage: determine an $X^* \in X$ such that the *expected* total cost $Ez^*(X,\underset{\sim}{\omega}) = f(X)+Eg^*(X,\underset{\sim}{\omega})$ is minimized:

$$Ez^*(X^*,\underset{\sim}{\omega}) = \min_{X \in X}\{Ez^*(X,\underset{\sim}{\omega})\}.$$

According to stochastic programming terminology, the first stage decision is made "here and now", given *imperfect* information about the second stage.

As an alternative, we formulate the so-called *distribution* model with an overall objective: determine a function $X^{\circ}: \Omega \to X$ such that for each $\omega \in \Omega$ the *actual* total cost is minimized:

$$z^*(X^{\circ}(\omega),\omega) = \min_{X \in X}\{z^*(X,\omega)\}.$$

Before the aggregate decision is taken, we "wait and see" until *perfect* information about the second stage is available.

2.2. Two-stage stochastic programming heuristics

There is little hope to develop practical optimization algorithms for the above stochastic programs. As to the two-stage decision model, the determination of $g^*(X,\omega)$ is often an NP-hard problem, so that a heuristic must be used at the second stage. Even if $g^*(X,\omega)$ can easily be determined, it seems impossible to obtain a tractable representation of $Eg^*(X,\underset{\sim}{\omega})$, 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 $\mathrm{E}g^*(X,\omega)$ by an estimate $g^{H_1}(X)$ and determine an $X^{H_1} \in X$ such that the estimated total cost $z^{H_1}(X) = f(X) + g^{H_1}(X)$ is minimized:

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

In some cases, even this approximate first stage problem is NP-hard and another heuristic device is needed to solve it (cf. Section 3.2).

At the second stage, we allocate the resources acquired, achieving an approximate cost $g^{H_2}(X^{H_1},\omega)$. 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 $\omega \in \Omega$ will be denoted by $z^{H_2}(X^{H_1},\omega) = f(X^{H_1}) + g^{H_2}(X^{H_1},\omega)$.

The success of this heuristic approach evidently depends on the quality of $g^{H_1}(X)$ as an estimate of $\mathrm{E}g^*(X,\omega)$ and of $g^{H_2}(X,\omega)$ as an approximation of $g^*(X,\omega)$.

In [Dempster *et al.* 1983, Marchetti Spaccamela *et al.* 1982, Frenk *et al.* 1983], the first stage heuristic is typically based on a *lower bound* on $g^*(X,\omega)$; the second stage heuristic obviously yields an *upper bound* on $g^*(X,\omega)$. The purpose of the analysis is then to show that the underestimate $z^{H_1}(X^{H_1})$, the optima $z^*(X^*,\omega)$ and $z^*(X^o(\omega),\omega)$ and the approximation $z^{H_2}(X^{H_1},\omega)$ are asymptotically equal in some probabilistic sense. In this context, good use can be made of probabilistic characterizations of optimal solution values to combinatorial problems, such as routing problems [Beardwood *et al.* 1959, Steele 1981] and location problems [Fisher & Hochbaum 1980, Hochbaum & Steele 1981, Papadimitriou 1981, Zemel 1983].

2.3. 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 $\chi_1, \chi_2, \ldots$ is said to *converge to a random variable* χ

(a) *in expectation* if $\lim_{n \to \infty} \mathrm{E}\chi_n = \mathrm{E}\chi$
     [notation: $\mathrm{E}\chi_n \to \mathrm{E}\chi$];

(b)   *in probability* if $\lim_{n\to\infty} \Pr\{|\underset{\sim}{x}_n - \underset{\sim}{x}| \le \varepsilon\} = 1$ for every $\varepsilon > 0$

[notation: $\underset{\sim}{x}_n \to \underset{\sim}{x}$ (ip)];

(c)   *with probability* 1 or *almost surely* if $\Pr\{\lim_{n\to\infty} \underset{\sim}{x}_n = \underset{\sim}{x}\} = 1$

[notation: $\underset{\sim}{x}_n \to \underset{\sim}{x}$ (wp1)].

Some well-known relations between these types of convergence are given in Section 2.4.

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.

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

$$\frac{E z^{H_2}(x^{H_1}, \underset{\sim}{\omega})}{E z^*(x^*, \underset{\sim}{\omega})} \ .$$

If this 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 $\varepsilon > 0$ and has the property that, for each ε, the ratio tends to a number less than $1 + \varepsilon$, then $(H_1(\varepsilon), H_2(\varepsilon))$ 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

$$\frac{z^{H_2}(x^{H_1}, \underset{\sim}{\omega})}{z^*(x^*, \underset{\sim}{\omega})} \quad \text{and} \quad \frac{z^{H_2}(x^{H_1}, \underset{\sim}{\omega})}{z^*(x^o(\underset{\sim}{\omega}), \underset{\sim}{\omega})} \ .$$

If the first of both ratios tends to 1 (or, for each $\varepsilon > 0$, to a number less than $1+\varepsilon$) 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, also the relative loss caused by imperfect information disappears in the limit.

Still other measures are based on a comparison of the aggregate deci-
sions X^{H_1}, X^* and $X^{\circ}(\underset{\sim}{\omega})$. In case X is a set of numbers, one can directly in-
vestigate the limiting behavior of the ratios

$$\frac{X^{H_1}}{X^*} \quad \text{and} \quad \frac{X^{H_1}}{X^{\circ}(\underset{\sim}{\omega})}$$

(cf. Section 3.1). The first of these ratios is a deterministic variable,
but the second one is random and its convergence analysis results in proba-
bilistic statements. Sometimes it may even be possible to obtain good bounds
on the differences $X^{H_1}-X^*$ and $X^{H_1}-X^{\circ}(\underset{\sim}{\omega})$. In case 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 dif-
ferences between $W(X^{H_1})$, $W(X^*)$ and $W(X^{\circ}(\underset{\sim}{\omega}))$ (cf. Section 3.2).

2.4. Relations between performance measures

Lemmas 1 and 2 give fundamental relations between the three types of conver-
gence of a sequence of random variables $\underset{\sim}{y}_1, \underset{\sim}{y}_2, \ldots$ to a random variable $\underset{\sim}{y}$.
We refer to [Serfling 1980] for proofs and for examples which show that the
inverse implications do not hold in general.

LEMMA 1. $\underset{\sim}{y}_n \rightarrow \underset{\sim}{y}$ (wp1) $\Rightarrow \underset{\sim}{y}_n \rightarrow \underset{\sim}{y}$ (ip). $\square$

LEMMA 2. *Suppose* $\underset{\sim}{y}_1, \underset{\sim}{y}_2, \ldots$ *is uniformly bounded (wp1), i.e., there exists a*
constant c such that for each n all realizations y_n *of* $\underset{\sim}{y}_n$ *satisfy* $|y_n| < c$
except for a set of realizations with probability measure 0. Then
$\underset{\sim}{y}_n \rightarrow \underset{\sim}{y}$ (ip) $\Rightarrow E\underset{\sim}{y}_n \rightarrow E\underset{\sim}{y}$. $\square$

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

$$\underset{\sim}{z}^{H} \text{ for } \underset{\sim}{z}^{H_2}(X^{H_1}, \underset{\sim}{\omega}), \quad \underset{\sim}{z}^* \text{ for } \underset{\sim}{z}^*(X^*, \underset{\sim}{\omega}), \quad \underset{\sim}{z}^{\circ} \text{ for } \underset{\sim}{z}^*(X^{\circ}(\underset{\sim}{\omega}), \underset{\sim}{\omega}).$$

To simplify the analysis, we make the following assumptions.

ASSUMPTION 1. *There exists a constant* $c_1 > 0$ *such that* $\tilde{z}^H/\tilde{z}^o < c_1$ (wp1).

ASSUMPTION 2. *There exists a constant* $c_2 > 0$ *such that* $\tilde{z}^*/E\tilde{z}^* < c_2$ (wp1).

ASSUMPTION 3. *There exists a constant* $c_3 > 0$ *such that* $\tilde{z}^*/E\tilde{z}^* > c_3$ (wp1).

ASSUMPTION 4. *There exists a constant* $c_4 > 0$ *such that* $\tilde{z}^o/E\tilde{z}^o > c_4$ (wp1).

In addition to these assumptions, we will use the basic properties of our models that $\tilde{z}^H \geq \tilde{z}^o$, $\tilde{z}^* \geq \tilde{z}^o$, $\tilde{z}^o > 0$ and $E\tilde{z}^H \geq E\tilde{z}^*$; but it need not be true that $\tilde{z}^H \geq \tilde{z}^*$ for every $\omega \in \Omega$. Note that under Assumption 1 also $\tilde{z}^H/\tilde{z}^* < c_1$, $\tilde{z}^H/E\tilde{z}^* < c_1$, $\tilde{z}^H/E\tilde{z}^o < c_1$, $\tilde{z}^o/E\tilde{z}^o < c_1$ (wp1). We will return in Section 3 to the question to what extent our assumptions are realistic in the applications we are considering.

Figure 1 shows which relations hold under these assumptions, and which do not. We will first illustrate some of the invalid implications by means of three examples, and next prove the valid implications in three theorems.
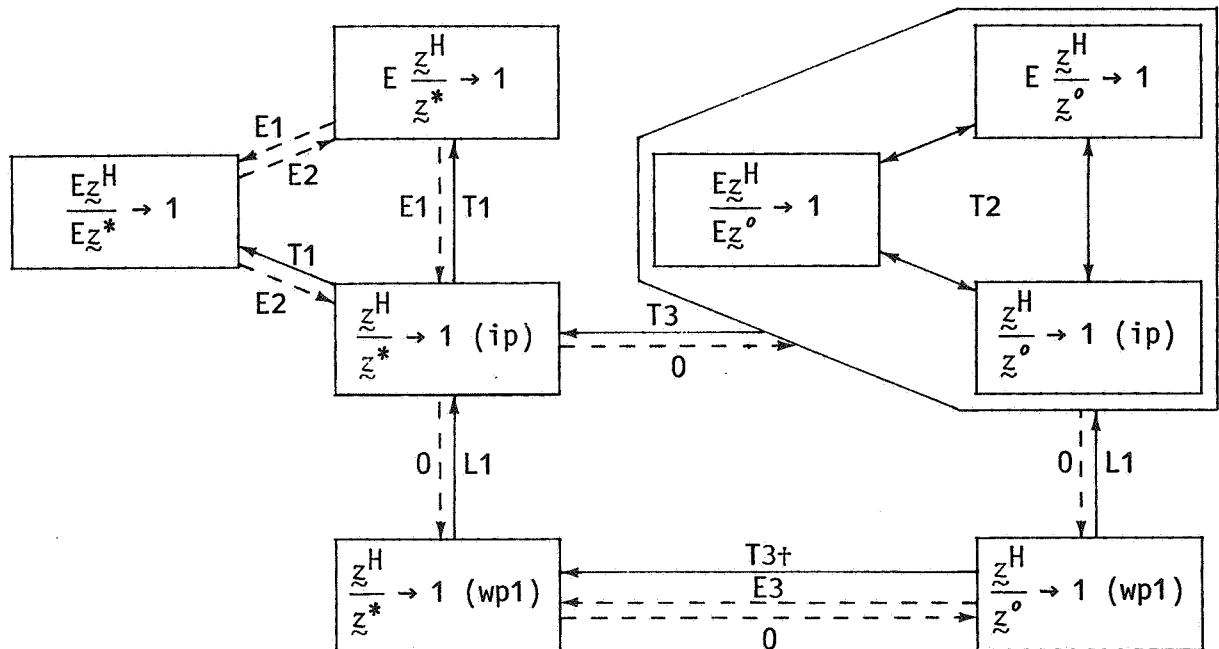

Figure 1. Relations between performance measures.

―――――→: valid implication; - - - →: invalid implication;

O: Obvious; E: Example; L: Lemma; T: Theorem;

†: if $\tilde{z}^H/\tilde{z}^*$ has a finite limit (wp1).

8

Example 1. $E(z^H/z^*) \to 1$ but $Ez^H/Ez^* \not\to 1$ and $z^H/z^* \not\to 1$ (ip).

Let $\Omega' \subset \Omega$ with $\Pr\{\omega \in \Omega'\} = \frac{1}{2}$. Define $z^H = 1$, $z^* = 2$ for $\omega \in \Omega'$ and $z^H = 6$, $z^* = 4$ for $\omega \notin \Omega'$. We have $E(z^H/z^*) = 1$ but $Ez^H/Ez^* = \frac{7}{6}$ and $z^H/z^* \in \{\frac{1}{2}, \frac{3}{2}\}$ for all $\omega \in \Omega$. $\square$

Example 2. $Ez^H/Ez^* \to 1$ but $E(z^H/z^*) \not\to 1$ and $z^H/z^* \not\to 1$ (ip).

Let $\Omega' \subset \Omega$ with $\Pr\{\omega \in \Omega'\} = \frac{1}{2}$. Define $z^H = 1$, $z^* = 2$ for $\omega \in \Omega'$ and $z^H = 2$, $z^* = 1$ for $\omega \notin \Omega'$. We have $Ez^H/Ez^* = 1$ but $E(z^H/z^*) = \frac{5}{4}$ and $z^H/z^* \in \{\frac{1}{2}, 2\}$ for all $\omega \in \Omega$. $\square$

Example 3. $z^H/z^\circ \to 1$ (wp1) but $z^H/z^* \not\to 1$ (wp1).

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

$$z_n^H = \begin{cases} 2 & (\omega \in I_n^H) \\ 1 & (\omega \notin I_n^H) \end{cases}, \quad z_n^* = \begin{cases} 2 & (\omega \in I_n^*) \\ 1 & (\omega \notin I_n^*) \end{cases}, \quad z_n^\circ = 1 \ (\omega \in \Omega),$$

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

$$\ell(n) = 2^{\lfloor \log_2 n \rfloor}, \quad I_n^H = \left[0, \frac{1}{\ell(n)}\right], \quad I_n^* = \left[\frac{n-\ell(n)}{\ell(n)}, \frac{n-\ell(n)+1}{\ell(n)}\right]$$
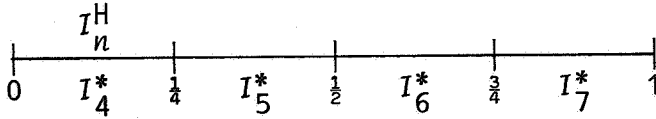
$n = 4,5,6,7$; $\ell(n) = 4$

Figure 2. Illustration of the intervals in Example 3.

(cf. Figure 2). For each $\omega \in \Omega$, $\lim_{n \to \infty} z_n^H = 1$ so that $\lim_{n \to \infty} z_n^H/z_n^\circ = 1$ as well; however, $\lim_{n \to \infty} z_n^*$ does not exist and neither does $\lim_{n \to \infty} z_n^H/z_n^*$. In probabilistic terms, we therefore have that $z_n^H/z_n^\circ \to 1$ with probability 1 but $z_n^H/z_n^* \to 1$ only in probability.

This example is due to H.C.P. Berbee. It will be shown in Theorem 3(ii) that, if z_n^H/z_n^* has a finite limit (wp1), then the implication is valid. $\square$

In the proofs of the following theorems, we adopt the common convention that the values assumed by a random variable on a set of probability measure 0 yields a zero contribution to its expected value.

Theorems 1 and 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.

(i) $z^H/z^* \to 1$ (ip) $\Rightarrow E(z^H/z^*) \to 1$, *under Assumption 1.*

(ii) $z^H/z^* \to 1$ (ip) $\Rightarrow Ez^H/Ez^* \to 1$, *under Assumptions 1 and 2.*

Proof. (i) By Assumption 1, we have $z^H/z^* < c_1$ (wp1). Hence, (i) follows from Lemma 2.

(ii) For every $\varepsilon > 0$ we define

$$\Omega'(\varepsilon) = \left\{ \omega : \left| \frac{z^H}{z^*} - 1 \right| > \varepsilon \right\}$$

and bound Ez^H from above by

$$Ez^H = \int_\Omega z^H dF(\omega) \le \int_\Omega \left(\left| \frac{z^H}{z^*} - 1 \right| + 1 \right) z^* dF(\omega),$$

so that, under Assumptions 1 and 2,

$$1 \le \frac{Ez^H}{Ez^*} \le 1 + \varepsilon + \int_{\Omega'(\varepsilon)} \left| \frac{z^H}{z^*} - 1 \right| \frac{z^*}{Ez^*} dF(\omega)$$

$$\le 1 + \varepsilon + (c_1+1)c_2 \Pr\{\omega \in \Omega'(\varepsilon)\}.$$

Since $z^H/z^* \to 1$ (ip), we have $\Pr\{\omega \in \Omega'(\varepsilon)\} \to 0$ for every $\varepsilon > 0$. It follows that $Ez^H/Ez^* \to 1$. $\square$

THEOREM 2.

(i) $E(z^H/z^o) \to 1 \Rightarrow z^H/z^o \to 1$ (ip).

(ii) $Ez^H/Ez^o \to 1 \Rightarrow z^H/z^o \to 1$ (ip), *under Assumption 4.*

(iii) $z^H/z^o \to 1$ (ip) $\Rightarrow E(z^H/z^o) \to 1$, *under Assumption 1.*

(iv) $z^H/z^o \to 1$ (ip) $\Rightarrow Ez^H/Ez^o \to 1$, *under Assumption 1.*

Proof. For every $\varepsilon > 0$ we define

$$\Omega'(\varepsilon) = \{\omega: \frac{z^H}{z^o} > 1+\varepsilon\}.$$

(i) Since $z^H/z^o \geq 1$ ($\omega \in \Omega$) and $z^H/z^o > 1+\varepsilon$ ($\omega \in \Omega'(\varepsilon)$), we can bound $E(\underset{\sim}{z}^H/\underset{\sim}{z}^o)$ from below by

$$E \frac{\underset{\sim}{z}^H}{\underset{\sim}{z}^o} = \int_\Omega \frac{z^H}{z^o} \, dF(\omega) \geq 1 + \varepsilon Pr\{\underset{\sim}{\omega} \in \Omega'(\varepsilon)\}.$$

It follows from $E(\underset{\sim}{z}^H/\underset{\sim}{z}^o) \to 1$ that $Pr\{\underset{\sim}{\omega} \in \Omega'(\varepsilon)\} \to 0$ for every $\varepsilon > 0$, i.e., $\underset{\sim}{z}^H/\underset{\sim}{z}^o \to 1$ (ip).

(ii) Under Assumption 4, we can similarly bound $E\underset{\sim}{z}^H/E\underset{\sim}{z}^o$ from below by

$$\frac{E\underset{\sim}{z}^H}{E\underset{\sim}{z}^o} = \int_\Omega \frac{z^H}{z^o} \frac{z^o}{E\underset{\sim}{z}^o} \, dF(\omega) \geq 1 + \varepsilon \int_{\Omega'(\varepsilon)} \frac{z^o}{E\underset{\sim}{z}^o} \, dF(\omega) \geq 1 + \varepsilon c_4 Pr\{\underset{\sim}{\omega} \in \Omega'(\varepsilon)\}.$$

It again follows that $\underset{\sim}{z}^H/\underset{\sim}{z}^o \to 1$ (ip).

(iii) By Assumption 1, we have $\underset{\sim}{z}^H/\underset{\sim}{z}^o < c_1$ (wp1). Hence, (iii) follows from Lemma 2.

(iv) Under Assumption 1, we can bound $E\underset{\sim}{z}^H/E\underset{\sim}{z}^o$ by

$$1 \leq \frac{E\underset{\sim}{z}^H}{E\underset{\sim}{z}^o} = \int_\Omega \frac{z^H}{E\underset{\sim}{z}^o} \, dF(\omega) \leq \int_{\Omega'(\varepsilon)} \frac{z^H}{E\underset{\sim}{z}^o} \, dF(\omega) + (1+\varepsilon) \int_\Omega \frac{z^o}{E\underset{\sim}{z}^o} \, dF(\omega)$$

$$\leq c_1 Pr\{\underset{\sim}{\omega} \in \Omega'(\varepsilon)\} + 1 + \varepsilon.$$

Since $\underset{\sim}{z}^H/\underset{\sim}{z}^o \to 1$ (ip), we have $Pr\{\underset{\sim}{\omega} \in \Omega'(\varepsilon)\} \to 0$ for every $\varepsilon > 0$. It follows that $E\underset{\sim}{z}^H/E\underset{\sim}{z}^o \to 1$. $\square$

Theorem 3 states the relations between the two-stage decision model and the distribution model.

THEOREM 3.

(i) $\underset{\sim}{z}^H/\underset{\sim}{z}^o \to 1$ (ip) $\Rightarrow \underset{\sim}{z}^H/\underset{\sim}{z}^* \to 1$ (ip), *under Assumptions 1 and 3.*

(ii) $\underset{\sim}{z}^H/\underset{\sim}{z}^o \to 1$ (wp1) $\Rightarrow \underset{\sim}{z}^H/\underset{\sim}{z}^* \to 1$ (wp1) *if $\underset{\sim}{z}^H/\underset{\sim}{z}^*$ has a finite limit* (wp1), *under Assumptions 1 and 3.*

Proof. (i) For every $\varepsilon > 0$ we define

$$\Omega_0(\varepsilon) = \{\omega: \frac{z^H}{z^o} > 1+\varepsilon^2\},$$

$$\Omega_1(\varepsilon) = \{\omega: \frac{z^H}{z^*} > 1+\varepsilon^2\},$$

$$\Omega_2(\varepsilon) = \{\omega: 1-\varepsilon \leq \frac{z^H}{z^*} \leq 1+\varepsilon^2\},$$

$$\Omega_3(\varepsilon) = \{\omega: \frac{z^H}{z^*} < 1-\varepsilon\}$$

and bound $E\underset{\sim}{z}^H$ from above by

$$E\underset{\sim}{z}^H = \int_\Omega z^H \, dF(\omega)$$

$$\leq \int_{\Omega_1(\varepsilon)} z^H dF(\omega) + (1+\varepsilon^2)\int_{\Omega_2(\varepsilon)} z^* dF(\omega) + (1-\varepsilon)\int_{\Omega_3(\varepsilon)} z^* dF(\omega),$$

so that, under Assumptions 1 and 3,

$$1 \leq \frac{E\underset{\sim}{z}^H}{E\underset{\sim}{z}^*} \leq \int_{\Omega_1(\varepsilon)} \frac{z^H}{E\underset{\sim}{z}^*} \, dF(\omega) + (1+\varepsilon^2) - (\varepsilon^2+\varepsilon)\int_{\Omega_3(\varepsilon)} \frac{z^*}{E\underset{\sim}{z}^*} \, dF(\omega)$$

$$\leq c_1 \Pr\{\underset{\sim}{\omega} \in \Omega_1(\varepsilon)\} + 1 + \varepsilon^2 - (\varepsilon^2+\varepsilon)c_3\Pr\{\underset{\sim}{\omega} \in \Omega_3(\varepsilon)\},$$

that is,

$$\Pr\{\underset{\sim}{\omega} \in \Omega_3(\varepsilon)\} \leq \frac{c_1}{(\varepsilon^2+\varepsilon)c_3} \Pr\{\underset{\sim}{\omega} \in \Omega_1(\varepsilon)\} + \frac{\varepsilon}{(\varepsilon+1)c_3}.$$

Since $\underset{\sim}{z}^H/\underset{\sim}{z}^o \to 1$ (ip), we have $\Pr\{\underset{\sim}{\omega} \in \Omega_0(\varepsilon)\} \to 0$ for every $\varepsilon > 0$ and, since $\Omega_0(\varepsilon) \supseteq \Omega_1(\varepsilon)$, $\Pr\{\underset{\sim}{\omega} \in \Omega_1(\varepsilon)\} \to 0$ for every $\varepsilon > 0$ as well. It is not difficult to see that this result together with the above upper bound on $\Pr\{\underset{\sim}{\omega} \in \Omega_3(\varepsilon)\}$ imply that also $\Pr\{\underset{\sim}{\omega} \in \Omega_3(\varepsilon)\} \to 0$ for every $\varepsilon > 0$. It follows that $\Pr\{\underset{\sim}{\omega} \in \Omega_1(\varepsilon) \cup \Omega_3(\varepsilon)\} \to 0$ for every $\varepsilon > 0$, i.e., $\underset{\sim}{z}^H/\underset{\sim}{z}^* \to 1$ (ip).

(ii) There exists a constant c such that $\underset{\sim}{z}^H/\underset{\sim}{z}^* \to c$ (wp1). Since $\underset{\sim}{z}^H/\underset{\sim}{z}^o \to 1$ (wp1) and $\underset{\sim}{z}^o \leq \underset{\sim}{z}^*$, we know that $c \leq 1$. Let n denote problem size. For every $\varepsilon > 0$ we define

$$\Omega'(\varepsilon) = \{\omega: \lim_{n\to\infty} \frac{z^H}{z^*} < 1-\varepsilon\}$$

and write

$$1 = \lim_{n\to\infty} E\, \frac{z^H}{z^*} = \int_{\Omega'(\varepsilon)} \lim_{n\to\infty} \frac{z^H}{z^*}\, dF(\omega) + \int_{\Omega\backslash\Omega'(\varepsilon)} \lim_{n\to\infty} \frac{z^H}{z^*}\, dF(\omega).$$

Under Assumptions 1 and 3, the first equality follows from Lemma 1, Theorem 3(i) and Theorem 1(i) (cf. Figure 1). The second equality holds under Assumption 1: z^H/z^* is uniformly bounded (wp1), so that for each $\Omega'' \subseteq \Omega$ the limit and the integral over Ω'' can be interchanged [Halmos 1973, p.114]. Therefore,

$$1 \le (1-\varepsilon)\Pr\{\omega \in \Omega'(\varepsilon)\} + 1 - \Pr\{\omega \in \Omega'(\varepsilon)\},$$

which implies that $\Pr\{\omega \in \Omega'(\varepsilon)\} = 0$ for every $\varepsilon > 0$. It follows that $c = 0.$ □

3. APPLICATIONS

3.1. A general two-level planning problem

We will consider a specific type of the distribution model formulated in Section 2.1: at the aggregate level, $X = \mathbb{N}$ and $f(X) = cX$ for a given constant $c > 0$, and the objective is to determine a function $X^o : \Omega \to \mathbb{N}$ such that for each $\omega \in \Omega$

$$z^*(X^o(\omega),\omega) = \min_{X\in\mathbb{N}} \{cX+g^*(X,\omega)\}.$$

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 have been investigated in [Dempster et al. 1983, Marchetti Spaccamela et al. 1982, Frenk et al. 1983]. They share some common 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 first stage and the overall decision.

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

$$\frac{w(\omega)}{X^\gamma} \leq g^*(X,\omega).$$

Secondly, there often is an estimate v of $w(\underset{\sim}{\omega})$ that is asymptotically accurate with probability 1 and that depends on the problem size and the probability distribution of $\underset{\sim}{\omega}$:

$$\frac{w(\underset{\sim}{\omega})}{v} \to 1 \text{ (wp1)}.$$

Such value estimates are available for various combinatorial optimization problems, as has been mentioned already in Section 2.2.

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

$$z^{H_1}(X) = cX + \frac{v}{X^\gamma} \leq cX + g^*(X,\underset{\sim}{\omega}) = z^*(X,\underset{\sim}{\omega}) \text{ (wp1)}.$$

Observing that z^{H_1} is a 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 $g^*(X^o(\omega),\omega)$ which is asymptotically equal to the above probabilistic lower bound with probability 1:

$$\frac{g^{H_2}(X^{H_1},\underset{\sim}{\omega})}{g^{H_1}(X^{H_1})} \to 1 \text{ (wp1)}.$$

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.

14

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 4. *If $X = \mathbb{N}$, $f(X) = cX$ ($c > 0$) and H_1 and H_2 are such that $g^{H_1}(X) = v/X^\gamma \leq g^*(X, \underset{\sim}{\omega})$ (wp1) and $g^{H_2}(X^{H_1}, \underset{\sim}{\omega})/g^{H_1}(X^{H_1}) \to 1$ (wp1), then*

(i) $z^{H_2}(X^{H_1}, \underset{\sim}{\omega})/z^*(X^\circ(\underset{\sim}{\omega}), \underset{\sim}{\omega}) \to 1$ *(wp1);*

(ii) $X^{H_1}/X^\circ(\underset{\sim}{\omega}) \to 1$ *(wp1).*

Proof. (i) We can bound $z^*(X, \underset{\sim}{\omega})$ from below (asymptotically with probability 1) and from above (deterministically) by

$$cX + g^{H_1}(X) = z^{H_1}(X) \leq z^*(X, \underset{\sim}{\omega}) \leq z^{H_2}(X, \underset{\sim}{\omega}) = cX + g^{H_2}(X, \underset{\sim}{\omega}) \quad (wp1),$$

so that

$$cX^{H_1} + g^{H_1}(X^{H_1}) \leq z^*(X^\circ(\underset{\sim}{\omega}), \underset{\sim}{\omega}) \leq cX^{H_1} + g^{H_2}(X^{H_1}, \underset{\sim}{\omega}) \quad (wp1).$$

The assumption that $g^{H_2}(X^{H_1}, \underset{\sim}{\omega})/g^{H_1}(X^{H_1}) \to 1$ (wp1) gives the desired result. (ii) Let n denote problem size. For each $\varepsilon > 0$ we define

$$\Omega'(\varepsilon) = \{\omega : \lim_{n \to \infty} \frac{X^{H_1}}{X^\circ(\omega)} < \frac{1}{1+\varepsilon}\}.$$

The unimodality of z^{H_1} implies that for $\underset{\sim}{\omega} \in \Omega'(\varepsilon)$ asymptotically

$$z^{H_1}((1+\varepsilon)X^{H_1}) \leq z^{H_1}(X^\circ(\underset{\sim}{\omega})) \leq z^*(X^\circ(\underset{\sim}{\omega}), \underset{\sim}{\omega}) \leq z^{H_2}(X^{H_1}, \underset{\sim}{\omega}) \quad (wp1).$$

A tedious but straightforward calculation shows that for each n

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

Hence, we have for $\underset{\sim}{\omega} \in \Omega'(\varepsilon)$ that

$$\lim_{n \to \infty} \frac{z^{H_2}(X^{H_1}, \underset{\sim}{\omega})}{z^{H_1}(X^{H_1})} > 1 \quad (wp1).$$

On the other hand, we know for $\underset{\sim}{\omega} \in \Omega$ that this limit is equal to 1 (wp1),

15

so that $\Pr\{\underset{\sim}{\omega} \in \Omega'(\varepsilon)\} = 0$ for every $\varepsilon > 0$. Similarly, $\Pr\{\omega: \lim_{n\to\infty} X^{H_1}/X^{\circ}(\omega) > 1/(1-\varepsilon)\} = 0$. It follows that $X^{H_1}/X^{\circ}(\underset{\sim}{\omega}) \to 1$ (wp1). $\square$

Theorem 4 deals with convergence with probability 1 in the distribution model. It is not hard to obtain analogous results for the other types of stochastic convergence, also in the context of the two-stage decision model.

As an example, we consider the hierarchical *scheduling* problem from [Dempster *et al.* 1983, Section 2]. At the aggregate level, one has to decide on the number X of *identical parallel machines* that are to be acquired, while knowing the number n of *jobs* that are to be processed. The job *processing times* $\underset{\sim}{\omega}_1, \dots, \underset{\sim}{\omega}_n$ are independently and identically distributed, with expectation μ and finite second moment. At the detailed level, a realization $\omega = (\omega_1, \dots, \omega_n)$ of the processing times becomes known, and one has to decide on a schedule in which each machine processes at most one job at a time, job j is processed during an uninterrupted period of length ω_j (j = 1,...,n), and no job is processed prior to time 0, so as to achieve a minimum value $g^*(X,\omega)$ of the maximum job completion time. We define $\omega_{sum} = \sum_{j=1}^n \omega_j$ and $\omega_{max} = \max_{1\le j\le n}\{\omega_j\}$.

First, it is obvious that $\omega_{sum}/X \le g^*(X,\omega)$, so that $w(\omega) = \omega_{sum}$ and $\gamma = 1$. Secondly, $\underset{\sim}{\omega}_{sum}/n\mu \to 1$ (wp1), so that $v = n\mu$. It follows that $g^{H_1}(X) = n\mu/X$ and $\hat{X} = \sqrt{n\mu/c}$.

Thirdly, any list scheduling algorithm, which considers the jobs in an arbitrarily fixed order and assigns each next one to the earliest available machine, achieves a schedule length $g^{H_2}(X,\omega) \le \omega_{sum}/X + \omega_{max}$. The finiteness of the second moment of the $\underset{\sim}{\omega}_j$ implies that $\underset{\sim}{\omega}_{max}/\sqrt{n} \to 0$ (wp1) and therefore $g^{H_2}(\hat{X},\underset{\sim}{\omega})/g^{H_1}(\hat{X}) \to 1$ (wp1).

The two-stage scheduling heuristic thus satisfies the properties expressed by Theorem 4. Under the perfectly reasonable assumption that there exist constants $\omega^L, \omega^U \in \mathbb{N}$ for which $\omega^L \le \omega_j \le \omega^U$ (j = 1,...,n), Assumptions 1-4 from Section 2.4 are valid, as the reader should verify. This implies, by Figure 1, that the heuristic satisfies a wide range of convergence properties. The extent to which Assumptions 1-4 are valid in the context of the *routing* model considered in [Marchetti Spaccamela *et al.* 1982] is a subject of further investigations.

3.2. A two-level scheduling problem

We finally consider an extension of the hierarchical scheduling problem discussed in Section 3.1. The difference is that, at the aggregate level, one has to decide upon the acquisition of a subset X out of a set M of *uniform* parallel machines; therefore, $X = 2^M$, the power set of M. For each machine $i \in M$, a *cost* c_i and a *speed* s_i are specified; we write $c(X) = \sum_{i \in X} c_i$, $s(X) = \sum_{i \in X} s_i$ $(X \in X)$. When, at the aggregate level, job j is scheduled on machine i, it has to be processed during a period of length ω_j/s_i. The further problem specification and notation are the same as for the identical machine problem.

We assume that there exist constants $c^L, c^U, s^L, s^U \in \mathbb{N}$ for which $c^L \leq c_i \leq c^U$ and $s^L \leq s_i \leq s^U$ $(i \in M)$. This will imply that the number of machines selected at the aggregate level grows as $\sqrt{n}$. It is then 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 selection algorithm. We therefore require that there exist constants $D, D' > 0$, $d' \geq d > 0$ such that $Dn^{\frac{1}{2}+d} \leq |M| \leq D'n^{\frac{1}{2}+d'}$.

In [Dempster *et al.* 1983, Section 3], the observation that $\omega_{sum}/s(X) \leq g^*(X,\omega)$ led to the choice of $g^{H1}(X) = n\mu/s(X)$ as an estimate of $Eg^*(X,\omega)$. As the minimization of $z^{H1}(X) = c(X)+n\mu/s(X)$ over $X \in X$ turned out to be NP-hard, a *greedy* heuristic G was proposed to find an approximation X^G to X^{H1}, for which

$$z^{H1}(X^G) \leq z^{H1}(X^{H1}) + c^U.$$

The use of a list scheduling algorithm H_2 at the detailed level resulted in a heuristic (G, H_2) that was proved to be asymptotically expectation-optimal and even asymptotically optimal in probability.

We will consider the distribution model rather than the two-stage decision model and prove the stronger result that the same heuristic is asymptotically clairvoyant with probability 1.

The following lemma establishes the asymptotic behavior of the optimal and heuristic solution values at the detailed level.

LEMMA 3. *For each* $X \in \mathcal{X}$ *with* $s(X) = O(\sqrt{n})$,

(i) $g^*(X,\underset{\sim}{\omega})/(n\mu/s(X)) \to 1$ (wp1);

(ii) $g^{H_2}(X,\underset{\sim}{\omega})/(n\mu/s(X)) \to 1$ (wp1).

Proof. For every realization ω of $\underset{\sim}{\omega}$ we have

$$\frac{\omega_{sum}}{s(X)} \le g^*(X,\omega) \le g^{H_2}(X,\omega) \le \frac{\omega_{sum}}{s(X)} + \frac{\omega_{max}}{s^L}.$$

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

$$\frac{\omega_{sum}-n\mu}{n\mu} + 1 \le \frac{g^*(X,\omega)}{n\mu/s(X)} \le \frac{g^{H_2}(X,\omega)}{n\mu/s(X)} \le \frac{\omega_{sum}-n\mu}{n\mu} + 1 + \frac{\omega_{max}s(X)}{n\mu s^L}.$$

By the strong law of large numbers, $(\underset{\sim}{\omega}_{sum}-n\mu)/n\mu \to 0$ (wp1) [Ash 1972]. Due to the finiteness of the second moment of the $\underset{\sim}{\omega}_j$, $\underset{\sim}{\omega}_{max}/\sqrt{n} \to 0$ (wp1). The assumption that $s(X) = O(\sqrt{n})$ implies (i) and (ii). $\square$

For every realization ω of $\underset{\sim}{\omega}$ we know that

$$z^*(X,\omega) \ge c(X) + \frac{\omega_{sum}}{s(X)}.$$

In order to prove that (G,H_2) is asymptotically clairvoyant with probability 1, we need a probabilistic analog of this deterministic inequality. Since $\underset{\sim}{\omega}_1,\ldots,\underset{\sim}{\omega}_n$ are independent random variables, each with the same fixed second moment, Komogorov's strong law of large numbers [Ash 1972, p. 274] implies that $(\underset{\sim}{\omega}_{sum}-n\mu)/n^\alpha \to 0$ (wp1) for any $\alpha > \frac{1}{2}$. Choosing a fixed $\alpha \in (\frac{1}{2},1)$, we have that asymptotically $\underset{\sim}{\omega}_{sum}-n\mu \ge -n^\alpha$ (wp1) and hence

$$z^*(X,\underset{\sim}{\omega}) \ge c(X) + \frac{n\mu-n^\alpha}{s(X)} \quad (wp1).$$

Let $X' \in \mathcal{X}$ denote the minimand of the right-hand side of this inequality. The following lemma states that the total speeds of the sets X' and X^G grow as $\sqrt{n}$.

LEMMA 4.

(i) $s(X') = \theta(\sqrt{n})$.

(ii) $s(X^G) = \theta(\sqrt{n})$.

Proof. (i) The proof proceeds along the same lines as that of Lemma 6(a) in [Dempster *et al.* 1983].

(ii) See Lemma 6(b), *ibidem.* □

We are now ready to prove our convergence result.

THEOREM 5. $z^{H_2}(X^G, \underline{\omega})/z^*(X^o(\underline{\omega}), \underline{\omega}) \to 1$ (wp1).

Proof. For every realization ω of $\underline{\omega}$, we have

$$z^{H_2}(X^G, \omega) \leq c(X^G) + \frac{\omega_{sum}}{s(X^G)} + \frac{\omega_{max}}{s^L}$$

$$= z^{H_1}(X^G) + \frac{\omega_{sum} - n\mu}{s(X^G)} + \frac{\omega_{max}}{s^L}$$

$$\leq z^{H_1}(X^{H_1}) + c^U + \frac{\omega_{sum} - n\mu}{s(X^G)} + \frac{\omega_{max}}{s^L}$$

$$\leq z^{H_1}(X') + c^U + \frac{\omega_{sum} - n\mu}{s(X^G)} + \frac{\omega_{max}}{s^L}$$

$$= c(X') + \frac{n\mu - n^\alpha}{s(X')} + \frac{n^\alpha}{s(X')} + c^U + \frac{\omega_{sum} - n\mu}{s(X^G)} + \frac{\omega_{max}}{s^L}.$$

We can bound the performance ratio from below (deterministically) and from above (asymptotically with probability 1) by

$$1 \leq \frac{z^{H_2}(X^G, \underline{\omega})}{z^*(X^o(\underline{\omega}), \underline{\omega})} \leq 1 + \frac{\dfrac{n^\alpha}{s(X')} + c^U + \dfrac{\underline{\omega}_{sum} - n\mu}{s(X^G)} + \dfrac{\underline{\omega}_{max}}{s^L}}{c(X') + \dfrac{n\mu - n^\alpha}{s(X')}}$$

$$\leq 1 + \frac{s^U}{c^L}\left(\frac{n^\alpha}{(s(X'))^2} + \frac{c^U}{s(X')} + \frac{\underline{\omega}_{sum} - n\mu}{s(X^G)s(X')} + \frac{\underline{\omega}_{max}}{s^L s(X')}\right) \quad (wp1).$$

Consider each of the four terms in the last factor above and recall Lemma 4. Since $\alpha < 1$, the first term tends to zero. The second one also tends to zero. By the strong law of large numbers, the third term tends to zero (wp1). Due to the finiteness of the second moment of the ω_j, the fourth term tends to zero as well (wp1). These observations prove the theorem. $\square$

ACKNOWLEDGEMENT

REFERENCES

1. R.B. ASH (1972). *Real Analysis and Probability*, Academic Press, New York.

2. J. BEARDWOOD, H.J. HALTON & J.M. HAMMERSLEY (1959). The shortest path through many points. *Proc. Cambridge Phil. Soc. 55*, 299-327.

3. G.R. BITRAN & A.C. HAX (1977). On the design of hierarchical production planning systems. *Decision Sci. 8*, 28-55.

4. S.P. BRADLEY, A.C. HAX & T.L. MAGNANTI (1977). *Applied Mathematical Programming*, Addison-Wesley, Reading, MA, 212-213.

5. M.A.H. DEMPSTER, M.L. FISHER, L. JANSEN, B.J. LAGEWEG, J.K. LENSTRA & A.H.G. RINNOOY KAN (1981). Analytical evaluation of hierarchical planning systems. *Oper. Res. 29*, 707-716.

6. 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.*, to appear.

7. M.L. FISHER & D.S. HOCHBAUM (1980). Probabilistic analysis of the planar K-median problem. *Math. Oper. Res. 5*, 27-34.

8. J.B.G. FRENK, A.H.G. RINNOOY KAN & L. STOUGIE (1983). A hierarchical scheduling problem with a well-solvable second stage. Report BW 177, Mathematisch Centrum, Amsterdam.

20

9.   P.R. HALMOS (1974). *Measure Theory*, Springer, New York.

10. D. HOCHBAUM & J.M. STEELE (1981). Steinhaus' geometric location problem
     for random samples in the plane. *Adv. in Appl. Probab. 14*, 56-67.

11. A. MARCHETTI SPACCAMELA, A.H.G. RINNOOY KAN & L. STOUGIE (1982). Hier-
     archical vehicle routing problems. Report BW 168, Mathematisch
     Centrum, Amsterdam.

12. C.H. PAPADIMITRIOU (1981). Worst case and probabilistic analysis of a
     geometric location problem. *SIAM J. Comput. 10*, 542-557.

13. R.F. SERFLING (1980). *Approximation Theorems of Mathematical Statistics*,
     Wiley, New York.

14. J.M. STEELE (1981). Subadditive Euclidean functionals and nonlinear
     growth in geometric probability. *Ann. Probab. 9*, 365-376.

15. E. ZEMEL (1983). Probabilistic analysis of geometric location problems.
     Unpublished manuscript.