Mathematical Programming 51 (1991) 349-358
North-Holland

On rates of convergence and asymptotic normality in the multiknapsack problem

Sara van de Geer
Centre for Mathematics and Computer Science, P.O. Box 4079, 1009 AB Amsterdam, Netherlands

Leen Stougie
Institute for Actuarial Sciences and Econometrics, University of Amsterdam, 1011 NH Amsterdam, Netherlands

Received 21 November 1988
Revised manuscript received July 1989

In Meanti et al. (1990) an almost sure asymptotic characterization has been derived for the optimal solution value as function of the knapsack capacities, when the profit and requirement coefficients of items to be selected from are random variables. In this paper we establish a rate of convergence for this process using results from the theory of empirical processes.

AMS 1980 Subject Classifications: 60F05, 60G50, 90C10.

Key words: Multiknapsack value function, Lagrangean relaxation, asymptotic characterization, uniform strong law of large numbers, uniform law of the iterated logarithm, Vapnik–Chervonenkis class, asymptotic normality.

1. Introduction

In Meanti et al. (1990) the optimal value of the *multiknapsack problem* is studied as a function of the knapsack capacities. They consider the following model. Let a_{ij} be the *amount of space* in the ith knapsack $(i=1,\ldots,m)$ *required* by the jth item $(j=1,\ldots,n)$. Item j yields a *profit* c_j $(j=1,\ldots,n)$ upon inclusion. The ith knapsack has *capacity* b_i $(i=1,\ldots,m)$. The multiknapsack problem is formulated as:

$$\begin{aligned}
(\mathrm{MK})\quad & \max \quad \sum_{j=1}^{n} c_j x_j \\
& \text{s.t.} \quad \sum_{j=1}^{n} a_{ij} x_j \leqslant b_i \quad (i=1,\ldots,m), \\
& \qquad\quad x_j \in \{0,1\} \quad (j=1,\ldots,n).
\end{aligned}$$

Meanti et al. (1990) have shown that if the coefficients c_j and a_{ij} $(j=1,\ldots,n,$ $i=1,\ldots,m)$ are generated by an appropriate random mechanism, then the sequence of optimal values of (MK), properly normalized, converges *with probability one* (w.p. 1) to a function of the b_i's, as n goes to infinity and m remains fixed. A crucial step in their proof of this result is the derivation of a *uniform strong law of large numbers*, using theory of convergence of convex functions.

We will show in this paper that results from *empirical process theory* can be applied to reprove this result. More interestingly, the application of empirical process theory allows for a rather straightforward derivation of a *rate of convergence* by establishing a *law of the iterated logarithm*. These results are presented in Section 3. Moreover, results from empirical process theory can be used to derive a central limit theorem as will be shown in Section 4.

First, in Section 2 we give an outline of Meanti et al. (1990) and indicate where our results fit in.

In Section 5 we discuss the interest of such results and the role that application of empirical process theory may play in this field of research.

2. Convergence of the multiknapsack value function

Meanti et al. (1990) assume that the profit coefficients $\boldsymbol{c}_j$, $j=1,\dots,n$, are i.i.d. nonnegative random variables with finite expectation, and that the vectors of requirement coefficients $\boldsymbol{a}_j=(\boldsymbol{a}_{1j},\dots,\boldsymbol{a}_{mj})^{\mathrm{T}}$, $j=1,\dots,n$, are i.i.d. nonnegative random vectors with finite expectations. The profit coefficients and requirement coefficients are independent from each other. Let $b_i=n\beta_i$, $i=1,\dots,m$, for $\beta=(\beta_1,\dots,\beta_m)^{\mathrm{T}}\in V:=\{\beta: 0\leqslant\beta_i\leqslant E\boldsymbol{a}_{i1}, i=1,\dots,m\}$. The asymptotic behaviour of the optimal value $\boldsymbol{z}_n^{\mathrm{I}}$ of (MK) is established as a function of β.

For any nonnegative vector of multipliers $\lambda=(\lambda_1,\dots,\lambda_m)^{\mathrm{T}}$ the optimal value of the *Lagrangean relaxation* of the *linear programming* (*continuous*) *relaxation* of (MK) is defined as

$$\boldsymbol{w}_n(\lambda)=\max\left\{\sum_{i=1}^{m}\lambda_i b_i+\sum_{j=1}^{n}\left(\boldsymbol{c}_j-\sum_{i=1}^{m}\lambda_i\boldsymbol{a}_{ij}\right)x_j \;\middle|\; 0\leqslant x_j\leqslant 1, j=1,\dots,n\right\}$$
$$=\sum_{i=1}^{m}\lambda_i b_i+\sum_{j=1}^{n}\left(\boldsymbol{c}_j-\sum_{i=1}^{m}\lambda_i\boldsymbol{a}_{ij}\right)\boldsymbol{x}_j^{\mathrm{L}}(\lambda),$$

where

$$\boldsymbol{x}_j^{\mathrm{L}}(\lambda)=\begin{cases}1 & \text{if } \boldsymbol{c}_j-\sum_{i=1}^{m}\lambda_i\boldsymbol{a}_{ij}>0,\\ 0 & \text{otherwise,}\end{cases}\qquad j=1,\dots,n.$$

Define

$$\boldsymbol{L}_n(\lambda):=\frac{1}{n}\boldsymbol{w}_n(\lambda)=\lambda^{\mathrm{T}}\beta+\frac{1}{n}\sum_{j=1}^{n}(\boldsymbol{c}_j-\lambda^{\mathrm{T}}\boldsymbol{a}_j)\boldsymbol{x}_j^{\mathrm{L}}(\lambda),$$

and let $\boldsymbol{\lambda}_n^*$ be a vector such that $\boldsymbol{L}_n(\boldsymbol{\lambda}_n^*)=\min_{\lambda\geqslant 0}\boldsymbol{L}_n(\lambda)$. Then, Meanti et al. (1990) show that

$$\boldsymbol{L}_n(\boldsymbol{\lambda}_n^*)-\frac{m}{n}\left(\max_{j=1,\dots,n}\boldsymbol{c}_j\right)\leqslant\frac{1}{n}\boldsymbol{z}_n^{\mathrm{I}}\leqslant\boldsymbol{L}_n(\boldsymbol{\lambda}_n^*). \tag{2.1}$$

Let the function $L(\lambda)$ be defined as

$$L(\lambda) := \lambda^{\mathrm{T}}\beta + E(\boldsymbol{c}_1 - \lambda^{\mathrm{T}}\boldsymbol{a}_1)\boldsymbol{x}_1^{\mathrm{L}}(\lambda)$$

and let λ^* be a minimizer of $L(\lambda)$. Theorem 3.1 in Meanti et al. (1990) states that

$$\lim_{n\to\infty} |\boldsymbol{L}_n(\boldsymbol{\lambda}_n^*) - L(\lambda^*)| = 0 \quad \text{w.p. 1.} \tag{2.2}$$

This, together with (2.1), implies their main result:

$$\lim_{n\to\infty} \left| \frac{1}{n} z_n^{\mathrm{I}} - L(\lambda^*) \right| = 0 \quad \text{w.p. 1.}$$

To prove result (2.2) they show that the strong law of large numbers, which implies that

$$\lim_{n\to\infty} |\boldsymbol{L}_n(\lambda) - L(\lambda)| = 0 \quad \text{w.p. 1}$$

holds uniformly over all λ in a *compact set* $S \subset \mathbb{R}^m$, using the *convexity* of the functions $\boldsymbol{L}_n(\lambda)$ and $L(\lambda)$. From this (2.2) follows almost immediately (cf. Meanti et al., 1990, Proof of Theorem 3.1). In the following section we show that results from empirical process theory can be applied to reprove the uniform strong law of large numbers, and, moreover, to establish a rate of convergence. For simplicity we will assume throughout the paper that $\boldsymbol{c}$ and $\boldsymbol{a}$ have bounded supports. Specifically, we will present a uniform law of the iterated logarithm that yields

$$\left(\frac{n}{\log\log n}\right)^{1/2} |\boldsymbol{L}_n(\boldsymbol{\lambda}_n^*) - L(\lambda^*)| = \mathrm{O}(1) \quad \text{w.p. 1.}$$

Furthermore, in Section 4, we use a general theorem from Pollard (1984) (see also Pollard, 1985) to derive asymptotic normality of the optimal Lagrangean multipliers $\boldsymbol{\lambda}_n^*$, of the optimal value of the Lagrangean relaxation $\boldsymbol{L}_n(\boldsymbol{\lambda}_n^*)$, and of the normalized optimal value of (MK) z_n^{I}. The asymptotic properties of $\boldsymbol{\lambda}_n^*$ are particularly interesting in view of its use in a heuristic procedure for solving the multiknapsack problem (see Rinnooy Kan et al., 1992).

3. Rate of convergence

Consider the functions $f_\lambda : \mathbb{R}^{m+1} \to \mathbb{R}$, for $\lambda \geqslant 0$, defined as

$$f_\lambda(c, a) = \begin{cases} \lambda^{\mathrm{T}}\beta + (c - \lambda^{\mathrm{T}}a)1_{\{(c,a):c>\lambda^{\mathrm{T}}a\}}(c, a), & c \geqslant 0,\ a \geqslant 0, \\ 0, & \text{otherwise,} \end{cases}$$

where $1_{\{(c,a):c>\lambda^{\mathrm{T}}a\}}(c, a)$ is the indicator function of the set $\{(c, a): c > \lambda^{\mathrm{T}}a\}$. Then $\boldsymbol{L}_n(\lambda)$ is the *mean value* of $f_\lambda(c, a)$ over n independent observations $(\boldsymbol{c}_1, \boldsymbol{a}_1), \ldots, (\boldsymbol{c}_n, \boldsymbol{a}_n)$,

$$\boldsymbol{L}_n(\lambda) = \frac{1}{n} \sum_{j=1}^{n} f_\lambda(\boldsymbol{c}_j, \boldsymbol{a}_j),$$

and $L(\lambda)$ is the *expectation* of $f_\lambda(\boldsymbol{c}, \boldsymbol{a})$,

$$L(\lambda) = Ef_\lambda(\boldsymbol{c}_1, \boldsymbol{a}_1).$$

Let $\mathscr{F}$ be the class of functions f_λ made up by all possible vectors $\lambda \geqslant 0$:

$$\mathscr{F} := \{f_\lambda : \lambda \geqslant 0\}.$$

The *graph* of a function $g : \mathbb{R}^d \to \mathbb{R}$ is formulated as

$$\text{graph } g = \{(t, x) \in \mathbb{R}^{d+1} : 0 \leqslant t \leqslant g(x) \vee g(x) \leqslant t \leqslant 0\}.$$

We will present some concepts and results from empirical process theory and show that the class of graphs of the functions in $\mathscr{F}$ has properties that allow direct application of these results.

Definition 1. Let $\mathscr{D}$ be a class of subsets of a space X. For $x_1, x_2, \ldots, x_n \in X$ define

$$\Delta^{\mathscr{D}}(x_1, \ldots, x_n) := \text{card}\{D \cap \{x_1, \ldots, x_n\} : D \in \mathscr{D}\}$$

and

$$m^{\mathscr{D}}(n) := \sup\{\Delta^{\mathscr{D}}(x_1, \ldots, x_n) : x_1, \ldots, x_n \in X\}.$$

Note that $m^{\mathscr{D}}(n) \leqslant 2^n$. The class $\mathscr{D}$ is called a *Vapnik-Chervonenkis class* if $m^{\mathscr{D}}(n) < 2^n$ for some $n \geqslant 1$ (cf. Vapnik and Chervonenkis, 1971).

For classes of functions we have a similar definition based on their graphs.

Definition 2. A class $\mathscr{G}$ of real-valued functions is called a *Vapnik-Chervonenkis graph class* if the graphs of the functions in $\mathscr{G}$ form a Vapnik-Chervonenkis class.

The following theorem from Alexander (1984) establishes a uniform law of the iterated logarithm for a Vapnik-Chervonenkis graph class of functions.

Theorem 3.1. *Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be a sequence of i.i.d. random variables taking values in a space $(X, \mathscr{A})$ and let $\mathscr{G}$ be a class of measurable real valued functions on X, such that*
(i) *$\mathscr{G}$ is a Vapnik-Chervonenkis graph class, and*
(ii) *the functions in $\mathscr{G}$ are uniformly bounded.*
Then, modulo measurability,

$$\sup_{g \in \mathscr{G}} \left(\frac{n}{\log \log n}\right)^{1/2} \left| \frac{1}{n} \sum_{j=1}^{n} g(\boldsymbol{x}_j) - Eg(\boldsymbol{x}_1) \right| = \mathrm{O}(1) \quad \textit{w.p. } 1. \qquad \square$$

For application of this theorem in our analysis we have to verify the two conditions (i) and (ii) for the class of functions $\{f_\lambda : \lambda \geqslant 0\}$. To show that the first condition is satisfied is rather straightforward, given the theory on Vapnik-Cheronenkis classes. It is known that classes of open and closed halfspaces in $\mathbb{R}^r$ say, i.e. $\mathscr{D}_1 = \{\{x : \theta^{\mathrm{T}} x > 0\}, \theta \in \mathbb{R}^r\}$ and $\mathscr{D}_2 = \{\{x : \theta^{\mathrm{T}} x \geqslant 0\}, \theta \in \mathbb{R}^r\}$ are Vapnik-Chervonenkis classes (see e.g. Dudley, 1984). Moreover, the Vapnik–Chervonenkis property is preserved under taking finite unions, intersections and complements Pollard (1984, p. 18). Now, the graph of a function f_λ is

$$\begin{aligned}&\{\{0\leqslant t\leqslant \lambda^{\mathrm{T}}\beta\}\cap\{c\leqslant\lambda^{\mathrm{T}}a\}\cap\{c\geqslant 0,\ a\geqslant 0\}\}\\ &\quad\cup\{\{0\leqslant t\leqslant\lambda^{\mathrm{T}}\beta+c-\lambda^{\mathrm{T}}a\}\cap\{c>\lambda^{\mathrm{T}}a\}\cap\{c\geqslant 0,\ a\geqslant 0\}\}\\ &\quad\cup\{\{t=0\}\cap\{c\geqslant 0,\ a\geqslant 0\}^{\mathrm{c}}\},\end{aligned}$$

which is clearly a finite union of finite intersections of halfspaces in $\mathbb{R}^{m+2}$.

As for the second condition of Theorem 3.1, we notice that, under the assumption of bounded support of $\boldsymbol{c}$ and $\boldsymbol{a}$, the functions f_λ are uniformly bounded only for λ in a bounded set. In Meanti et al. (1990, Lemma 3.1), it is shown that the interesting values of λ, i.e., those values that are candidates for minimizing $\boldsymbol{L}_n(\lambda)$ and $L(\lambda)$, are in the set $S:=\{\lambda:\lambda^{\mathrm{T}}\beta\leqslant E\boldsymbol{c}_1+1,\ \lambda\geqslant 0\}$. Therefore, we arrive at the following lemma.

Lemma 3.2. *If $\boldsymbol{c}$ and $\boldsymbol{a}$ have bounded support, then*

$$\sup_{\lambda\in S}\left(\frac{n}{\log\log n}\right)^{1/2}|\boldsymbol{L}_n(\lambda)-L(\lambda)|=\mathrm{O}(1)\quad w.p.\ 1.\qquad\square$$

From this lemma the following theorem follows easily.

Theorem 3.3. *If $\boldsymbol{c}$ and $\boldsymbol{a}$ have bounded support, then*

$$\left(\frac{n}{\log\log n}\right)^{1/2}|\boldsymbol{L}_n(\boldsymbol{\lambda}_n^*)-L(\lambda^*)|=\mathrm{O}(1)\quad w.p.\ 1.$$

Proof. Let λ^* be a minimum of $L(\lambda)$. If $L(\lambda^*)\leqslant\boldsymbol{L}_n(\boldsymbol{\lambda}_n^*)$, then $|\boldsymbol{L}_n(\boldsymbol{\lambda}_n^*)-L(\lambda^*)|\leqslant\boldsymbol{L}_n(\lambda^*)-L(\lambda^*)$ since $\boldsymbol{\lambda}_n^*$ minimizes $\boldsymbol{L}_n(\lambda)$. Otherwise, if $L(\lambda^*)>\boldsymbol{L}_n(\boldsymbol{\lambda}_n^*)$, then $|\boldsymbol{L}_n(\boldsymbol{\lambda}_n^*)-L(\lambda^*)|\leqslant L(\boldsymbol{\lambda}_n^*)-\boldsymbol{L}_n(\boldsymbol{\lambda}_n^*)$. Hence,

$$|\boldsymbol{L}_n(\boldsymbol{\lambda}_n^*)-L(\lambda^*)|\leqslant\sup_{\lambda\in S}|\boldsymbol{L}_n(\lambda)-L(\lambda)|.$$

This inequality together with Lemma 3.2 completes the proof. $\square$

Regarding (2.1), this theorem leads easily to the rate of convergence of the normalized optimal value z_n^{I} of the multiknapsack problem.

Theorem 3.4. *If $\boldsymbol{c}$ and $\boldsymbol{a}$ have bounded support, then*

$$\left(\frac{n}{\log\log n}\right)^{1/2}\left|\frac{1}{n}z_n^{\mathrm{I}}-L(\lambda^*)\right|=\mathrm{O}(1)\quad w.p.\ 1.\qquad\square$$

In fact, this rate of convergence is sharp in most cases (see Remark 1 at the end of Section 4).

4. Asymptotic normality

For the derivation of the results in this section we need two extra assumptions on the function $L(\lambda)$. The first one is that $L(\lambda)$ has a unique minimum λ^*. From Meanti et al. (1990, p. 242) we know that in this case $\boldsymbol{\lambda}_n^*$ converges to λ^* with probability 1. Moreover, Lemma 4.4 from the same paper gives sufficient conditions on the distribution of $\boldsymbol{c}$ and $\boldsymbol{a}$ for this to hold.

The second assumption that we need is that λ^* is an interior point of the feasible region of λ, which means that $\lambda^* > 0$. This is equal to assuming that none of the m knapsack constraints has positive expected slack. That this is not such a severe restriction comes from the fact that constraints with positive expected slack do not influence $L(\lambda^*)$.

Under these conditions Theorem VII.5 from Pollard (1984, p. 141) is tailored for establishing asymptotic normality of $\boldsymbol{\lambda}_n^*$. Before we get to the application of this theorem we derive some preliminary results. The purpose of this is to provide the reader with some insight in the matter and to use these results later on to establish asymptotic normality of $\boldsymbol{L}_n(\boldsymbol{\lambda}_n^*)$ and z_n^{I}/n.

Suppose that $\boldsymbol{L}(\lambda)$ has second derivative matrix V in λ^*. Under the assumption that λ^* is an interior point the Taylor expansion for λ close to λ^* gives

$$L(\lambda) - L(\lambda^*) = \tfrac{1}{2}(\lambda - \lambda^*)^{\mathrm{T}} V(\lambda - \lambda^*) + \mathrm{o}(\|\lambda - \lambda^*\|^2) \tag{4.1}$$

where $\|\lambda - \lambda^*\| = ((\lambda - \lambda^*)^{\mathrm{T}}(\lambda - \lambda^*))^{1/2}$. If V is non-singular, (4.1) implies that for small values of $\|\lambda - \lambda^*\|$ and some positive constant η,

$$L(\lambda) - L(\lambda^*) \geqslant \eta \|\lambda - \lambda^*\|^2. \tag{4.2}$$

Let us compare this with the behaviour of $\boldsymbol{L}_n(\lambda) - L(\lambda^*)$. For this purpose we use another theorem from empirical process theory. $\boldsymbol{y}_n = \mathrm{O}_{\mathrm{p}}(\alpha_n)$ is used as shorthand notation to say that for every $\varepsilon > 0$ there exists an $M < \infty$ such that for n large enough $\Pr\{|\boldsymbol{y}_n| > M\alpha_n\} < \varepsilon$.

Theorem 4.1. *Under the conditions of Theorem* 3.1 *we have that, modulo measurability*

$$\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{j=1}^{n} g(\boldsymbol{x}_j) - Eg(\boldsymbol{x}_1) \right| = \mathrm{O}_{\mathrm{p}}(n^{-1/2}). \qquad \square$$

This theorem is a consequence of general results on so-called Donsker-classes of functions (see Dudley, 1984; Giné and Zinn, 1984; Pollard, 1984). It provides us with the following lemma.

Lemma 4.2. *If $\boldsymbol{a}$ has a distribution with bounded support,*

$$\sup_{\lambda > 0} \frac{(\boldsymbol{L}_n(\lambda) - \boldsymbol{L}_n(\lambda^*)) - (L(\lambda) - L(\lambda^*))}{\|\lambda - \lambda^*\|} = \mathrm{O}_{\mathrm{p}}(n^{-1/2}). \tag{4.3}$$

Proof. Define the class $\mathscr{F}_1$ as

$$\mathscr{F}_1 = \{(f_\lambda - f_{\lambda^*})/\|\lambda - \lambda^*\| : \lambda \in \mathbb{R}_+^m\}.$$

$\mathscr{F}_1$ is a Vapnik-Chervonenkis graph class, since $\mathscr{F} = \{f_\lambda : \lambda \in \mathbb{R}_+^m\}$ is a Vapnik-Chervonenkis graph class, f_{λ^*} is a fixed function and the property is preserved under scaling.

It remains to show that $\mathscr{F}_1$ is uniformly bounded. Consider the case that $\lambda^{\mathrm{T}} a < \lambda^{*\mathrm{T}} a$. Then,

$$\begin{aligned}
&\frac{f_\lambda(c,a) - f_{\lambda^*}(c,a)}{\|\lambda - \lambda^*\|} \\
&= \frac{-(\lambda - \lambda^*)^{\mathrm{T}} a}{\|\lambda - \lambda^*\|} 1_{\{(c,a):\, \lambda^{*\mathrm{T}} a < c\}}(c,a) + \frac{(c - \lambda^{\mathrm{T}} a)}{\|\lambda - \lambda^*\|} 1_{\{(c,a):\, \lambda^{*\mathrm{T}} a \geq c, \lambda^{\mathrm{T}} a < c\}}(c,a) \\
&\quad + \frac{(\lambda - \lambda^*)^{\mathrm{T}} \beta}{\|\lambda - \lambda^*\|} \\
&= \frac{-(\lambda - \lambda^*)^{\mathrm{T}} a}{\|\lambda - \lambda^*\|} 1_{\{(c,a):\, \lambda^{*\mathrm{T}} a < c\}}(c,a) + \frac{(c - \lambda^{*\mathrm{T}} a)}{\|\lambda - \lambda^*\|} 1_{\{(c,a):\, \lambda^{*\mathrm{T}} a \geq c, \lambda^{\mathrm{T}} a < c\}}(c,a). \\
&\quad - \frac{(\lambda - \lambda^*)^{\mathrm{T}} a}{\|\lambda - \lambda^*\|} 1_{\{(c,a):\, \lambda^{*\mathrm{T}} a \geq c, \lambda^{\mathrm{T}} a < c\}}(c,a) + \frac{(\lambda - \lambda^*)^{\mathrm{T}} \beta}{\|\lambda - \lambda^*\|}.
\end{aligned}$$

Hence,

$$\frac{|f_\lambda(c,a) - f_{\lambda^*}(c,a)|}{\|\lambda - \lambda^*\|} \leq \frac{|(\lambda - \lambda^*)^{\mathrm{T}} a|}{\|\lambda - \lambda^*\|} + \frac{|(\lambda - \lambda^*)^{\mathrm{T}} \beta|}{\|\lambda - \lambda^*\|} \leq \iota^{\mathrm{T}} a + \iota^{\mathrm{T}} \beta,$$

where ι is the m-dimensional vector consisting of all ones. The case $\lambda^{\mathrm{T}} a > \lambda^{*\mathrm{T}} a$ can be handled similarly and yields the same bound. Since $\boldsymbol{a}$ is assumed to have bounded support, and hence $\beta < E\boldsymbol{a}_1$ is bounded, we have that $\mathscr{F}_1$ is uniformly bounded. Now, application of Theorem 4.1 to the class $\mathscr{F}_1$ finishes the proof. □

Since $\boldsymbol{\lambda}_n^*$ minimizes $\boldsymbol{L}_n(\lambda)$ we have that

$$(\boldsymbol{L}_n(\boldsymbol{\lambda}_n^*) - \boldsymbol{L}_n(\lambda^*)) - (L(\boldsymbol{\lambda}_n^*) - L(\lambda^*)) \leq -(L(\boldsymbol{\lambda}_n^*) - (L(\lambda^*)).$$

Combine this with (4.1) and (4.3) to see that

$$\|\boldsymbol{\lambda}_n^* - \lambda^*\| \mathrm{O_p}(n^{-1/2}) \leq -\eta \|\boldsymbol{\lambda}_n^* - \lambda^*\|^2$$

or

$$\|\boldsymbol{\lambda}_n^* - \lambda^*\| \leq \frac{1}{\eta} \mathrm{O_p}(n^{-1/2}) = \mathrm{O_p}(n^{-1/2}). \tag{4.4}$$

Thus we arrive at a $\sqrt{n}$-rate of convergence. Theorem VII.5 from Pollard (1984) can be used to give the limiting distribution. For a thorough explanation of the concepts playing a role in the theorem we refer the interested reader to Chapter

VII of Pollard's book. Here we will restrict ourselves to a brief exposition on the crucial condition of the theorem, which concerns the remainder term r_λ of the linear approximation of f_λ near λ^*:

$$f_\lambda(c, a) - f_{\lambda^*}(c, a) = (\lambda - \lambda^*)^{\mathrm{T}} \Delta(c, a) + \|\lambda - \lambda^*\| r_\lambda(c, a), \tag{4.5}$$

where $\Delta(c, a)$ is given by

$$\Delta(c, a) = \beta - a 1_{\{(c,a):\, \lambda^{*\mathrm{T}} a < c\}}(c, a).$$

Using (4.5) as a definition of $r_\lambda(c, a)$, it is easy to verify that the class $\{r_\lambda : \lambda \in \mathbb{R}^m_+\}$ is a Vapnik–Chervonenkis graph class. Since $\boldsymbol{a}$ has bounded support, uniform boundedness of this class follows easily from boundedness of the class $\mathscr{F}_1$ defined in the proof of Lemma 4.2, and boundedness of $\{(\lambda - \lambda^*)^{\mathrm{T}} \Delta(c, a)/\|\lambda - \lambda^*\| : \lambda \in \mathbb{R}^m_+\}$. Let the distribution of $\boldsymbol{c}$ and $\boldsymbol{a}$ be such that $\int |r_{\lambda^*_n}(c, a)|^2 \, \mathrm{d}F(c, a) \to 0$ w.p. 1, e.g. assume F to be continuous in c and a. From Pollard (1984, pp. 151–152) we know that the above properties of r are sufficient to imply condition (v) of Theorem VII.5, which is a so-called stochastic equicontinuity condition on the remainder term. It roughly says that for $\lambda^*_n \to \lambda^*$ the impact of the remainder term on the asymptotic behaviour of $\lambda^*_n - \lambda^*$ becomes negligible.

Theorem 4.3. *Suppose that $L(\lambda)$ has a unique minimum λ^* and a non-singular second derivative matrix V at λ^*. Specifically, assume that the distribution function of $\boldsymbol{c}$ and $\boldsymbol{a}$ is continuous. Then $\sqrt{n}(\boldsymbol{\lambda}^*_n - \lambda^*)$ converges in law to a multivariate normal distribution with mean 0 and covariance matrix $V^{-1}\Sigma V^{-1}$, where Σ is the covariance matrix of the random vector $\Delta(c, a)$.*

Proof. The proof is a direct application of Pollard's Theorem VII.5 to our special case. We showed above that under the assumptions the stochastic equicontinuity condition is satisfied. The other conditions of the theorem are satisfied trivially or by assumption. □

As a byproduct of the results obtained so far we will derive asymptotic normality of $\boldsymbol{L}_n(\lambda^*_n)$. Remember that $\boldsymbol{L}_n(\lambda^*)$ is just a normalized sum of i.i.d. random variables, so that $\sqrt{n}(\boldsymbol{L}_n(\lambda^*) - L(\lambda))$ converges to a normal law. The following lemma shows that $\boldsymbol{L}_n(\boldsymbol{\lambda}^*_n)$ is close enough to $\boldsymbol{L}_n(\lambda^*)$ to exhibit the same asymptotic behaviour.

Lemma 4.4. *Under the conditions of Theorem* 4.3,

$$|\boldsymbol{L}_n(\boldsymbol{\lambda}^*_n) - \boldsymbol{L}_n(\lambda^*)| = \mathrm{O}_{\mathrm{p}}(n^{-1}).$$

Proof. Combining (4.1) and (4.4) we have that

$$|L(\boldsymbol{\lambda}^*_n) - L(\lambda^*)| = \mathrm{O}_{\mathrm{p}}(n^{-1}). \tag{4.6}$$

Using (4.3) and (4.4) we see that

$$\begin{aligned}&|(L_n(\boldsymbol{\lambda}_n^*)-L_n(\lambda^*))-(L(\boldsymbol{\lambda}_n^*)-L(\lambda^*))|\\&=\|\boldsymbol{\lambda}_n^*-\lambda^*\|\frac{|(L_n(\boldsymbol{\lambda}_n^*)-L_n(\lambda^*))-(L(\boldsymbol{\lambda}_n^*)-L(\lambda^*))|}{\|\boldsymbol{\lambda}_n^*-\lambda^*\|}\\&=\mathrm{O_p}(n^{-1}).\end{aligned}$$

Together with (4.6) this implies that

$$\begin{aligned}|(L_n(\boldsymbol{\lambda}_n^*)-L_n(\lambda^*)|&\leqslant|(L_n(\boldsymbol{\lambda}_n^*)-L_n(\lambda^*))-(L(\boldsymbol{\lambda}_n^*)-L(\lambda^*))|+|L(\boldsymbol{\lambda}_n^*)-L(\lambda^*)|\\&=\mathrm{O_p}(n^{-1}).\qquad\Box\end{aligned}$$

Corollary. *Suppose that the conditions of Theorem* 4.3 *are satisfied. Let* σ^2 *be the variance of* $f_{\lambda^*}(\boldsymbol{c},\boldsymbol{a})$. *Then* $\sqrt{n}(\mathrm{L}_n(\boldsymbol{\lambda}_n^*)-L(\lambda^*))$, $\sqrt{n}(z_n^1/n-L(\lambda^*))$ *and* $\sqrt{n}(L_n(\lambda^*)-L(\lambda^*))$ *are asymptotically equivalent and converge in law to a normal random variable with mean zero and variance* σ^2. □

Remark 1. In a similar way as above it can be shown that

$$|L_n(\boldsymbol{\lambda}_n^*)-L_n(\lambda^*)|=\mathrm{O}\left(\frac{\log\log n}{n}\right)\quad\text{w.p. 1.}$$

Using the ordinary law of the iterated logarithm on $L_n(\lambda^*)$ the above implies that the almost sure rate of convergence in Theorem 3.4 is sharp:

$$\limsup_{n\to\infty}\frac{\sqrt{n}}{\sqrt{2\log\log n}}|L_n(\boldsymbol{\lambda}_n^*)-L(\lambda^*)|=\sigma\quad\text{w.p. 1.}$$

Remark 2. It is also possible to show that $n(L_n(\boldsymbol{\lambda}_n^*)-L_n(\lambda^*))$ is asymptotically chi-square distributed.

5. Postlude

We observe that there exist other examples of applications of empirical process theory in the research area of probabilistic value analysis of combinatorial problems. In a way analogous to the one in this paper an almost sure characterization of a covering problem has been established (Piersma, 1987). In Rhee and Talagrand (1988, 1989) empirical process theory has been used in probabilistic analyses of the optimal value of respectively a Euclidean matching problem and a median location problem. We also mention the probabilistic value analysis of a minimum flowtime scheduling problem (Marchetti Spaccamela et al., 1992).

Furthermore the approach of Pollard (1984) towards proving central limit theorems, turns out to be well-suited for value functions expressable as empirical process on Vapnik-Chervonenkis graph classes.

In view of this, it seems worthwhile to try for insights in the specific structures of combinatorial optimization problems that allow for such applications. That wider applicability is possible is intuitively supported by the fact that theorems from empirical process theory, like Theorem 3.1, heavily depend on combinatorial properties of classes of functions.

References

K.S. Alexander, "Probability inequalities for empirical processes and a law of the iterated logarithm," *Annals of Probability* 12 (1984) 1041-1067.

R.M. Dudley, "A course on empirical processes," *Lecture Notes in Mathematics (Lectures given at Ecole d'Eté de Probabilités de St. Flour 1982)* (Springer, Berlin, 1984) pp. 1-142.

E. Giné and J. Zinn, "On the central limit theorem for empirical process," *Annals of Probability* 12 (1984) 929-989.

A. Marchetti Spaccamela, W.T. Rhee, L. Stougie and S.A. van de Geer, "A probabilistic value analysis of the minimum weighted flowtime scheduling problem," *Operations Research Letters* 11 (1992), to appear.

M. Meanti, A.H.G. Rinnooy Kan, L. Stougie and C. Vercellis, "A probabilistic analysis of the multiknapsack value function," *Mathematical Programming* 46 (1990) 237-247.

N. Piersma, "Empirical process theory and combinatorial optimization: a survey," Thesis, University of Amsterdam (Amsterdam, 1987).

D. Pollard, *Convergence of Stochastic Processes* (Springer, New York, 1984).

D. Pollard, "New ways to prove central limit theorems," *Econometric Theory* 1 (1985) 295-314.

W.T. Rhee and M. Talagrand, "Exact bounds for the stochastic upward matching problem," *Transactions of the American Mathematical Society* 307 (1988) 109-125.

W.T. Rhee and M. Talagrand, "A concentration inequality for the k-median problem," *Mathematics of Operations Research* 14 (1989) 189-202.

A.H.G. Rinnooy Kan, L. Stougie and C. Vercellis, "A class of generalized greedy algorithms for the multiknapsack problem," to appear in: *Discrete Applied Mathematics* (1992).

V.N. Vapnik and Y.A. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," *Theory of Probability and its Applications* 16 (1971) 264-280.