A HIERARCHICAL SCHEDULING PROBLEM WITH A WELL-SOLVABLE SECOND STAGE

J.B.G. FRENK* and A.H.G. RINNOOY KAN

Econometric Institute, Erasmus University, Burgemeester Oudlaan 50, NL-3000 DR Rotterdam, The Netherlands

and

L. STOUGIE

Centrum voor Wiskunde en Informatica, Kruislaan 413, NL-1098 SJ Amsterdam, The Netherlands

Abstract

In the hierarchical scheduling model to be considered, the decision at the aggregate level to acquire a number of identical machines has to be based on probabilistic information about the jobs that have to be scheduled on these machines at the detailed level. The objective is to minimize the sum of the acquisition costs and the expected average completion time of the jobs. In contrast to previous models of this type, the second part of this objective function corresponds to a well-solvable scheduling problem that can be solved to optimality by a simple priority rule. A heuristic method to solve the entire problem is described, for which strong asymptotic optimality results can be established.

Keywords and phrases

Hierarchical planning models, identical machine scheduling.

1. Introduction

Hierarchical planning problems involve a sequence of interrelated decisions to be taken over time at an increasing level of detail and with an increasing amount of information.

In a *scheduling* context, for instance, the first decisions in such a sequence typically correspond to the acquisition of certain resources, whereas later decisions

*Now at the Department of Industrial Engineering and Operations Research, University of California, Berkeley, California, USA.

involve the precise allocation of these resources over time; the initial decisions at the *aggregate level,* however, usually have to be based on incomplete information on what the exact demand on the resources will be at the *detailed level.*

In several papers [5,6,11,12], including one that appears elsewhere in this volume [11], it has been argued that the natural way to formulate such a problem is as a *multi-stage stochastic programming problem,* in which each stage corresponds to a decision level, the problem parameters of which may initially be known only in probability. The objective will then be to set the decision variables at each level in such a way that the overall decision is optimal *in expectation.*

The resulting stochastic programming problem is difficult to solve for two reasons. In the first place, the problems that have to be solved at the detailed level usually correspond to *NP-hard* [9] combinatorial optimization problems, for which truly efficient (in the sense of *polynomially bounded* [9]) solution methods are very unlikely to exist. And secondly, the stochastic nature of the problem gives rise to additional computational challenges. Hence, the natural way to solve these problems is by means of *stochastic programming heuristics* [5,6,11,12]. Such heuristics are usually based on sharp *a priori* estimates of the optimal detailed level objective function value as a function of the aggregate level decision variables, and were shown to have strong properties of asymptotic optimality in various specific cases.

The hierarchical scheduling model studied in this paper derives its interest from the fact that the problem at the detailed level is not NP-hard but solvable in polynomially bounded time by a simple priority rule. However, the stochastic nature of the problem still forces us to resort to a heuristic solution method. In sect. 2, we introduce the model in more detail, and describe and motivate the heuristic solution method. In sect. 3, we develop and apply some advanced tools from probability theory to prove strong properties of asymptotic optimality for the heuristic solution, including an estimate of the rate at which its value converges to the value of the optimal solution. In fact, we show that the relative loss that can be ascribed to imperfect information at the aggregate level asymptotically tends to 0 almost surely (a.s.), which is the strongest possible result under the circumstances. Some concluding remarks are contained in sect. 4.

2. The scheduling model and the heuristic

Consider the following hierarchical planning problem. At the *aggregate level* a decision has to be made about the number m of *identical machines* that have to be acquired at cost c each. The machines will be used to process n jobs, whose *processing times* p_j $(j = 1, \ldots, n)$ are not yet known precisely at this level. Let us assume that these processing times can be conceived of as independent, identically distributed *random variables* with a continuous common *distribution function* $F(x)$ and (finite) expected value μ.

After m has been chosen, a realization $p = (p_1, \ldots, p_n)$ of the processing times is given and the jobs now have to be scheduled from time 0 onwards on the m machines acquired so as to minimize the *average value* $\bar{C}(m,p)$ of the job *completion times* C_j $(j = 1, \ldots, n)$. Let us denote the optimal value of $\bar{C}(m,p)$ for fixed m by $\bar{C}^0(m,p)$. Initially, before a realization of the processing times is given, this is a *random variable.* (All such variables will be underlined in the sequel.) Hence, the overall objective function $\underline{Z}(m,\underline{p})$ is given by

$$\underline{Z}(m,\underline{p}) \triangleq cm + \underline{\bar{C}}^0(m,\underline{p}) \ . \tag{1}$$

This objective reflects the trade-off between the cost of acquiring extra machines and the (possible) benefits of having these extra machines available at the detailed level. We shall want to find the value m^0 such that

$$E[\underline{Z}(m^0,\underline{p})] = \min_m \{E[\underline{Z}(m,\underline{p})]\} = \min_m \{cm + E[\underline{\bar{C}}^0(m,\underline{p})]\} \ . \tag{2}$$

As announced in the introduction, it is a peculiar and an unusual feature of this scheduling model that the optimal detailed level objective function value $\bar{C}^0(m,\underline{p})$ can be calculated in polynomial time for each realization of $\underline{p}$. Indeed, as demonstrated in [2], an optimal schedule can be constructed by assigning each job to the first available machine in order of increasing processing times. If $\underline{p}^{(1)} \leqslant \underline{p}^{(2)} \leqslant \ldots \leqslant \underline{p}^{(n)}$ are the *order statistics* of $\underline{p}_1, \underline{p}_2, \ldots, \underline{p}_n$, the optimality of the above *SPT rule* implies that

$$\underline{\bar{C}}^0(m,\underline{p}) = \frac{1}{n} \sum_{j=1}^{n} \left\lceil \frac{n-j+1}{m} \right\rceil \underline{p}^{(j)} \ . \tag{3}$$

The analysis of the expected value of (3) as a function of m is, however, not a trivial task. To find a suitable value of m at the aggregate level, we will still have to rely on a heuristic approach. As in previous cases [5,6,12], this *stochastic programming heuristic* will be based on a *lower bound* on the detailed level objective (3) whose relative error is vanishingly small. In developing such a bound, we solve an open problem posed in [4, p. 290].

A lower bound and a corresponding upper bound are given by the obvious inequalities

$$\frac{1}{n} \sum_{j=1}^{n} \frac{n-j+1}{m} \underline{p}^{(j)} \leqslant \underline{\bar{C}}^0(m,\underline{p}) \leqslant \frac{1}{n} \sum_{j=1}^{n} \frac{n-j+m+1}{m} \underline{p}^{(j)} \ . \tag{4}$$

Let us calculate the expected value of the above lower bound rewritten as

$$\frac{1}{m}\sum_{j=1}^{n}\underline{p}_j-\frac{1}{nm}\sum_{j=1}^{n}(j-1)\underline{p}^{(j)}\ . \tag{5}$$

The expected value of the first term in (5) is equal to $n\mu/m$. The expected value of the second term is calculated as follows:

$$\begin{aligned}\sum_{j=1}^{n}(j-1)E\underline{p}^{(j)} &= n\int_0^\infty\sum_{j=1}^{n}(j-1)\binom{n-1}{j-1}F(x)^{j-1}(1-F(x))^{n-j}x\,\mathrm{d}F(x)\\ &= n(n-1)\int_0^\infty\sum_{k=0}^{n-2}\binom{n-2}{k}F(x)^k(1-F(x))^{n-2-k}xF(x)\mathrm{d}F(x)\\ &= n(n-1)\int_0^\infty xF(x)\mathrm{d}F(x).\end{aligned} \tag{6}$$

Now, as a heuristic choice m^H for m at the aggregate level we propose the value minimizing the lower bound on $E\underline{Z}(m,\underline{p})$ given by

$$cm+\frac{1}{m}\left(n\mu-(n-1)\int_0^\infty xF(x)\mathrm{d}F(x)\right)\ . \tag{7}$$

i.e. the most favorable integer round-off of

$$\left(\frac{n\mu-(n-1)\nu}{c}\right)^{1/2} \tag{8}$$

with

$$\nu\triangleq\int_0^\infty xF(x)\mathrm{d}F(x)\ .$$

Subsequently at the detailed level, we schedule the jobs on the m^H machines acquired using the SPT rule. Thus, the heuristic solution value is given by

$$\underline{Z}(m^H,\underline{p}) = cm^H + \underline{\bar{C}}^0(m^H,\underline{p}) \ . \tag{9}$$

We analyze the quality of this heuristic in the next section, and conclude this section by observing 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 $\nu = 3/(4\lambda)$.

3. Analysis of the heuristic

To analyze the asymptotic behaviour of the bounds in (4), we rewrite these inequalities as

$$\frac{1}{n^2 m} \sum_{j=1}^{n} \underline{p}_j + \frac{1}{nm} \sum_{j=1}^{n} \left(1 - \frac{j}{n}\right) \underline{p}^{(j)} \leqslant \frac{1}{n} \underline{\bar{C}}^0(m,\underline{p})$$

$$\leqslant \frac{m+1}{n^2 m} \sum_{j=1}^{n} \underline{p}_j + \frac{1}{nm} \sum_{j=1}^{n} \left(1 - \frac{j}{n}\right) \underline{p}^{(j)} \tag{10}$$

and observe that

$$\underline{T}_n \triangleq \frac{1}{n} \sum_{j=1}^{n} \left(1 - \frac{j}{n}\right) \underline{p}^{(j)} \tag{11}$$

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

$$\frac{1}{n} \sum_{j=1}^{n} J\left(\frac{j}{n}\right) \underline{p}^{(j)} , \tag{12}$$

with $J(t) = 1 - t$.

In the appendix we establish the following general almost surely (a.s.) convergence result for such statistics.

THEOREM 1.

If $J : [0,1] \rightarrow IR$ is a continuous function, then

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} J\left(\frac{j}{n}\right) \underline{p}^{(j)} = \int_0^\infty xJ(F(x))\mathrm{d}F(x) \qquad \text{(a.s.)}. \tag{13}$$

As a special case, we obtain that

$$\lim_{n \to \infty} \underline{T}_n = \mu - \nu \qquad \text{(a.s.)}. \tag{14}$$

To analyze the quality of our heuristic, we now compare the upper bound on $\underline{Z}(m^H, \underline{p})$ given by (cf. (10))

$$cm^H + \frac{n}{m^H} \underline{T}_n + \frac{1}{n} \frac{m^H + 1}{m^H} \sum_{j=1}^{n} \underline{p}_j \tag{15}$$

to the solution value that could be realized in the case of *perfect information*, i.e. in the case that the realization $(p_1, \ldots, p_n)$ is already known when the aggregate level decision has to be made. The number of machines to acquire then clearly depends on these values and may be written as $\underline{m}^0(\underline{p})$. From (4), we derive that

$$\underline{Z}(\underline{m}^0(\underline{p}), \underline{p}) \geqslant \min_m \left\{cm + \frac{n}{m} \underline{T}_n\right\} \tag{16}$$

and hence

$$1 \leqslant \frac{\underline{Z}(m^H, \underline{p})}{\underline{Z}(\underline{m}^0(\underline{p}), \underline{p})} \leqslant \frac{cm^H + \frac{n}{m^H} \underline{T}_n + \frac{1}{n} \frac{m^H + 1}{m^H} \sum_{j=1}^{n} \underline{p}_j}{2\sqrt{cn\underline{T}_n}} \tag{17}$$

From the definition of m^H (cf. (8)) and (14) we deduce that, almost surely,

$$\lim_{n \to \infty} \frac{cm^H}{2\sqrt{cn\underline{T}_n}} = \lim_{n \to \infty} \frac{1}{2} \left(\frac{c(n\mu - (n-1)\nu)}{cn(\mu - \nu)}\right)^{1/2} = \frac{1}{2} \tag{18}$$

and similarly that

$$\lim_{n \to \infty} \frac{\frac{n}{m^H} \underline{T}_n}{2\sqrt{cn\underline{T}_n}} = \frac{1}{2} \qquad \text{(a.s.)} . \tag{19}$$

Since, for n sufficiently large,

$$\frac{\frac{1}{n} \frac{m^H + 1}{m^H} \sum_{j=1}^{n} \underline{p}_j}{2\sqrt{cn\underline{T}_n}} \leqslant \frac{\frac{2}{n} \sum_{j=1}^{n} \underline{p}_j}{2\sqrt{cn\underline{T}_n}} , \tag{20}$$

and since the strong law of large numbers [1] implies that this latter term converges to 0 almost surely, we have arrived at the following result.

THEOREM 2

$$\lim_{n \to \infty} \frac{\underline{Z}(m^H,\underline{p})}{\underline{Z}(\underline{m}^0(\underline{p}),\underline{p})} = 1 \qquad \text{(a.s.)} . \tag{21}$$

Hence in the terminology of [11], the heuristic is *asymptotically clairvoyant almost surely:* the relative loss due to imperfect information indeed goes to 0 almost surely. In particular, under some additional boundedness conditions, this result implies [11] the following corollary indicating that the heuristic is also *asymptotically optimal in expectation*

COROLLARY 1

$$\lim_{n \to \infty} \frac{E\underline{Z}(m^H,\underline{p})}{E\underline{Z}(m^0,\underline{p})} = 1. \tag{22}$$

If the second moment $E\underline{p}_j^2$ may be assumed to be finite, it turns out that we can even establish the rate at which $\underline{Z}^H(m,\underline{p})/\underline{Z}^H(\underline{m}^0(\underline{p}),\underline{p})$ converges to 1, something that was not done in previous cases. For this purpose, we again make use of a general result that is established in the appendix.

THEOREM 3

If $J : [0,1] \to IR$ is a continuously differentiable function, then

$$\limsup_{n\to\infty} \frac{\left|\frac{1}{n}\sum_{j=1}^{n} J\left(\frac{j}{n}\right) \underline{p}^{(j)} - \int_0^\infty xJ(F(x))\mathrm{d}F(x)\right| \sqrt{n}}{\sqrt{\log\log n}} < \infty \quad \text{(a.s.)} \tag{23}$$

We use Theorem 3 to analyze the convergence of the ratio $\underline{Z}(m^H,\underline{p})/\underline{Z}(\underline{m}^0(\underline{p}),\underline{p})$. The right-hand side of inequality (17) is our starting point. From the definition of m^H we have

$$\frac{cm^H}{(cn\underline{T}_n)^{1/2}} - 1 = \frac{(c(n\mu-(n-1)\nu))^{1/2} - (cn\underline{T}_n)^{1/2}}{(cn\underline{T}_n)^{1/2}}$$

$$= \frac{\frac{1}{n}\nu + (\mu-\nu) - \underline{T}_n}{\underline{T}_n^{1/2}\left(\left(\mu - \frac{(n-1)}{n}\nu\right)^{1/2} + \underline{T}_n^{1/2}\right)}. \tag{24}$$

Hence, Theorem 3 applied once again to the special case that $J(t) = 1 - t$ yields that

$$\limsup_{n\to\infty} \left|\frac{cm^H}{(cn\underline{T}_n)^{1/2}} - 1\right| \left(\frac{n}{\log\log n}\right)^{1/2} < \infty \quad \text{(a.s.)}. \tag{25}$$

From (19) we have that

$$\lim_{n\to\infty} \frac{\dfrac{n\underline{T}_n}{m^H}}{(cn\underline{T}_n)^{1/2}} = 1 \quad \text{(a.s.)}. \tag{26}$$

We observe that

$$\frac{cm^H}{(cn\underline{T}_n)^{1/2}} = \frac{(cn\underline{T}_n)^{1/2}}{\frac{n\underline{T}_n}{m^H}} . \tag{27}$$

Together, (25), (26) and (27) imply

$$\limsup_{n \to \infty} \left| \frac{\frac{n\underline{T}_n}{m^H}}{(cn\underline{T}_n)^{1/2}} - 1 \right| \left(\frac{n}{\log\log n} \right)^{1/2}$$

$$= \limsup_{n \to \infty} \left| \frac{\frac{n\underline{T}_n}{m^H}}{(cn\underline{T}_n)^{1/2}} \right| \left| 1 - \frac{cm^H}{(cn\underline{T}_n)^{1/2}} \right| \left(\frac{n}{\log\log n} \right)^{1/2} < \infty \qquad \text{(a.s.)} \tag{28}$$

Finally

$$\limsup_{n \to \infty} \frac{\frac{2}{n} \sum_{j=1}^{n} \underline{p}_j}{(cn\underline{T}_n)^{1/2}} \left(\frac{n}{\log\log n} \right)^{1/2} = 0 \qquad \text{(a.s.)} \tag{29}$$

and we have arrived at the following strong extension of Theorem 2.

THEOREM 4

$$\limsup_{n \to \infty} \left| \frac{\underline{Z}(m^H, \underline{p})}{\underline{Z}(\underline{m}^0(\underline{p}), \underline{p})} - 1 \right| \left(\frac{n}{\log\log n} \right)^{1/2} < \infty \qquad \text{(a.s.)} . \tag{30}$$

We finally prove that not only the value of the heuristic but also the solution at the aggregate level itself almost surely converges to the optimal one. Indeed, we establish the rate of convergence in the following theorem.

THEOREM 5

$$\limsup_{n \to \infty} \left| \frac{\underline{m}^0(\underline{p})}{mH} - 1 \right| n^{1/4} < \infty \qquad \text{(a.s.)} . \tag{31}$$

Proof: We define the function

$$\underline{Z}^{LB}(m,\underline{p}) \triangleq cm + \frac{n\underline{T}_n}{m} , \tag{32}$$

which for fixed p is a unimodal function of m. It is minimized by

$$\underline{m}^{LB}(\underline{p}) \triangleq \left(\frac{n\underline{T}_n}{c}\right)^{1/2} . \tag{33}$$

We have

$$\underline{Z}^{LB}(\underline{m}^{LB},\underline{p}) \leqslant \underline{Z}(\underline{m}^0(\underline{p}),\underline{p}) \leqslant \underline{Z}^{LB}(\underline{m}^{LB},\underline{p}) + \frac{2}{n} \sum_{j=1}^{n} \underline{p}_j . \tag{34}$$

We now compute $\underline{m}_1$ and $\underline{m}_2$ such that

$$\underline{Z}^{LB}(\underline{m}_1,\underline{p}) = \underline{Z}^{LB}(\underline{m}_2,\underline{p}) = \underline{Z}^{LB}(\underline{m}^{LB},\underline{p}) + \frac{2}{n} \sum_{j=1}^{n} \underline{p}_j . \tag{35}$$

To do so, we solve the equality

$$cm + \frac{n\underline{T}_n}{m} = 2(cn\underline{T}_n)^{1/2} + \frac{2}{n} \sum_{j=1}^{n} \underline{p}_j \tag{36}$$

rewritten as

$$cm^2 - \left(2(cn\underline{T}_n)^{1/2} + \frac{2}{n} \sum_{j=1}^{n} \underline{p}_j\right) m + n\underline{T}_n = 0 , \tag{37}$$

to find two roots

$$\underline{m}_1 = \left(\frac{n\underline{T}_n}{c}\right)^{1/2} + \frac{1}{cn} \sum_{j=1}^{n} \underline{p}_j - \left(\frac{2}{cn} \sum_{j=1}^{n} \underline{p}_j (cn\underline{T}_n)^{1/2} + \left(\frac{1}{cn} \sum_{j=1}^{n} \underline{p}_j\right)^2 \right)^{1/2} \tag{38}$$

and

$$\underline{m}_2 = \left(\frac{n\underline{T}_n}{c}\right)^{1/2} + \frac{1}{cn}\sum_{j=1}^{n} \underline{p}_j + \left(\frac{2}{cn}\sum_{j=1}^{n} \underline{p}_j(cn\underline{T}_n)^{1/2} + \left(\frac{1}{cn}\sum_{j=1}^{n} \underline{p}_j\right)^2\right)^{1/2} . \tag{39}$$

The definitions of $\underline{m}_1$ and $\underline{m}_2$ and (34) imply that $\underline{m}_1 \leqslant \underline{m}^0(\underline{p}) \leqslant \underline{m}_2$ and hence

$$\frac{\underline{m}_1}{mH} \leqslant \frac{\underline{m}^0(\underline{p})}{mH} \leqslant \frac{\underline{m}_1}{mH} + \frac{\underline{m}_2 - \underline{m}_1}{mH} . \tag{40}$$

Now

$$\frac{(cn\underline{T}_n)^{1/2}}{cmH} + \frac{\left(\frac{2}{cn}\sum_{j=1}^{n} \underline{p}_j(cn\underline{T}_n)^{1/2}\right)^{1/2}}{mH} \leqslant \frac{\underline{m}_1}{mH} \leqslant \frac{(cn\underline{T}_n)^{1/2}}{cmH} + \frac{\left(\frac{1}{cn}\sum_{j=1}^{n} \underline{p}_j\right)}{mH} \tag{41}$$

and

$$\frac{\underline{m}_2 - \underline{m}_1}{mH} \leqslant \frac{2}{mH}\left(\frac{2}{cn}\sum_{j=1}^{n} \underline{p}_j(cn\underline{T}_n)^{1/2}\right)^{1/2} + \frac{2}{mH}\left(\frac{1}{cn}\sum_{j=1}^{n} \underline{p}_j\right) . \tag{42}$$

As

$$\lim_{n\to\infty} \left(\frac{2}{cn}\sum_{j=1}^{n} \underline{p}_j\right)^{1/2} \frac{(cn\underline{T}_n)^{1/4}}{n^{1/4}} < \infty \qquad \text{(a.s.)} , \tag{43}$$

(40), (41), (42) together with (19) and the strong law of large numbers imply the theorem.

4. Concluding remarks

In sect. 3 and the appendix an analysis is given of the stochastic programming heuristic for a hierarchical planning problem whose detailed level decision is easy rather than computationally intractable. The analysis is based on a sharp *a priori* estimate of the solution value produced by a simple *greedy-like* priority rule, an estimate that is based on results from the theory of *order statistics.* This theory may be of similar use in analyzing the performance of other greedy-like solution methods (for an analysis in the classical case of the minimum spanning tree, see [8]). As in the case treated here, such results might find natural application in the context of hierarchical problems as well.

References

[1] P. Billingsley, *Probability and Measure* (John Wiley and Sons, New York, 1979).

[2] R.W. Conway, W.L. Maxwell and L.W. Miller, *Theory of Scheduling* (Addison-Wesley, 1967).

[3] K.L. Chung, An estimate concerning the Kolmogorov limit distribution, Trans. Amer. Math. Soc. 67(1949)36.

[4] M.A.H. Dempster, *A Stochastic Approach to Hierarchical Planning and Scheduling,* in: *Deterministic and Stochastic Scheduling,* ed. M.A.H. Dempster, J.K. Lenstra and A.H.G. Rinnooy Kan (D. Reidel Publishing Company, 1982) 271.

[5] M.A.H. Dempster, M.L. Fisher, L. Jansen, B.J. Lageweg, J.K. Lenstra and A.H.G. Rinnooy Kan, Analytical evaluation of hierarchical planning systems, Oper. Res. 29(1981)707.

[6] M.A.H. Dempster, M.L. Fisher, L. Jansen, B.J. Lageweg, J.K. Lenstra and A.H.G. Rinnooy Kan, Analysis of heuristics for stochastic programming: results for hierarchical scheduling problems, Math. Oper. Res. (1983)

[7] W. Feller, *An Introduction to Probability Theory and Its Applications,* vol. 1, 3rd edition (John Wiley and Sons, New York, 1968).

[8] A.M. Frieze, On the value of a random minimum spanning tree problem, Techn. Rep. (Dept. of Computer Sci. and Statistics, University of London, 1982).

[9] M.R. Garey and D.S. Johnson, *Computers and Intractibility: A Guide to the Theory of NP-Completeness* (Freeman, San Francisco, 1979).

[10] P. Hartman and A. Wintner, On the law of the iterated logarithm, Amer. J. of Math. 63 (1941)169.

[11] J.K. Lenstra, A.H.G. Rinnooy Kan and L. Stougie, A framework for the probabilistic analysis of hierarchical planning systems, to be submitted for publication.

[12] A. Marchetti Spaccamela, A.H.G. Rinnooy Kan and L. Stougie, Hierarchical vehicle routing problems, Networks (1984).

Appendix

Proof of Theorem 1

If we denote the *empirical distribution function* by

$$\underline{F}_n(x) \triangleq \frac{1}{n} |\{j | \underline{p}_j \leqslant x\}| , \tag{A.1}$$

then

$$\underline{U}_n \triangleq \frac{1}{n} \sum_{j=1}^{n} J\left(\frac{j}{n}\right) \underline{p}^{(j)} = \int_0^\infty xJ(\underline{F}_n(x)) \mathrm{d}\underline{F}_n(x) . \tag{A.2}$$

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

$$\underline{F}_n(F^{-1}(y)) = \frac{1}{n} |\{j | F(\underline{p}_j) \leqslant y\}| . \tag{A.3}$$

However, $\underline{v}_j = F(\underline{p}_j)$ is uniformly distributed on [0, 1] [7] and hence

$$\underline{F}_n(F^{-1}(y)) = \underline{V}_n(y) , \tag{A.4}$$

where $\underline{V}_n(y)$ is the empirical distribution function of n uniformly, independently distributed random varianbles. Thus, if we substitute $x = F^{-1}(y)$ in (A.2), we obtain that

$$\begin{aligned} \underline{U}_n &= \int_0^1 F^{-1}(y) J(\underline{V}_n(y)) \mathrm{d}\underline{V}_n(y) \\ &= \int_0^1 F^{-1}(y)(J(\underline{V}_n(y)) - J(y)) \mathrm{d}\underline{V}_n(y) + \int_0^1 F^{-1}(y) J(y) \mathrm{d}\underline{V}_n(y) . \end{aligned} \tag{A.5}$$

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

$$\lim_{n \to \infty} \sup_{y \in [0,1]} |\underline{V}_n(y) - y| = 0 \qquad \text{(a.s.)} \tag{A.6}$$

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

$$\limsup_{n \to \infty} \left| \frac{\int_0^1 F^{-1}(y)(J(\underline{V}_n(y)) - J(y))\,\mathrm{d}\underline{V}_n(y)}{\int_0^1 F^{-1}(y)\,\mathrm{d}\underline{V}_n(y)} \right| \leqslant \epsilon \qquad \text{(a.s.)}. \tag{A.7}$$

Because of the strong law of large numbers [1, p. 250], the denominator in (A.7) converges to μ (a.s.), and hence

$$\limsup_{n \to \infty} \left| \int_0^1 F^{-1}(y)(J(\underline{V}_n(y)) - J(y))\,\mathrm{d}\underline{V}_n(y) \right| = 0 \qquad \text{(a.s.)}. \tag{A.8}$$

We again invoke the strong law of large numbers to analyze the second term on the right-hand side of (A.5)

$$\begin{aligned} \lim_{n \to \infty} \int_0^1 F^{-1}(y)J(y)\,\mathrm{d}\underline{V}_n(y) &= \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} F^{-1}(\underline{v}_i)J(\underline{v}_i) \\ &= E(F^{-1}(\underline{v}_i)J(\underline{v}_i)) = \int_0^1 F^{-1}(y)J(y)\,\mathrm{d}y \quad \text{(a.s.)}. \end{aligned} \tag{A.9}$$

Together (A.8) and (A.9) imply the theorem.

Proof of Theorem 3

Using (A.5), we write

$$\begin{aligned} \underline{U}_n - \int_0^1 F^{-1}(y)J(y)\,\mathrm{d}y &= \int_0^1 F^{-1}(y)(J(\underline{V}_n(y)) - J(y))\,\mathrm{d}\underline{V}_n(y) \\ &\quad + \int_0^1 F^{-1}(y)J(y)\,\mathrm{d}\underline{V}_n(y) - \int_0^1 F^{-1}(y)J(y)\,\mathrm{d}y \end{aligned} \tag{A.10}$$

and analyze the right-hand side of (A.10) 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

$$J(\underline{V}_n(y)) - J(y) = J'(\underline{W}_n(y))(\underline{V}_n(y) - y), \tag{A.11}$$

with

$$\underline{W}_n(y) \triangleq \theta \underline{V}_n(y) + (1-\theta)y. \tag{A.12}$$

Since $\underline{V}_n(y)$ is an increasing function and $F^{-1}(y) \geqslant 0$, we may conclude, after substitution of (A.11) in the first term of the right-hand side of (A.10), that

$$\left| \int_0^1 F^{-1}(y)(J(\underline{V}_n(y)) - J(y)) \mathrm{d}\underline{V}_n(y) \right|$$
$$\leqslant \sup_{y \in [0,1]} |\underline{V}_n(y) - y| \int_0^1 F^{-1}(y) \, |J'(\underline{W}_n(y))| \, \mathrm{d}\underline{V}_n(y). \tag{A.13}$$

Now, since F is continuous ([3])

$$\limsup_{n \to \infty} \frac{\sqrt{n} \limsup_{n \to \infty} |\underline{V}_n(y) - y|}{\sqrt{2 \log\log n}} = \frac{1}{2} \qquad \text{(a.s.)} \tag{A.14}$$

Furthermore, there exists a constant M such that

$$\int_0^1 F^{-1}(y) \, |J'(\underline{W}_n(y))| \, \mathrm{d}\underline{V}_n(y) \leqslant M \int_0^1 F^{-1}(y) \mathrm{d}\underline{V}_n(y), \tag{A.15}$$

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

$$\lim_{n \to \infty} \frac{1}{\mu} \int_0^1 F^{-1}(y) \mathrm{d}\underline{V}_n(y) = 1. \tag{A.16}$$

Hence,

$$\limsup_{n \to \infty} \frac{\left| \int_0^1 F^{-1}(y)(J(\underline{V}_n(y)) - J(y)) \mathrm{d}\underline{V}_n(y) \right| \sqrt{n}}{\sqrt{\log\log n}} < \infty \qquad \text{(a.s.)} \tag{A.17}$$

The last two terms of the right-hand side of (A.10) can be rewritten as

$$\frac{1}{n} \sum_{j=1}^{n} F^{-1}(\underline{v}_i)J(\underline{v}_i) - \int_0^1 F^{-1}(y)J(y)\,\mathrm{d}y\,. \tag{A.18}$$

If $E\underline{p}_j^2 < \infty$, we may apply the law of the iterated logarithm [10] to find that

$$\limsup_{n \to \infty} \frac{\left| \int_0^1 F^{-1}(y)J(y)\,\mathrm{d}\underline{V}_n(y) - \int_0^1 F^{-1}(y)J(y)\,\mathrm{d}y \right| \sqrt{n}}{\sqrt{\log\log n}} < \infty \quad \text{(a.s.)} \tag{A.19}$$

Together (A.17) and (A.19) imply the theorem.