Available online at www.sciencedirect.com
SCIENCE DIRECT®
European Journal of Operational Research 155 (2004) 439–454
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/dsw

Discrete Optimization

Case study on statistically estimating minimum makespan for flow line scheduling problems

Amy D. Wilson [a], Russell E. King [b,*], James R. Wilson [b]

[a] *Wyeth Vaccines, Sanford, NC 27330, USA*
[b] *Department of Industrial Engineering, North Carolina State University, P.O. Box 7906, Raleigh, NC 27695-7906, USA*

Received 2 January 2002; accepted 18 November 2002

Abstract

Lower bounds are typically used to evaluate the performance of heuristics for solving combinatorial minimization problems. In the absence of tight analytical lower bounds, optimal objective-function values may be estimated statistically. In this paper, extreme value theory is used to construct confidence-interval estimates of the minimum makespan achievable when scheduling nonsimilar groups of jobs on a two-stage flow line. Experimental results based on randomly sampled solutions to each of 180 randomly generated test problems revealed that (i) least-squares parameter estimators outperformed standard analytical estimators for the Weibull approximation to the distribution of the sample minimum makespan; (ii) to evaluate each Weibull fit reliably, both the Anderson–Darling and Kolmogorov–Smirnov goodness-of-fit tests should be used; and (iii) applying a local improvement procedure to a large sample of randomly generated initial solutions improved the probability that the resulting Weibull fit yielded a confidence interval covering the minimum makespan.

Keywords: Heuristics; Scheduling; Confidence-interval estimation; Flexible flow line

1. Introduction

Because of the difficulty in easily obtaining optimal solutions to large combinatorial optimization problems, heuristics are often employed to obtain solutions that are hoped to be near-optimal. It is a challenge, however, to assess how close to optimality the solutions of a heuristic are because optimal solutions are not readily available for the problems for which a heuristic is justified. An upper bound on a heuristic solution's deviation from optimality can be obtained by comparing that solution's objective-function value to a lower bound on the optimal objective-function value for a minimization problem or an upper bound on the optimal objective-function value for a maximization problem. Ideally, such bounds should be easy to compute and very close to the optimal value of the objective function. Unfortunately, tight analytical (or easily computable) bounds on the optimal objective-function value for complex combinatorial optimization problems are often difficult to develop because the formulation of

* Corresponding author. Tel.: +1-919-515-5186; fax: +1-919-515-1543.
E-mail address: king@eos.ncsu.edu (R.E. King).

0377-2217/$ - see front matter
doi:10.1016/S0377-2217(02)00910-4

such bounds typically require relaxations of some of the complications of the problem that give it its true character. In this situation, a better idea of heuristic performance may be ascertained by comparison to a statistical estimate of the optimal objective-function value.

A $100(1-\alpha)\%$, $0 < \alpha < 1$, confidence interval for the optimal objective-function value z^* of a minimization problem can be calculated from a set of objective-function values corresponding to solutions of the problem that are randomly sampled according to a probability distribution defined on the problem's feasible region. The resulting confidence interval for z^* usually has the form $[\hat{z}^*, z_{\min}]$, where $z_{\min}$ is the minimum value observed in the sample. As discussed in Derigs (1985), there are two main approaches for estimating the minimum value of a target population and for constructing a confidence interval for that value. The first of these techniques, the limiting distribution approach, constructs a confidence interval for z^* using the order statistics of a random sample of size m, $\{z_i : i = 1, \ldots, m\}$, taken from the target population. This approach yields a lower limit of the $100(1-\alpha)\%$ confidence interval for z^* having the form $\hat{z}^* = z_{[1]} - C_\alpha(z_{[k]} - z_{[1]})$, where $z_{[1]} \leqslant z_{[2]} \leqslant \cdots \leqslant z_{[m]}$ are the order statistics of the sample; $1 \leqslant k \leqslant m$; and C_α is a positive constant. The upper limit of the confidence interval is $z_{\min} = z_{[1]}$. Different values of C_α and k are derived by Robson and Whitlock (1964), Van der Watt (1980), and Boender et al. (1982).

The second of these approaches, based on extreme value theory, constructs a confidence-interval estimate of the minimum value of the target population using the asymptotic distribution of extreme values as derived by Fisher and Tippett (1928). Consider n independent random samples each of size m taken from the target population. Notice that the smallest value in the overall sample consisting of nm observations may be regarded as the smallest of the n minimum values taken from each of the original samples of size m. Thus as m gets large, the smallest value in each sample of size m, when suitably standardized, should tend to the same asymptotic distribution as for the smallest value in a sample of size nm, when suitably standardized—provided that such a limiting distribution exists when random samples are taken from the given target population. The asymptotic distribution, if it exists, is characterized by the following *min-stable property*: the asymptotic distribution is nondegenerate; and when the minimum $x_{[1]}$ in a random sample $\{x_i : i = 1, \ldots, n\}$ of size n taken from that asymptotic distribution is standardized using appropriate constants $\lambda_n > 0$ and ξ_n, the standardized random variable $\lambda_n(x_{[1]} - \xi_n)$ has the same distribution as the original n observations (Leadbetter et al., 1983, Theorems 1.4.1 and 1.4.2).

Fisher and Tippett (1928) derive three possible forms for the asymptotic distribution using the functional relation that it must satisfy based on the min-stable property. If the target population has a finite minimum value (or "threshold") and Gnedenko's necessary and sufficient condition holds for that target population (Leadbetter et al., 1983, Theorem 1.6.2), then as m gets large the asymptotic distribution of the smallest value in a random sample of size m from the target population is a three-parameter Weibull distribution. Dannenbring (1977) notes that because of the discrete nature of combinatorial optimization problems, there is no guarantee that Gnedenko's condition can be satisfied, especially in the presence of adverse tail behavior. Although the Fisher–Tippett theorem requires a continuous target distribution, Ovacik et al. (2000), Derigs (1985), and Los and Lardinois (1982) assert that this continuity requirement may be approximately satisfied for many combinatorial optimization problems because of the large number of possible solutions. Gumbel (1958) provides a more detailed discussion of the Fisher–Tippett theorem and the conditions for its application.

The three-parameter Weibull distribution is defined by the threshold (or location) parameter, a, the scale parameter, b, and the shape parameter, c. The probability density function (p.d.f.) and the cumulative distribution function (c.d.f.) of the three-parameter Weibull distribution are respectively given by

$$f(x) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{c-1} \exp\left\{-\left(\frac{x-a}{b}\right)^{c}\right\}$$
$$\text{and} \quad F(x) = 1 - \exp\left\{-\left(\frac{x-a}{b}\right)^{c}\right\}, \tag{1}$$

for all $x \geqslant a$, where $b > 0$ and $c > 0$; and for all $x < a$, $f(x) = 0$ and $F(x) = 0$.

In practice, the exact parameters of the appropriate Weibull distribution are not known and must be estimated from a sample. Given fitted parameters $\hat{a}$, $\hat{b}$, and $\hat{c}$, estimated from a random sample $\{x_i : i = 1, \ldots, n\}$ of size n taken from the three-parameter Weibull distribution (1), Golden and Alt (1979) derived an approximate $100(1 - e^{-n})\%$ confidence interval for the threshold parameter a having the form $[\hat{x}^*, x_{\min}]$, where the lower limit of this confidence interval is given by $\hat{x}^* = x_{\min} - \hat{b}$. In theory, $\hat{x}^*$ can be used as a lower bound for the Weibull threshold parameter a; and if each x_i is actually generated as the sample minimum for a random sample of sufficiently large size m taken from a target population with minimum value z^*, then $a = z^*$ and $\hat{x}^*$ can be used as a statistical lower bound for z^*.

In this paper, confidence-interval estimation methods along the lines described above are developed and applied to the problem of scheduling nonsimilar groups of jobs on a two-stage flow line to minimize makespan. The motivation for statistical estimation of the minimum makespan was the observation that analytical lower bounds developed on the optimal objective-function value for this problem were not tight, and the quality of these lower bounds worsened as problem size increased.

The rest of this paper is organized as follows. Previous comparative studies of various confidence-interval estimation methods are discussed in Section 2. The particular problem considered in this research is outlined in Section 3, and in Section 4 the numerical problem instances used in this research are described. Section 5 contains a complete discussion of the methods used in this paper to estimate the Weibull parameters and to test the goodness of fit of the Weibull approximation to the empirical distribution of the sample data. In Section 6, a local improvement procedure is outlined, and the effect of its incorporation in the sampling method on the performance of the confidence-interval estimator for the problem's minimum makespan is discussed. Section 7 provides additional insight into the performance of the confidence-interval estimation methods developed in this paper by comparing the constructed confidence intervals to near-optimal solutions obtained from a genetic algorithm. Finally, conclusions and recommendations for future research are provided in Section 8.

2. Comparative studies of interval estimation methods

Golden and Alt (1979) compared their confidence-interval estimation method to the limiting distribution approach of Robson and Whitlock (1964) by applying both techniques to the traveling salesman problem. Golden and Alt found that their confidence interval always covered (contained) the optimal objective-function value while the Robson–Whitlock confidence interval failed to cover the optimal objective-function value on two out of six problems tested. Derigs (1985) evaluated the limiting distribution approaches of Robson and Whitlock (1964), Boender et al. (1982), and Van der Watt (1980) as well as the extreme value theory approaches of Golden and Alt (1979) and Los and Lardinois (1982). Derigs applied these confidence-interval estimation procedures to both the traveling salesman problem and the quadratic assignment problem. He found that the only confidence interval to cover the optimal objective-function value on all problems tested was that of Golden and Alt. More recently Ovacik et al. (2000) sought to minimize $L_{\max}$, the maximum job lateness, when scheduling jobs on a single machine with sequence dependent set-up times; and for this problem, they found that the method of Golden and Alt outperformed the limiting distribution approach of Boender et al. in terms of confidence-interval coverage probability.

There is potential for confidence intervals generated using the method of Golden and Alt to be too wide to be practically meaningful. However, for the problems studied in this paper, it is essential to have a reliable confidence-interval estimation method since the optimal objective-function values for these problems are not known. In view of the findings of the previous studies cited above, the focus in this paper is on using extreme value theory together with the confidence-interval estimation method of Golden and Alt to yield a

reliable statistical lower bound on the minimum makespan for the given scheduling problem.

3. Problem description

The problem of scheduling nonsimilar groups of jobs on a two-stage flow line is motivated by the cutting and sewing operations of upholstered furniture manufacturing. These operations occur in two stages. Pieces of fabric are cut at the first stage and sewn together at the second. Each order, or job, processed through the flow line represents a particular combination of fabric type and style. To minimize changeovers at each stage of the line, jobs are grouped by fabric type at cutting and by style at sewing. The composition of the groups at each stage is different, and the groups are therefore referred to as "nonsimilar".

Each job is characterized by its processing times at the two stages, the first-stage group to which it belongs, and the second-stage group to which it belongs. There are multiple machines in parallel at each of the two stages of the line. Groups cannot be split across machines at either stage, and jobs must be processed within their prespecified groups at each stage. On each machine at each stage, jobs within a group cannot start processing until all jobs in the previous group have completed processing. At the first stage, jobs within each group and all groups are processed sequentially with no idle time. Jobs proceed individually from the first stage to the second. On each machine at the second stage, a group may begin processing prior to the arrival of all its jobs from the first stage; however, each group's start time at its second-stage machine must be such that once the group begins processing, all jobs in that group may be processed without incurring any idle time at the second-stage machine. The objective of the scheduling is to minimize makespan for a given set of jobs.

The solution of this problem is NP-complete. When each group at each stage contains a single job, this problem reduces to the flow line problem with parallel machines at each stage, which has been shown to be NP-complete even when there are only two stages with multiple machines at only one stage (Gupta, 1988). The nature of the complexity of this problem can be clearly understood by considering the four elements that are needed to define a complete solution: (i) the assignment of groups to machines at the first stage, (ii) the sequence of groups on each machine at the first stage, (iii) the sequence of jobs within each group at the first stage, and (iv) the assignment of groups to machines at the second stage. Given an assignment of groups to machines at the second stage, makespan is minimized at the second stage by processing the groups in increasing order of earliest start time, processing jobs within each group in the order in which they completed at the first stage. See Wilson (2001) for a more complete discussion of this problem and its complexity.

Because of the complex combinatorial nature of this scheduling problem, a randomized search heuristic was developed to approximate its solution. Deterministic lower bounds on the minimum makespan were developed and tested for this problem in order to evaluate the heuristic solutions; but these lower bounds were not tight, and the quality of these lower bounds worsened as the problem size increased. Therefore, statistical estimation of the minimum makespan was considered in order to more accurately determine the deviation from optimality of the makespan resulting from heuristic solutions. Details of the heuristic and the evaluation of the deterministic lower bounds can be found in Wilson (2001). Of the 180 test problems considered in this research, the randomized search heuristic found the best known solution for 153 problems. The remaining 27 best known solutions were found with the local improvement procedure discussed in Section 6.1 and Appendix A.

4. Experimental design

The method used to generate data for experimentation was designed to capture the essential characteristics of the original application to upholstered furniture manufacturing. For each style, a random number of fabric types is generated. Each unique fabric type and style combination is represented by a job. It is assumed that processing requirements at each stage are dependent more on

Table 1
Problems used in experimentation

N_1	N_2	No. of jobs; [Avg., Range]
10	10	57; [45, 74]
10	15	86; [80, 93]
30	15	220; [180, 244]
15	30	228; [204, 260]

style than on fabric type. Therefore, first- and second-stage processing times, p_{1j} and p_{2j}, are generated for each style j. In order to ensure that not all jobs within a given style have the same processing time, a quantity multiplier, qty, is generated randomly for each job. Jobs of style j have processing time equal to $p_{1j} \times \text{qty}$ at the first stage and equal to $p_{2j} \times \text{qty}$ at the second stage.

The selected combinations of N_1, the number of first-stage groups (also the number of fabric types), and N_2, the number of second-stage groups (also the number of styles), are summarized in Table 1 along with the mean and range of the number of jobs actually generated. For each combination of N_1 and N_2, M_1 and M_2, the number of machines at the first and second stage, respectively, were set so that $M_1 = M_2 = 2$ and $\text{qty} \sim U[1, 5]$, where in general $U[\beta_1, \beta_2]$ denotes the (continuous) uniform distribution on the interval with left-hand endpoint β_1 and right-hand endpoint β_2. The processing times p_{1j} and p_{2j} were generated randomly over all combinations of the intervals [1, 10], [5, 15], and [10, 20] for each problem size. Five instances of each problem were generated for a total of 180 problems. For problems with two machines at each of the two stages, the total number of possible solutions to the scheduling problem is given by

$$(\lfloor N_1/2 \rfloor N_1!)\left(\prod_{i=1}^{N_1} \text{njob}_i!\right)\left\{\sum_{j=1}^{\lfloor N_2/2 \rfloor} \binom{N_2}{j}\right\}, \tag{2}$$

where njob_i is the number of jobs in first-stage group i for $i = 1, \ldots, N_1$.

5. Fitting the parameters of the Weibull distribution

Without knowing the true c.d.f. of the target population of makespans associated with all feasible solutions of a given instance of the scheduling problem, Gnedenko's necessary and sufficient conditions for the existence of a limiting distribution for the suitably standardized sample minimum makespan could not be verified. If the target population is continuous with p.d.f. $g(z)$, corresponding c.d.f. $G(z) = \int_{-\infty}^{z} g(u)\,\mathrm{d}u$ for all real z and minimum value $z^* \equiv \min\{z : G(z) > 0\}$, then a more easily verified sufficient condition for a suitably standardized sample minimum makespan $\delta_m(z_{[1]} - \zeta_m)$ to converge in distribution to a Weibull random variable as the sample size m tends to infinity is that

$$\lim_{z \to z^*+} \frac{(z - z^*)g(z)}{G(z)} = r, \tag{3}$$

for some constant $r > 0$; see Theorem 1.6.1 of Leadbetter et al. (1983). On the other hand, if there is a nonzero probability mass concentrated at z^* so that $G(z) = 0$ for $z < z^*$ and $G(z^*) > 0$, then it is known that there is no standardized version $\delta_m(z_{[1]} - \zeta_m)$ of the sample minimum makespan $z_{[1]}$ that converges to *any* limiting nondegenerate distribution, including a three-parameter Weibull distribution, as the sample size m tends to infinity; see Corollary 1.5.2 of Leadbetter et al. Unfortunately, this is precisely the situation with all combinatorial minimization problems. Thus, in general, the three-parameter Weibull distribution (1) is at best an approximation to the distribution of the sample minimum objective-function value $z_{[1]}$.

For given values of m and n and for a given instance of the two-stage flow line scheduling problem, a set of n sample minimum makespans $\{x_i : i = 1, \ldots, n\}$ is computed from nonoverlapping (mutually exclusive) random samples each of size m; and the latter random samples are taken from the target population of makespans with minimum value z^* and c.d.f. $G(z)$. In view of the discussion in the preceding paragraph, it was necessary to verify the three-parameter Weibull c.d.f. $F(x)$ specified by (1) can yield an adequate approximation to the empirical c.d.f. of the sample data $\{x_i : i = 1, \ldots, n\}$,

$$F_n(x) \equiv \frac{\#\{x_i : x_i \leqslant x \text{ and } i \leqslant n\}}{n} \quad \text{for all real } x, \tag{4}$$

when suitable estimates $\hat{a}$, $\hat{b}$, and $\hat{c}$ of the Weibull parameters are computed from the $\{x_i\}$. In order to verify the adequacy of the Weibull approximation to the empirical c.d.f. (4), large samples of completely random solutions to the scheduling problem were generated. In conformance with the requirements of the Fisher–Tippett theorem, a total sample of size nm was generated where $n = m = 500$.

Though the solutions to each given instance of the scheduling problem (and therefore the n sample minimum values of makespan) were generated independently of one other, in general those solutions (and their corresponding makespan values) are stochastically dependent because they are all constructed to solve the same instance of the scheduling problem, which was also randomly generated according to the problem-generation scheme detailed in Section 4. The real issue here is whether the degree of stochastic dependence between the resulting observations of makespan is sufficiently strong to invalidate a distribution-fitting scheme that is based on the assumption of independent and identically distributed observations of sample minimum makespan. Therefore, to verify the "approximate independence" of the $n = 500$ sample minimum values of makespan each based on a random sample of $m = 500$ solutions of the given scheduling problem, the "runs test" was used as described in Law and Kelton (1991). If the $n = 500$ sample minimum values of makespan passed the runs test for independence at a confidence level of 0.95, then a Weibull distribution was fitted to this data set.

Various methods have been used to fit the parameters of the Weibull distribution to a set of sample minimum objective-function values. Dannenbring (1977) used both the method of moments and the method of least squares. Golden and Alt (1979), Los and Lardinois (1982), Derigs (1985), and Gonsalvez et al. (1987) solved the maximum likelihood equations using the gradient search technique of Harter and Moore (1965, 1967). However, Derigs noted that this procedure can have convergence difficulties when the true shape parameter, c, is greater than or equal to three. Vasko and Wilson (1984) used nonlinear regression to minimize the sum of squared differences between the empirical and fitted c.d.f.'s. Rather than using the standard definition (4) of the empirical c.d.f., Vasko and Wilson took $F_n(x_{[i]}) = (i-d)/(n-2d+1)$, for $i = 1, \ldots, n$, where d is a real number in the interval [0, 0.5].

Analytical estimators of Weibull parameters were also used by several authors. Zanakis (1979) found that the following analytical parameter estimators had the best performance in practice:

$$\hat{a} = \frac{x_{[1]}x_{[2]} - x_{[2]}^2}{x_{[1]} + x_{[n]} - 2x_{[2]}}, \tag{5}$$

$$\hat{b} = x_{[\lceil 0.63n \rceil]} - \hat{a}, \tag{6}$$

$$\hat{c} = \frac{\ln\{\ln(1-q_1)/\ln(1-q_2)\}}{\ln\{(x_{[\lceil nq_1 \rceil]} - \hat{a})/(x_{[\lceil nq_2 \rceil]} - \hat{a})\}}, \tag{7}$$

where $q_1 = 0.97366$ and $q_2 = 0.16731$. These estimators were used by Ovacik et al. (2000). Bruijs (1984) used the analytical estimator (6) for b but used the analytical estimator for a derived by Cooke (1979). Regardless of the method used to fit the parameters of the Weibull distribution, the Kolmogorov–Smirnov (KS) test (e.g. Law and Kelton, 1991) has been used exclusively in all previously cited work to evaluate goodness of fit for the resulting Weibull distribution.

Initially, it was attempted to use the Harter–Moore technique to estimate the Weibull parameters. However, convergence difficulties were encountered and the technique was abandoned. Instead, the analytical parameter estimators (5)–(7) of Zanakis were compared against a least-squares procedure for estimating the Weibull c.d.f. (1). This approach to least-squares estimation was selected because of its successful implementation in other distribution-fitting applications involving estimation of threshold parameters as well as scale and shape parameters (Swain et al., 1988; Wagner and Wilson, 1996). The following sum of squares was minimized as a function of $\hat{a}$, $\hat{b}$, and $\hat{c}$ using the Nelder–Mead simplex search procedure (Olsson, 1974; Olsson and Nelson, 1975), with starting values given by Eqs. (5)–(7):

$$SS(\hat{a}, \hat{b}, \hat{c}) \equiv \sum_{i=1}^{n} \left\{ \hat{F}(x_{[i]}; \hat{a}, \hat{b}, \hat{c}) - \frac{i}{n+1} \right\}^2, \tag{8}$$

where $\hat{F}(x;\hat{a},\hat{b},\hat{c})$ for all real x denotes the c.d.f. of the fitted Weibull distribution with threshold parameter $\hat{a}$, scale parameter $\hat{b}$, and shape parameter $\hat{c}$. The basis for applying the principle of least squares to (8) to fit a Weibull distribution to the $\{x_i\}$ is detailed in Swain et al. (1988, Sections 1–2). Notice that this approach to least-squares estimation of Weibull parameters coincides in principle with the method of Vasko and Wilson (1984) if $d = 0$ is used in their formulation of the empirical c.d.f.; on the other hand, Vasko and Wilson neither specified nor justified the value of d used in their study.

In the Nelder–Mead search procedure used to minimize the sum of squares (8), the initial step size for each parameter was set equal to 20% of the magnitude of the corresponding initial parameter estimate. The estimate $\hat{a}$ of the Weibull threshold parameter was also constrained to be greater than or equal to the deterministic lower bound for the given scheduling problem. The reflection, expansion, and contraction coefficients of the Nelder–Mead procedure were set equal to 1.0, 2.0, and 0.5, respectively. For more information on the performance of the Nelder–Mead search procedure on continuous optimization problems (such as the estimation of Weibull parameters), see Humphrey and Wilson (2000).

Because graphical comparison of the empirical c.d.f.'s and the fitted c.d.f.'s produced by the analytical parameter estimators indicated that the fitted distributions would pass the KS test despite significant differences between the empirical and fitted c.d.f.'s in the tail regions, the Anderson–Darling (AD) goodness-of-fit test (e.g. Law and Kelton, 1991) was also used. The AD test, being specifically designed to detect discrepancies between empirical and fitted distributions in the tail areas, has a greater ability to detect those cases in which the fitted distribution does not adequately track the tail regions of the empirical distribution. A confidence level of 0.95 was used for both goodness of fit tests.

Of the 180 problems examined, 11 did not pass the independence test. On all of the 169 problems for which a set of $n = 500$ "approximately" independent sample minimum values of makespan were obtained, the least-squares procedure produced a Weibull fit that passed both the KS and AD goodness-of-fit tests. The fits yielded by the analytical parameter estimators, however, failed the KS test 4 times; and they failed the AD test an additional 19 times for a total of 23 rejected fits. In addition, the least-squares procedure produced fits that had lower values for both goodness-of-fit statistics on 165 out of the 169 problems.

In general it was observed that the primary failing of the Weibull fit based on the analytical parameter estimators was the inability of the fitted distribution to adequately approximate the tail regions of the empirical distribution of $n = 500$ sample minimum makespans. Fig. 1 shows the empirical c.d.f. along with the Weibull c.d.f.'s fitted using the analytical and least-squares parameter estimators for a representative problem. This figure is illustrative of the general tendency for the Weibull fits based on analytical parameter estimators to underestimate the left-hand tail of the empirical distribution. This is a concern especially in the current context because it is the extreme left-hand tail of the empirical distribution (and of the underlying theoretical c.d.f.) that is of particular interest. Though these comparative results are expected since analytical estimators of Weibull parameters should be viewed as rough approximations at best, it is useful to note that the analytical parameter estimators worked well as starting values for the least-squares procedure. Fig. 1 also motivates the use of the AD test in addition to the KS test for these problems.

Because the Weibull fit based on analytical parameter estimators tended to inadequately approximate the left-hand tail of the associated empirical distribution, the resulting confidence interval was also frequently questionable. For the scheduling problem graphed in Fig. 1, the lower limit of the Golden–Alt confidence interval based on the analytical parameter estimators was $\hat{z}^* = \hat{x}^* = x_{\min} - \hat{b} = 4629$. The lower limit of the Golden–Alt confidence interval as determined by the least-squares parameter estimators, however, was $\hat{z}^* = 4272$. The best known makespan for the problem, 4709, is greater than both lower limits. However, the confidence-interval lower limit based on analytical parameter estimators suggests that the best known makespan is only 80 time units

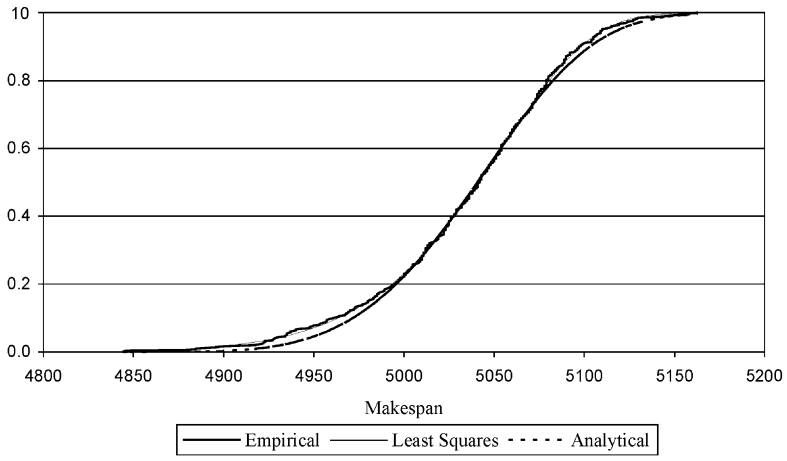

Fig. 1. C.d.f. for a selected problem with $N_1 = 15$, $N_2 = 30$, and $p_1 = p_2 = [5, 15]$.

away from the optimal makespan while the confidence-interval lower limit based on the least-squares parameter estimators indicates that the best known makespan may be up to 437 time units away from the optimal makespan. This general pattern, which was consistent for all problems, demonstrates the effect of using Weibull parameter estimators that do not produce a sufficiently accurate approximation to the empirical distribution of sample minimum makespans in order to yield a reliable confidence interval for z^*.

For all the problems for which a set of "approximately" independent sample minimum makespans was generated, an excellent Weibull fit was produced by the method of least squares. Also, when using the least-squares estimate $\hat{b}$ of the Weibull scale parameter b, the best known solution was contained in *every* confidence interval generated using the method of Golden and Alt. On the basis of these results, it was concluded that the Fisher–Tippett result did indeed apply at least as an adequate approximation for this problem and could be used to generate sufficiently accurate confidence-interval estimates of the minimum makespan for the given scheduling problem.

6. Effect of the method of sampling

A criticism of generating confidence intervals using completely random samples is that a random sample may not adequately penetrate the tail areas of the distribution and will therefore provide an interval that overestimates the optimal solution (for minimization problems). Dannenbring (1977) saw this effect in the point estimators he studied for the optimal solution to a flow shop problem. He found that the optimal solution was most frequently overestimated for those problems for which a random sample did not adequately penetrate the tail region. Bruijs (1984) similarly generated interval estimates for the quadratic assignment problem using random samples and found that they too had a tendency to overestimate the best known solutions. In generating the intervals, however, he used only the ten best random solutions from a sample of 10,000.

To get a better picture of the tail of the distribution, and thus a better estimate of the optimum, a sampling method that incorporates an effective heuristic can be used. Authors have generated samples of independent minimum values by applying a heuristic, usually a local improvement procedure, to a collection of randomly generated initial solutions (Golden and Alt, 1979; Los and Lardinois, 1982; Vasko and Wilson, 1984; Bruijs, 1984; Derigs, 1985; Gonsalvez et al., 1987). McRoberts (1971) suggested that the result of applying a local improvement procedure could be interpreted as a sample minimum and that if a local improvement procedure were repeated n times to n randomly generated initial solutions, a

collection of n independent extreme values would be obtained.

Potential theoretical problems that arise from the use of heuristic solutions as the sample of smallest values to which to fit the asymptotic distribution are outlined by Los and Lardinois (1982) and further discussed in Derigs (1985). Despite the facts that local optima are not extreme values of equal-sized samples of statistically independent observations and cannot be considered independent a priori, after they are generated their independence can be tested and their use in this context has been empirically validated by the authors listed above.

An additional statement made by Los and Lardinois (1982) in regard to the use of local minima is that there is an error in the application of the theory if there are repeated values in the sample. They state that having repeated values in the sample is equivalent to repeating the same implicit sample (that is, repeating an initial solution from which the local improvement procedure starts), which would violate an independence assumption. Their confidence-interval estimate is generated using only unique minima to avoid this issue. Ovacik et al. (2000) similarly obtain unique minima which are used to generate confidence intervals. This, however, is not necessary. Because for these discrete optimization problems the objective-function values to which distributions are fit do not necessarily have a one-to-one relationship with the solutions themselves, eliminating repeated values from the sample effectively changes the actual probability distribution of objective-function values. If several distinct solutions have the same objective-function value, the fact that the same objective-function value is repeated in the sample does not imply that the same implicit sample was repeated; rather, it accurately represents the distribution of all possible solutions as they translate to their corresponding objective-function values.

In the majority of previous research, the n sample minima used to fit the parameters of the Weibull distribution are generated from the application of a single local improvement procedure to n randomly generated initial solutions. There have been, however, some variants in how the sample of n minima are obtained. In order to remove the heuristic relation among the n values generated by applying the same local improvement procedure to all n randomly generated initial solutions, Los and Lardinois (1982) instead apply one of several different local improvement procedures to each initial solution. They also suggest generating nm local optima and using the n smallest values from each sample of size m as the sample. Similarly, Gonsalvez et al. (1987) in their application to the multicovering problem randomly selected one of several alternative heuristics at each step of their solution procedure so that the local minima obtained were not dependent on any particular heuristic. Finally, Ovacik et al. (2000) generated a collection of local minima by taking the values from which nonimproving moves were made in a simulated annealing run. As suggested by Los and Lardinois (1982), they selected a batch size m and used the n minima of the samples of size m as their sample.

Because the method of sampling has an effect on the resulting confidence interval, several methods are compared in this research. The confidence intervals generated by the random samples (RAN) as discussed in Section 5 are compared with confidence intervals generated with two different sampling schemes using local minima. These three sampling methods differ in two distinct ways. They differ either by whether or not a heuristic is incorporated in the sampling method and by how the actual sampling itself is accomplished. Section 6.1 explains the two sample generation methods that make use of the local improvement procedure, and Section 6.2 contains the results of the comparisons among the three sampling methods.

6.1. Sample generation methods using the local improvement procedure

Several different local improvement procedures have been implemented for flow line problems (Osman and Potts, 1989; Taillard, 1990; Ho and Chang, 1991; Zegordi et al., 1995; Gupta et al., 1997). The local improvement procedure developed in this research tries to improve makespan by iteratively moving jobs earlier in the schedule at the first stage. The details of the procedure are provided in Appendix A.

The local improvement procedure was used in two different ways to generate samples with which to calculate confidence intervals for the global minimum makespan for each of the test problems. First, parallel to how the sample was generated in the random case, the local improvement procedure was run on a total of nm random initial solutions where $n = m = 500$. The n minima of the samples of size m formed the sample of extreme values. This sampling method is referred to as LLI.

The second way in which the local improvement procedure was used was as a part of a bounded local improvement heuristic, denoted BLI. In this heuristic, a decision is made as to whether a random solution is discarded or further explored as it is developed based on lower bounds for its ultimate makespan that are calculated given partial schedule information. Complete details of this heuristic and the lower bounds given partial information are provided in Appendix B.

Data sets were generated with BLI in the following manner. A total of 50,000 iterations of the algorithm were run. This number of iterations was selected because it was large enough to generate sufficiently large samples. At each iteration where a complete solution was accepted and subsequently improved by the local improvement procedure, the makespan value of the improved solution was saved in the sample of minima. It was expected that this sample would be independent because for this problem the lower bound and the final value for a solution are not correlated. A solution with a low lower bound can turn out to have a high makespan value.

6.2. Comparison of the confidence intervals generated by each sampling method

For both RAN and LLI, 500 data points were used to generate the confidence intervals. If the sample did not pass the independence test, then no statistical confidence interval was calculated. RAN produced 169 independent samples, and LLI produced 166 independent samples. The least-squares parameter estimators were consistently better than the analytical parameter estimators for LLI as they were for RAN. The least-squares procedure failed to yield an acceptable fit on only one out of the 166 problems for which an interval was generated. The analytical estimators, however, failed the KS test 14 times and the AD test an additional 33 times for a total of 47 rejected fits.

The BLI samples were used to calculate confidence intervals in a manner more consistent with how local minima have been used in previous research. The total sample sizes (S) produced by this method ranged from 21 to 1701 with an average of 243. Given a batch size m, the sample size of minima, n, was set equal to $\min(\lfloor S/m \rfloor, 500)$. Initially a fit was attempted with $m = 1$. If that did not pass all statistical tests (independence, KS, and AD) at the 95% confidence level, then the batch size, m, was increased by one until a good fit was obtained, as long as the resulting sample size n remained greater than or equal to 25. In the cases where S was less than 25, a fit with $m = 1$ was still attempted. Although the confidence level associated with the Golden and Alt (1979) confidence interval is dependent on the sample size, for the smallest sample the confidence level is $(100 - 7.58 \times 10^{-10})\%$, so the confidence level for all intervals regardless of sample size can be assumed to be very close to 100%. BLI produced 179 independent samples and the least-squares estimators produced a good Weibull fit for each of them. Of the 179 confidence intervals calculated, 124 used $m = 1$ and 45 used $m = 2$. The maximum value used for m was 7. The average sample size n over all problems was 165.

The confidence intervals generated using each of the three sampling methods always contained the best known solution. In one case with BLI, the lower limit of the confidence interval was equal to the best known solution. Also, for each problem, the confidence intervals were always overlapping.

Table 2 shows the number of times that the deterministic lower bound (LB_{det}) and each sampling method produced the highest lower bound value (including ties) for each problem. As shown in Table 2, for the larger problems, BLI produced a lower bound that was greater than the other sampling methods. In this context, however, this does not imply that BLI yielded better lower bounds, and in fact, it could mean that the lower bounds calculated with BLI may not fall below the true optimal solution.

Table 2
Number of times each method produced the highest lower bound value

Problem	LB_{det}	RAN	LLI	BLI
$N_1 = 10, N_2 = 10$	19	6	13	8
$N_1 = 10, N_2 = 15$	9	6	7	23
$N_1 = 30, N_2 = 15$	1	14	11	19
$N_1 = 15, N_2 = 30$	0	12	5	28

Table 3
Number of times each method produced the lowest lower bound value[a]

Problem	RAN	LLI	BLI
$N_1 = 10, N_2 = 10$	14 (8)	6 (4)	25 (17)
$N_1 = 10, N_2 = 15$	23 (11)	11 (2)	11 (5)
$N_1 = 30, N_2 = 15$	13 (8)	23 (10)	9 (4)
$N_1 = 15, N_2 = 30$	17 (0)	22 (3)	6 (1)

[a] Entry in parentheses is the number of times the given statistical lower bound was less than the deterministic lower bound out of the number of times it had the lowest lower bound value.

Though the idea in generating the various lower bounds was to obtain bounds that were as large as possible, simply examining the bounds on the basis of how often they produced the greatest value does not provide a complete picture of their behavior. Table 3 gives the number of times that each sampling method produced the lowest lower bound value of the three sampling methods. The number in parentheses in each entry of the table represents the number of times the given statistical lower bound was less than the deterministic lower bound out of the number of times it had the lowest lower bound value. Table 3 highlights the strength of the statistical estimation methods to provide a better estimate of the optimal solution than the deterministic bound for this problem. LLI provides the lowest value most frequently while still being larger than the deterministic lower bound as problem size increases. That this would be the case is intuitive, as this sample should best represent the tail area of the distribution of smallest values.

Table 4
Average percentage deviations of the best known solution from the lower bound for each sampling method

p_1	p_2	$N_1 = 10, N_2 = 10$				$N_1 = 10, N_2 = 15$			
		Det	RAN	LLI	BLI	Det	RAN	LLI	BLI
[1, 10]	[1, 10]	15.9	9.9	9.8	13.5	26.4	12.3	14.9	5.8
	[5, 15]	6.9	5.0	4.9	3.3	12.7	7.7	6.1	5.1
	[10, 20]	1.9	1.9	1.4	1.9	3.9	3.9	3.8	3.0
[5, 15]	[1, 10]	4.4	3.2	4.4	4.4	7.6	5.3	4.8	3.6
	[5, 15]	21.2	10.8	8.7	11.0	33.6	8.2	8.1	6.1
	[10, 20]	12.5	7.5	5.9	6.3	20.8	9.7	8.4	7.2
[10, 20]	[1, 10]	0.9	0.9	0.9	0.9	1.1	1.1	1.1	1.1
	[5, 15]	11.5	11.5	10.1	5.9	16.9	9.7	5.1	3.6
	[10, 20]	24.8	10.7	8.4	9.7	35.5	15.3	16.5	8.5
		$N_1 = 30, N_2 = 15$				$N_1 = 15, N_2 = 30$			
[1, 10]	[1, 10]	30.4	18.5	16.8	6.4	43.8	5.8	17.8	2.7
	[5, 15]	16.5	4.7	8.9	4.7	21.9	6.1	5.1	3.1
	[10, 20]	5.6	3.1	3.6	4.0	10.9	3.9	3.0	2.3
[5, 15]	[1, 10]	18.1	6.9	17.4	2.5	19.4	10.4	6.9	1.8
	[5, 15]	35.4	13.5	8.4	5.0	46.3	4.9	6.7	10.6
	[10, 20]	24.2	4.3	13.5	8.7	29.5	5.7	11.3	2.2
[10, 20]	[1, 10]	10.1	8.6	5.9	3.4	7.8	1.7	3.0	1.3
	[5, 15]	27.4	8.3	17.0	10.8	28.7	6.8	4.8	3.5
	[10, 20]	38.2	5.9	8.0	4.4	46.0	6.5	7.5	2.8

Tables 2 and 3 give some idea about the relative strength of the lower bounds generated by each sampling method. However, it is also important to examine what each of these bounds imply about heuristic performance. If it is assumed that these confidence-interval lower limits of the form $\hat{z}^* = \hat{x}^* = x_{\min} - \hat{b}$ are valid lower bounds, what would be concluded about how close the heuristic solution values are to optimality? Table 4 gives the average percentage deviation of the lower bounds from the best known solution for each sampling method. The percentage deviation is given by

$$\frac{\mathrm{BK} - \max(\mathrm{LB}_{\mathrm{stat}}, \mathrm{LB}_{\mathrm{det}})}{\max(\mathrm{LB}_{\mathrm{stat}}, \mathrm{LB}_{\mathrm{det}})} \times 100\%, \tag{9}$$

where BK is the best known solution, and $\mathrm{LB}_{\mathrm{stat}}$ is the statistically estimated lower bound. In Table 4, p_1 indicates first-stage processing times and p_2 indicates second-stage processing times.

BLI frequently gave the best picture of heuristic performance. Due to the smaller sample sizes, these intervals would be expected to be less wide than the others. There is, however, a fundamental difference in the samples produced by each of the three methods. Fig. 2 shows the empirical c.d.f.'s for each of the three sampling methods for a representative problem. This figure clearly illustrates the effect of including a heuristic in the sampling method as both the BLI and LLI distributions encompass smaller values with greater probability than RAN. This figure also highlights the effect of the sampling scheme itself. Both RAN and LLI use the same sampling scheme, and their c.d.f.'s have extended tails whereas the BLI c.d.f. has a very steep tail. Due to this tail shape, it would not be expected that the lower limit of the confidence interval for a would be much less than $\hat{a}$. Because LLI appears to give the best picture of the smallest values of the population, it is most likely the more reliable sampling method.

7. Invalidation of some confidence intervals with a genetic algorithm

In order to further test the validity of the confidence intervals generated using the various sampling methods outlined in Sections 5 and 6, a genetic algorithm was implemented. Details of the genetic algorithm and its design may be found in Wilson (2001). The genetic algorithm provided better solutions for the problems examined than were previously known. These new solutions invalidated some of the confidence intervals for the optimal solutions derived in Section 6. Table 5 shows the number of confidence intervals invalidated by the genetic algorithm for each problem size and sampling method. Since the genetic algorithm is not guaranteed to find the optimal solu-

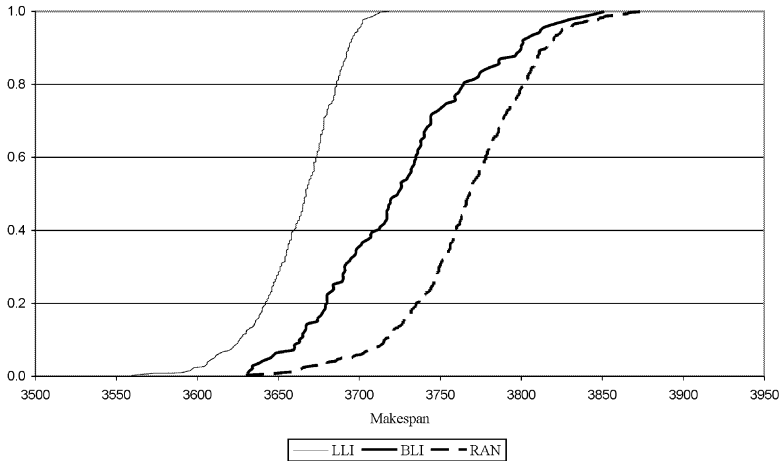

Fig. 2. Empirical c.d.f. comparison among the three sampling methods for a representative problem with $N_1 = 15$, $N_2 = 30$, and $p_1 = p_2 = [1, 10]$.

Table 5
Number of invalidated confidence intervals

Sampling method	$N_1 = 10$, $N_2 = 10$	$N_1 = 10$, $N_2 = 15$	$N_1 = 30$, $N_2 = 15$	$N_1 = 15$, $N_2 = 30$
RAN	0	1	12	12
LLI	0	0	6	4
BLI	0	2	14	27

tion, the numbers shown in Table 5 should be regarded as an optimistic picture of the true performance of the estimators. Recall that a total of 45 confidence intervals were generated by each sampling method for each problem size.

In Section 6 it was concluded that LLI was the most reliable sampling method. The results shown in Table 5 validate this conclusion, as of the three sampling methods it had the fewest confidence intervals invalidated by the genetic algorithm. BLI performed the worst, having a majority of its confidence intervals invalidated on the larger problems. These results further demonstrate the importance of the method of sampling in generating confidence intervals for optimal solutions and also show that statistical estimates of optimal solutions are not necessarily going to be consistently reliable as lower bounds even when generated from samples satisfying the requirements of the Fisher–Tippett result.

8. Conclusions and recommendations

Though this research has provided additional empirical evidence that using the extreme value theory approach in generating statistical estimates for the optimal objective-function value in large combinatorial minimization problems can be effective, this work has also revealed some caveats about the general use of such an approach. It was shown that the quality of a confidence-interval estimate of the minimum objective-function value depends critically on (i) the method for generating sample minimum objective-function values, and (ii) the method used to estimate the Weibull parameters from the set of sample minima. Recommendations for practitioners encountering similar estimation problems based on this case study are summarized below.

First, apply the method of least squares using the Nelder–Mead simplex search procedure to estimate the parameters of the Weibull distribution. In estimating the parameters of the Weibull distribution to approximate the empirical distribution of sample minimum objective-function values, the method of least squares using the Nelder–Mead simplex search procedure generally provided excellent results. The code for the Nelder–Mead procedure is readily available (Olsson, 1974), easy to implement, and takes little CPU time to execute (0.05 second for a sample of 500 on a 700 mHz computer).

Second, use the AD test coupled with the KS test to evaluate the goodness of fit for the Weibull approximation to the empirical distribution of the set of sample minimum objective-function values. In testing the goodness of fit for the Weibull approximation, the KS test was ineffective at detecting poor fits in the regions of interest of the empirical distribution. The AD test should also be employed to ensure that tail regions of the empirical distribution are adequately approximated.

Finally, employ a variety of sampling methods and validate the resulting confidence intervals with several heuristics. The method used to generate the set of n sample minimum objective-function values each based on a sample of size m is a crucial aspect of the application of these statistical estimation methods. Even though a particular minimization application may satisfy the requirements of the Fisher-Tippett theorem, yield sample minimum objective-function values that pass an independence test, and yield a Weibull distribution that closely approximates the corresponding empirical distribution, there is no guarantee for any finite values of m and n that the resulting Weibull distribution will accurately approximate the underlying *theoretical* distribution whose lower limit is the global minimum objective-function value z^*.

For the scheduling problem considered in this research, it was the case that a large sample of batched local minima obtained from the application of a local improvement procedure provided the most consistent results of all the sampling methods examined. It still cannot be concluded,

however, that similar sampling procedures will provide good results for other minimization problems with different characteristics. In applying these confidence-interval estimation methods to other problems for which optimal objective-function values are not known, it is imperative to use a variety of sampling methods and to validate the intervals with several heuristics.

It still remains unclear how good these statistical estimators can be expected to be in general and whether or not the confidence-interval lower limit may be regarded as a lower bound at all. The results of the genetic algorithm demonstrated the difficulty in using statistical estimators as lower bounds. Further, without knowing the optimal solutions there cannot be true validation of any interval estimates for any problem for which they are calculated. Despite these issues, statistical interval estimates of optimal solutions may provide practitioners more useful guidance on the expected performance of a given heuristic in a particular minimization problem than could be obtained from an easily calculated lower bound which is so loose as to be useless.

As an example, a gold standard for evaluating the experimental performance of computationally efficient heuristics for a given problem may be based on using a "most accurate" but possibly computationally inefficient heuristic as the local improvement procedure in order to obtain a more reliable statistical estimator of the form $\hat{z}^* = \hat{x}^* = x_{\min} - \hat{b}$ based on extreme value theory, least-squares estimation of Weibull parameters, and the Golden–Alt confidence interval. Since there is a substantial body of empirical evidence (including that for the problem at hand) that a genetic algorithm which has been tuned to the structure of a particular optimization problem generally delivers superior estimates of the optimal objective-function value, it would satisfy the criteria of this "most accurate" heuristic. Although computationally expensive, the resulting estimator of z^* could provide the basis for a comprehensive experimental evaluation of the average deviation from optimality for a given heuristic in a carefully selected suite of test problems.

It remains a topic of further research to determine how to best use the information provided by statistical interval estimates of optimal solutions in heuristic evaluation when optimal solutions are unknown.

Acknowledgements

The authors wish to thank three anonymous referees for their valuable comments and suggestions.

Appendix A. Local improvement procedure

1. Randomly generate a complete solution.
2. Determine the completion times of the machines at the second stage, t_i, $i = 1, \ldots, m_2$. Find the value of the solution, $C_{\text{sol}} = \max_{i=1,\ldots,m_2}(t_i)$.
3. Sort the completion times of the machines at the second stage in descending order. Let $[i]$ denote the machine in the ith position of the sorted order. Let $B_{[i]}$ be the total processing time of the last group scheduled on machine $[i]$ and let $R_{[i]}$ be the ready time of the last group scheduled on machine $[i]$.
4. Find the first machine $[i]$ for which $t_{[1]} - t_{[i]} > B_{[1]}$ and $R_{[1]} \leqslant t_{[i]}$. Denote this machine as M. Allocate the last group scheduled on machine [1] to the last position on machine M. Return to step 2. If there is no such machine, continue to step 5.
5. Sort the jobs in each second stage group i in ascending order of their ready times to the second stage, r_{ij}. Let $[j]$ denote the job in the jth position of the sorted order for group i and let p_{ij} be the second stage processing time of job j in second stage group i. For each group find the last job such that $r_{i[j]} - p_{i[j-1]} > r_{i[j-1]}$. Denote this job by J_i for each group i. The job to move is the job that completes the latest at the first stage of $J_i, i = 1, \ldots, N_2$.
6. If the selected job is not first in the job sequence within its first stage group, exchange it with the immediately preceding job. Calculate the makespan of the new schedule, C_{new}. If $C_{\text{new}} < C_{\text{sol}}$, set $C_{\text{sol}} = C_{\text{new}}$ and go to step 5. Else, go to step 7.
7. If the first stage group to which the selected job belongs is not the first in the group sequence on

its machine, exchange the group with its immediate predecessor. Calculate the makespan of the new schedule, C_{new}. If $C_{new} < C_{sol}$, set $C_{sol} = C_{new}$ and go to step 5. Else, go to step 8.

8. Move the selected job's first stage group to the machine at the first stage that completes earliest other than the one to which it is currently assigned. Insert the group on the new machine in the latest position possible where it will start earlier. Calculate the makespan of the new schedule, C_{new}. If $C_{new} < C_{sol}$, set $C_{sol} = C_{new}$ and go to step 5. Else, C_{sol} is accepted as the local optimal and the algorithm terminates.

Appendix B. Bounded local improvement procedure

The bounded local improvement procedure is identical to the local improvement procedure of Appendix A with the exception of step 1. Step 1 is divided into steps a, b, and c as outlined below. Expressions for each of the lower bounds are provided following the algorithm. A single iteration of the local improvement procedure establishes the initial value of C_{best}. Each time the local improvement procedure is executed, if $C_{sol} < C_{best}$, C_{best} is set equal to C_{sol}.

a. Randomly generate an assignment of groups to machines at the first stage. Calculate LB_{alloc}. If $LB_{alloc} \geqslant C_{best}$, repeat this step. Otherwise, continue to step b.
b. Randomly generate a sequence of groups on each machine at the first stage. Calculate LB_{grp}. If $LB_{grp} \geqslant C_{best}$, go to step a. Otherwise, continue to step c.
c. Randomly generate a sequence of jobs for each group at the first stage. Calculate LB_{first}. If $LB_{first} \geqslant C_{best}$, go to step a. Otherwise, continue to step 2 of the local improvement procedure.

The lower bounds are calculated as follows:

$$LB_{alloc} = \max_{i=1,\ldots,m_1} \left(\sum_{k \in G_i} A_k + \min_{j \in J_i} (pt_j) \right),$$

where G_i = set of first stage groups allocated to machine i at the first stage, J_i = set of jobs allocated to machine i at the first stage, A_k = total first stage processing time of first stage group k, pt_j = second stage processing time of job j.

$LB_{grp} = \max(LB_1, LB_2)$ where LB_1 and LB_2 are calculated as shown below. Release times to the second stage are calculated for each job assuming that it is processed first within its first stage group. Earliest start times are then calculated for each second stage group i, ES_i, such that there would be no idle time once the processing of a second stage group began.

$$LB_1 = ES_{m_2} + \left(\sum_{i=1}^{N_2} B_i - \sum_{i=1}^{m_2-1} x_i \right) \Big/ m_2,$$

$$LB_2 = \max_{h=1,\ldots,N_2} \left(ES_h + \left(\sum_{i=h}^{N_2} B_i \right) \Big/ y \right),$$

$$y = \min(N_2 + 1 - h, m_2),$$

where B_i = total second stage processing time of second stage group i, ES_{m_2} = the m_2th largest earliest start time,

$$x_i = \begin{cases} ES_{m_2} - ES_i & \text{if } ES_i + B_i > ES_{m_2}, \\ B_i & \text{otherwise,} \end{cases}$$

and the second stage groups are sorted such that $ES_1 \leqslant ES_2 \leqslant \cdots \leqslant ES_{N_2}$.

LB_{first}, the lower bound given all solution information except the second stage allocation, is calculated in the same manner as LB_{grp}. The exception is that because the job sequence within each first stage group is known, the release times of the jobs and therefore the earliest start times of the second stage groups do not need to be estimated.

References

Boender, C.G.E., Rinnooy Kan, A.H.G., Stougie, L., Timmer, G.T., 1982. A stochastic method for global optimization. Mathematical Programming 22, 125–140.

Bruijs, P.A., 1984. On the quality of heuristic solutions to a 19×19 quadratic assignment problem. European Journal of Operational Research 17, 21–30.

Cooke, P., 1979. Statistical inference for bounds of random variables. Biometrika 66, 367–374.

Dannenbring, D.G., 1977. Procedures for estimating optimal solution values for large combinatorial problems. Management Science 23 (12), 1273–1283.

Derigs, U., 1985. Using confidence limits for the global optimum in combinatorial optimization. Operations Research 33 (5), 1024–1049.

Fisher, R.A., Tippett, L.H.C., 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proceedings of the Cambridge Philosophical Society 24, 180–190.

Golden, B.L., Alt, F.B., 1979. Interval estimation of a global optimum for large combinatorial problems. Naval Research Logistics Quarterly 26 (1), 69–77.

Gonsalvez, D.J., Hall, N.G., Rhee, W.T., Siferd, S.P., 1987. Heuristic solutions and confidence intervals for the multi-covering problem. European Journal of Operational Research 31, 94–101.

Gumbel, E.J., 1958. Statistics of Extremes. Columbia University Press, New York.

Gupta, J.N.D., 1988. Two-stage, hybrid flowshop scheduling problem. Journal of the Operational Research Society 39 (4), 359–364.

Gupta, J.N.D., Hariri, A.M.A., Potts, C.N., 1997. Scheduling a two-stage hybrid flow shop with parallel machines at the first stage. Annals of Operations Research 69, 171–191.

Harter, H.L., Moore, A.H., 1965. Maximum-likelihood estimation of the parameters of gamma and weibull populations from complete and censored samples. Technometrics 7 (4), 639–643.

Harter, H.L., Moore, A.H., 1967. Errata. Technometrics 9 (1), 195.

Ho, J.C., Chang, Y., 1991. A new heuristic for the n-job, M-machine, flow-shop problem. European Journal of Operational Research 52, 194–202.

Humphrey, D.G., Wilson, J.R., 2000. A revised simplex search procedure for stochastic simulation response-surface optimization. INFORMS Journal on Computing 12 (4), 272–283.

Law, A.M., Kelton, W.D., 1991. Simulation Modeling and Analysis, second ed. McGraw Hill, Inc., New York.

Leadbetter, M.R., Lindgren, G., Rootzén, H., 1983. Extremes and Related Properties of Random Sequences and Processes. Springer-Verlag, New York.

Los, M., Lardinois, C., 1982. Combinatorial programming, statistical optimization and the optimal transportation network problem. Transportation Research B 16 (2), 89–124.

McRoberts, K.L., 1971. A search model for evaluating combinatorially explosive problems. Operations Research 19, 1331–1349.

Olsson, D.M., 1974. A sequential simplex program for solving minimization problems. Journal of Quality Technology 6 (1), 53–57.

Olsson, D.M., Nelson, L.S., 1975. The Nelder–Mead simplex procedure for function minimization. Technometrics 17 (1), 45–51.

Osman, I., Potts, C.N., 1989. Simulated annealing for permutation flow-shop scheduling. Omega 17 (6), 551–557.

Ovacik, I.M., Rajagopalan, S., Uzsoy, R., 2000. Integrating interval estimates of global optima and local search methods for combinatorial optimization problems. Journal of Heuristics 6, 481–500.

Robson, D.S., Whitlock, J.H., 1964. Estimation of a truncation point. Biometrika 51, 33–39.

Swain, J.J., Venkatraman, S., Wilson, J.R., 1988. Least-squares estimation of distribution functions in Johnson's translation system. Journal of Statistical Computation and Simulation 29, 271–297.

Taillard, E., 1990. Some efficient heuristic methods for the flow shop sequencing problem. European Journal of Operational Research 47, 65–74.

Van der Watt, P., 1980. A note on estimation of bounds of random variables. Biometrika 67, 712–714.

Vasko, F.J., Wilson, G.R., 1984. An efficient heuristic for large set covering problems. Naval Research Logistics Quarterly 31, 163–171.

Wagner, M.A.F., Wilson, J.R., 1996. Using univariate Bézier distributions to model simulation input processes. IIE Transactions 28 (9), 699–712.

Wilson, A.D., 2001. Scheduling nonsimilar groups of jobs on a flow line. Ph.D. dissertation, North Carolina State University.

Zanakis, S.H., 1979. A simulation study of some simple estimators for the three-parameter Weibull distribution. Journal of Statistical Computation and Simulation 9, 101–116.

Zegordi, S.H., Itoh, K., Enkawa, T., 1995. Minimizing makespan for flow shop scheduling by combining simulated annealing with sequencing knowledge. European Journal of Operational Research 85, 515–531.