Journal of Global Optimization (2005) 31:317–332

Single-Machine Scheduling with Exponential Processing Times and General Stochastic Cost Functions*

XIAOQIANG CAI[1] and XIAN ZHOU[2]
[1]*Department of Systems Engineering & Engineering Management, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong*
[2]*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*

(Received 30 March 2004; accepted in revised form 1 April 2004)

Abstract. We study a single-machine stochastic scheduling problem with n jobs, in which each job has a random processing time and a general stochastic cost function which may include a random due date and weight. The processing times are exponentially distributed, whereas the stochastic cost functions and the due dates may follow any distributions. The objective is to minimize the expected sum of the cost functions. We prove that a sequence in an order based on the product of the rate of processing time with the expected cost function is optimal, and under certain conditions, a sequence with the weighted shortest expected processing time first (WSEPT) structure is optimal. We show that this generalizes previous known results to more general situations. Examples of applications to practical problems are also discussed.

Key words: Due dates, Exponential processing times, Single machine, Stochastic cost functions, Stochastic scheduling

1. Introduction

We address the following problem. A number of n jobs are to be processed on a single-machine, which are all available at time zero. The processing times p_i of job i, $i = 1, 2, \dots, n$, are independent random variables. The cost functions are stochastic processes which can include various deterministic and/or stochastic attributes such as due dates d_i, weights w_i, etc. Let $\lambda = (i_1, \dots, i_n)$ be a permutation of the integers $\{1, 2, \dots, n\}$, referred to as a *sequence*, that determines the order to process the jobs, with $i_k = i$ if and only if job i is the kth to be processed. The problem is to find an optimal

* This work was partially supported by the Research Grants Council of Hong Kong under Earmarked Grants No. CUHK4418/99E and No. PolyU 5081/00E.

sequence λ^* that minimizes the following objective function:

$$\mathrm{ETC}(\lambda) = \mathrm{E}\left[\sum_{i=1}^{n} f_i(C_i(\lambda))\right] \tag{1.1}$$

over all sequences λ, where for each $i \in \{1, 2, \ldots, n\}$,

$\{f_i(t) : t \geqslant 0\}$ is a stochastic process independent of $p_1, \ldots, p_n$,
$C_i(\lambda)$ is the completion time of job i under a sequence λ, and
$\mathrm{E}(X)$ represents the expectation of a random variable X.

In particular, we consider the objective function of the form:

$$\begin{aligned} \mathrm{ETC}(\lambda) &= \mathrm{E}\left[\sum_{i:C_i(\lambda)>d_i} w_i g(C_i(\lambda) - d_i)\right] \\ &= \mathrm{E}\left[\sum_{i=1}^{n} w_i g(C_i(\lambda) - d_i) I_{\{C_i(\lambda)>d_i\}}\right], \end{aligned} \tag{1.2}$$

where

$g(\cdot)$ is a non-negative and non-decreasing (deterministic) function defined on $[0, \infty)$,
w_i, $i = 1, 2, \ldots, n$, are deterministic weights,
d_i, $i = 1, 2, \ldots, n$, are random due dates independent of the processing times $\{p_i\}$, and
I_A is the indicator of an event A which takes value 1 if A occurs and 0 otherwise.

Both (1.1) and (1.2) are general objective functions, although (1.2) is a special case of (1.1). Many stochastic scheduling problems studied previously in the literature have performance measures covered by (1.2). Some typical examples are given below:

(a) Expected weighted sum of completion times (let $f_i(t) = w_i t$, see Rothkopf, 1966).
(b) Expected weighted sum of squared completion times (let $f_i(t) = w_i t^2$, see Bagga and Kalra, 1981).
(c) Expected total weighted tardiness (let $f_i(t) = w_i t I_{\{t>d_i\}}$, see Pinedo, 1983).
(d) Expected weighted number of tardy jobs (let $f_i(t) = w_i I_{\{t>d_i\}}$, see Pinedo, 1983; Boxma and Forst, 1986).
(e) Weighted lateness probability, that is, $\sum_{i=1}^{n} w_i \Pr(C_i(\lambda) > d_i)$ (which is equivalent to the objective function in (d) above; see Sarin et al., 1991).

In this paper, we study the problem where processing times $p_1, \dots, p_n$ follow exponential distributions, i.e., p_i has a density function of the form $\lambda_i e^{-\lambda_i x}$, or equivalently, a cumulative distribution function (cdf) $1 - e^{-\lambda_i x}$, $i = 1, 2, \dots, n$. The exponential distribution is often used to model uncertain times, and is justified in the case with a high level of uncertainty (see Cai and Zhou, 2000). The parameter λ_i represents the 'rate' of p_i and is reciprocal to the mean of p_i.

The cost functions $f_j(\cdot)$, $j = 1, 2, \dots, n$, in (1.1) are general stochastic processes, and the due dates d_j in (1.2) may follow arbitrary distributions with cdf F_j, $j = 1, 2, \dots, n$. For brevity, denote

- the model in (1.1) as $1 \mid p_j \sim \exp(\lambda_j) \mid E\left[\sum f_j(C_j)\right]$, and
- the model in (1.2) as $1 \mid p_j \sim \exp(\lambda_j), d_j \sim F_j \mid E\left[\sum w_j g(C_j - d_j) I_{\{C_j > d_j\}}\right]$.

We only consider the *non-preemptive* scheduling problems in this paper. In other words, the machine will process each job continuously until it is completed, with job preemption in the middle of processing not allowed.

Stochastic scheduling problems with random processing times and/or random due dates have been topics of research for decades. See, for example, Pinedo (2002) and Righter (1994). Among these problems, models with exponential processing times are commonly addressed in the literature and, consequently, some very elegant results on their solutions have been derived. Some relevant works on problems with exponential processing times are briefly reviewed below.

Derman et al. (1978) considered the problem of minimizing the weighted number of tardy jobs on a single machine. They showed that the weighted shortest expected processing time (WSEPT) sequence is optimal when all jobs have a common random due date which follows an arbitrary distribution.

Glazebrook (1979) examined a parallel-machine problem. He showed that the shortest expected processing time (SEPT) sequence minimizes the expected mean flowtime.

Weiss and Pinedo (1980) investigated multiple non-identical machine problems. They addressed a performance measure that involves a cost rate $g(U_t)$ at time $t \geqslant 0$, where U_t is the set of uncompleted jobs at the time t. They showed that several cost functions, including expected sum of weighted completion times, expected makespan, and expected lifetime of a series system, are covered by the performance measure and are minimized by a SEPT sequence or a longest expected processing time (LEPT) sequence.

Pinedo (1983) examined four problems, namely, minimization of the expected weighted sum of completion times of jobs with random arrival times on a single machine, minimization of the expected weighted sum of job tardinesses on a single machine, minimization of the expected weighted number of tardy jobs on a single machine, and minimization of the expected weighted number of tardy jobs on parallel machines. The first is a preemptive version of the problem, while the rest are non-preemptive. He showed that WSEPT sequences are optimal under certain (compatability) conditions.

A single-machine scheduling problem with the expected weighted number of tardy jobs as objective function was investigated by Boxma and Forst (1986). They derived optimal sequences in cases with various due date and processing time distributions, including exponential due dates, exponential processing times, independently and identically distributed (i.i.d.) due dates, i.i.d. processing times, etc.

Kampke (1989) generalized the work of Weiss and Pinedo (1980). Sufficient conditions for optimal priority policies, which may be more general than the SEPT or the LEPT, were derived.

In Pinedo (2002), the WSEPT sequence has been shown to minimize the performance measure $\mathrm{E}\big[\sum w_i h(C_i)\big]$, where $h(\cdot)$ is a general function. This covers objective functions in which all jobs have a common deterministic due date, such as the expected sum of weighted tardinesses. Moreover, the performance measure $\mathrm{E}\big[\sum w_i h_i(C_i)\big]$ is also studied, where $h_i(\cdot)$ is a job-dependent cost function. Pinedo defined an order $h_j \geqslant_s h_k$ (termed as h_j is *steeper* than h_k) between the cost functions by $dh_j(t) \geqslant dh_k(t)$ if the differentials exist; or $h_j(t+\delta) - h_j(t) \geqslant h_k(t+\delta) - h_k(t)$ otherwise, for all $t \geqslant 0$ and $\delta > 0$. It is shown that if $\lambda_j w_j \geqslant \lambda_i w_i \iff h_j \geqslant_s h_k$, then the WSEPT sequence minimizes $\mathrm{E}\big[\sum w_i h_i(C_i)\big]$.

In this paper, we address the problem $1 \mid p_j \sim \exp(\lambda_j) \mid \mathrm{E}\big[\sum f_j(C_j)\big]$ with general stochastic cost functions $f_i(t)$, which is non-decreasing in t. In particular, we consider the problem $1 \mid p_j \sim \exp(\lambda_j), d_j \sim F_j \mid \mathrm{E}\big[\sum w_j g(C_j - d_j) I_{\{C_j > d_j\}}\big]$, where the due dates are random variables and $g(\cdot)$ is a general non-decreasing function. This is a general objective function, which subsumes many performance measures commonly studied in the literature, as we have shown above. The main result for the problem $1 \mid p_j \sim \exp(\lambda_j) \mid \mathrm{E}\big[\sum f_j(C_j)\big]$ is that a sequence in the order based on the increments of $\lambda_j \mathrm{E}[f_j(t)]$ is optimal. This extends the results of Pinedo (2002) by *dropping* the compatibility condition (such as $\lambda_j w_j \geqslant \lambda_i w_i \iff h_j \geqslant_s h_k$), and by allowing *stochastic* cost functions. For $1 \mid p_j \sim \exp(\lambda_j), d_j \sim F_j \mid \mathrm{E}\big[\sum w_j g(C_j - d_j) I_{\{C_j > d_j\}}\big]$, we will present two results:

(1) When all due dates d_i have a common distribution, the WSEPT sequence is optimal without requiring any additional conditions.
(2) Otherwise, if $g(\cdot)$ is convex on $[0, \infty)$ with $g(0) = 0$, then a sequence in the non-increasing order of $\{\lambda_j w_j F_j(x)\}$ is optimal. In particular, if $\lambda_i w_i \geqslant \lambda_j w_j$ implies that d_i is stochastically less than or equal to d_j

(the definition of the stochastic order is given after Theorem 3), then the WSEPT sequence is optimal.

To highlight the generality and applicability of our results, we provide two practical examples in Section 3. Finally, Section 4 provides some concluding remarks.

2. Main results

We assume, throughout this paper, that each $f_i(t)$ has a finite mean function $m_i(t) = \mathrm{E}[f_i(t)]$ for all $t \geqslant 0$ which is non-decreasing in t.

THEOREM 1. *For the problem* $1 \mid p_j \sim \exp(\lambda_j) \mid \mathrm{E}\left[\sum f_j(C_j)\right]$, *if* $i > j$ *implies that* $\lambda_i m_i(t)$ *has increments no more than those of* $\lambda_j m_j(t)$ *at any* t, *i.e.,*

$$\lambda_i[m_i(t) - m_i(s)] \leqslant \lambda_j[m_j(t) - m_j(s)] \quad \forall t > s, \tag{2.1}$$

or equivalently,

$$\int_0^\infty \phi(s)\lambda_i dm_i(s) \leqslant \int_0^\infty \phi(s)\lambda_j dm_j(s) \tag{2.2}$$

for any non-negative measurable function $\phi(s)$ on $[0, \infty)$, *where the integrals are in Lebesgue–Stieltjes sense, then the sequence* $(1, 2, \ldots, n)$ *is optimal. In other words, a sequence in non-increasing order of the increments of* $\{\lambda_i m_i(t)\}$ *is optimal.*

Proof. First, by taking $\phi(s) = I_{[s,t]}$ in (2.2) we see that (2.2) implies (2.1). Conversely, for any non-negative measurable function $\phi(s)$, we can construct functions $\phi_1(s) \leqslant \phi_2(s) \leqslant \cdots$, with each $\phi_k(s)$ being a linear combination of functions of form $I_{[s,t]}$, such that $\phi_k(s) \to \phi(s)$ as $k \to \infty$.

Hence an application of the monotone convergence theorem shows that (2.1) implies (2.2). This establishes the equivalence between (2.1) and (2.2).

Next, since $\{f_i(t)\}$ are independent of $\{p_i\}$, we have, for $i, j = 1, 2, \ldots, n$ and $t \geqslant 0$,

$$\begin{aligned}\mathrm{E}[f_i(t+p_j)] = \mathrm{E}\{\mathrm{E}[f_i(t+p_j)|p_j]\} &= \int_0^\infty \mathrm{E}[f_i(t+x)|p_j = x]\lambda_j \mathrm{e}^{-\lambda_j x}\mathrm{d}x \\ &= \int_0^\infty \mathrm{E}[f_i(t+x)]\lambda_j \mathrm{e}^{-\lambda_j x}\mathrm{d}x = \int_0^\infty m_i(t+x)\lambda_j \mathrm{e}^{-\lambda_j x}\mathrm{d}x.\end{aligned} \tag{2.3}$$

Furthermore, by convolution it can be shown that

$$\text{the density of } p_i + p_j = \begin{cases} \frac{\lambda_i \lambda_j}{\lambda_j - \lambda_i} \left(\mathrm{e}^{-\lambda_i x} - \mathrm{e}^{-\lambda_j x} \right) & \text{if } \lambda_i \neq \lambda_j \\ \lambda_i^2 x \mathrm{e}^{-\lambda_i x} & \text{if } \lambda_i = \lambda_j. \end{cases} \tag{2.4}$$

(Note that the second part of (2.4) is equal to the limit of the first part as λ_j converges to λ_i.) Thus, when $\lambda_i \neq \lambda_j$, by (2.4) together with an argument similar to (2.3) we obtain

$$\mathrm{E}[f_i(t + p_i + p_j)] = \frac{\lambda_i \lambda_j}{\lambda_j - \lambda_i} \int_0^\infty m_i(t + x) \left(\mathrm{e}^{-\lambda_i x} - \mathrm{e}^{-\lambda_j x} \right) \mathrm{d}x. \tag{2.5}$$

Let $\lambda = \{..., i, j, ...\}$ be an arbitrary job sequence, $\lambda' = \{..., j, i, ...\}$ be the sequence by interchanging two consecutive jobs i, j in λ, and C denote the completion time of the job prior to job i under λ. Then, for the objective function ETC in (1.1),

$$\begin{aligned} \mathrm{ETC}(\lambda) - \mathrm{ETC}(\lambda') = &\, \mathrm{E}[f_i(C + p_i)] + \mathrm{E}[f_j(C + p_i + p_j)] \\ & - \mathrm{E}[f_j(C + p_j)] - \mathrm{E}[f_i(C + p_i + p_j)]. \end{aligned} \tag{2.6}$$

Since $p_1, \ldots, p_n$ are mutually independent, conditional on $C = t$ we have

$$\mathrm{E}[f_i(C + p_i) | C = t] = \mathrm{E}[f_i(t + p_i) | C = t] = \mathrm{E}[f_i(t + p_i)]$$

and similarly, $\mathrm{E}[f_i(C + p_i + p_j) | C = t] = \mathrm{E}[f_i(t + p_i + p_j)]$. Hence a combination of (2.6) with (2.3) and (2.5) yields that, conditional on $C = t$,

$$\begin{aligned} \mathrm{ETC}(\lambda) - \mathrm{ETC}(\lambda') = &\, \mathrm{E}[f_i(t + p_i)] + \mathrm{E}[f_j(t + p_i + p_j)] \\ & - \mathrm{E}[f_j(t + p_j)] - \mathrm{E}[f_i(t + p_i + p_j)] \\ = & \int_0^\infty m_i(t + x) \left\{ \lambda_i \mathrm{e}^{-\lambda_i x} - \frac{\lambda_i \lambda_j}{\lambda_j - \lambda_i} \left(\mathrm{e}^{-\lambda_i x} - \mathrm{e}^{-\lambda_j x} \right) \right\} \mathrm{d}x \\ & - \int_0^\infty m_j(t+x) \left\{ \lambda_j \mathrm{e}^{-\lambda_j x} - \frac{\lambda_i \lambda_j}{\lambda_j - \lambda_i} \left(\mathrm{e}^{-\lambda_i x} - \mathrm{e}^{-\lambda_j x} \right) \right\} \mathrm{d}x \\ = & \int_0^\infty [\lambda_i m_i(t+x) - \lambda_j m_j(t+x)] \frac{\lambda_j \mathrm{e}^{-\lambda_j x} - \lambda_i \mathrm{e}^{-\lambda_i x}}{\lambda_j - \lambda_i} \, \mathrm{d}x \\ = &\, a_{ij}(t), \quad \text{say}. \end{aligned} \tag{2.7}$$

Extend the domain of each $m_i(t)$ to $(-\infty, \infty)$ by defining $m_i(t) = 0$ for $t < 0$. Then $m_i(\cdot)$ is a non-decreasing function on $(-\infty, \infty)$. Hence we can write $m_i(t + x) = \int_{-\infty}^{t+x} \mathrm{d}m_i(s)$, $i = 1, \ldots, n$. An application of Fubini's Theorem then gives

$$\begin{aligned}
a_{ij}(t) &= \int_0^\infty \int_{-\infty}^{t+x} [\lambda_j \mathrm{d}m_j(s) - \lambda_i \mathrm{d}m_i(s)] \frac{\lambda_i \mathrm{e}^{-\lambda_i x} - \lambda_j \mathrm{e}^{-\lambda_j x}}{\lambda_j - \lambda_i}\,\mathrm{d}x \\
&= \int_{-\infty}^{t} \int_0^\infty \frac{\lambda_i \mathrm{e}^{-\lambda_i x} - \lambda_j \mathrm{e}^{-\lambda_j x}}{\lambda_j - \lambda_i}\,\mathrm{d}x \left[\lambda_j \mathrm{d}m_j(s) - \lambda_i \mathrm{d}m_i(s)\right] + \\
&\quad + \int_t^\infty \int_{s-t}^\infty \frac{\lambda_i \mathrm{e}^{-\lambda_i x} - \lambda_j \mathrm{e}^{-\lambda_j x}}{\lambda_j - \lambda_i}\,\mathrm{d}x \left[\lambda_j \mathrm{d}m_j(s) - \lambda_i \mathrm{d}m_i(s)\right] \\
&= \int_t^\infty \frac{\mathrm{e}^{-\lambda_i (s-t)} - \mathrm{e}^{-\lambda_j (s-t)}}{\lambda_j - \lambda_i} \left[\lambda_j \mathrm{d}m_j(s) - \lambda_i \mathrm{d}m_i(s)\right]. \qquad (2.8)
\end{aligned}$$

It is easy to see that

$$\frac{\mathrm{e}^{-\lambda_i (s-t)} - \mathrm{e}^{-\lambda_j (s-t)}}{\lambda_j - \lambda_i} \geqslant 0 \qquad \text{for all} \quad s \geqslant t.$$

Hence by (2.7)–(2.8) together with condition (2.2) we have, condition on $C = t$,

$$i > j \implies \mathrm{ETC}(\lambda) - \mathrm{ETC}(\lambda') = a_{ij}(t) \geqslant 0 \quad \forall t \geqslant 0,$$

which in turn implies, unconditionally, $\mathrm{ETC}(\lambda) - \mathrm{ETC}(\lambda') \geqslant 0$.

Thus, we have shown that $\mathrm{ETC}(\lambda) \geqslant \mathrm{ETC}(\lambda')$ for $i > j$ when $\lambda_i \neq \lambda_j$. The same holds when $\lambda_i = \lambda_j$ as well, which can be similarly proven using the second part of (2.4), or considering the limit as λ_j converges to λ_i. It follows that the sequence λ' is better than λ if $i > j$. In other words, if $i > j$ while job i is ahead of job j, then the objective function can be reduced by switching i and j. This means that any sequence other than $(1, 2, ..., n)$ can be improved. Consequently the sequence $(1, 2, ..., n)$ is optimal. □

Note that condition (2.2) is what we need to prove Theorem 1, while condition (2.1) is usually easier to check in specific cases. Also, (2.1) does not require $m_i(t)$ to be differentiable at all $t \geqslant 0$. For example, $m_i(t)$ may be discontinuous at some points. Then (2.1) assumes that

$$\lambda_i[m_i(t+) - m_i(t-)] \leqslant \lambda_j[m_j(t+) - m_j(t-)] \quad \text{for } i > j$$

at any discontinuity t (which can also be written as $\lambda_i \mathrm{d}m_i(t) \leqslant \lambda_j \mathrm{d}m_j(t)$ in that sense). It is also possible for $m_i(t)$ to have different left and right derivatives at some t. In such a case, (2.1) requires $\lambda_i \mathrm{d}m_i(t+) \leqslant \lambda_j \mathrm{d}m_j(t+)$ and $\lambda_i \mathrm{d}m_i(t-) \leqslant \lambda_j \mathrm{d}m_j(t-)$ for $i > j$.

Remark 1. Theorem 1 extends the results of Pinedo (2002). Condition (2.1) or (2.2) is in fact equivalent to '$\lambda_j m_j$ is steeper than $\lambda_i m_i$' in Pinedo's terminology. Hence Theorem 1 says that the sequence in a reverse steepness order of $\{\lambda_i m_i(t), i = 1, \ldots, n\}$ is optimal. Note that in Pinedo (2002), who considers deterministic cost functions f_i only, an agreeable condition is needed between the steepness of $f_i(t)/w_i$ and the order of $\lambda_i w_i$, i.e., $\lambda_i w_i \geqslant \lambda_j w_j$ implies that $f_i(t)/w_i$ is steeper than $f_j(t)/w_j$. In our Theorem 1, such an agreeable condition can be replaced by a weaker condition (2.1) (or equivalently, (2.2)). In addition, Theorem 1 is more general than the results of Pinedo (2002), in that it allows stochastic cost functions, so that the parameters such as due dates, weights, etc., can be random variables in our theorem instead of being restricted to the deterministic case.

The following example shows an application of Theorem 1.

EXAMPLE 1. Let $f_i(t) = w_i h(t)$, where w_i is a deterministic weight and $h(t)$ is a non-decreasing stochastic process. Then $m_i(t) = \mathrm{E}[f_i(t)] = w_i \mathrm{E}[h(t)]$ is non-decreasing in t. Furthermore, if $\lambda_i w_i > \lambda_j w_j$, then

$$\begin{aligned}\lambda_i[m_i(t) - m_i(s)] &= \lambda_i w_i \{\mathrm{E}[h(t)] - \mathrm{E}[h(s)]\} \\ &\geqslant \lambda_j w_j \{\mathrm{E}[h(t)] - \mathrm{E}[h(s)]\} = \lambda_j[m_j(t) - m_j(s)] \quad \forall t > s.\end{aligned}$$

Hence by Theorem 1, a sequence in non-increasing order of $\{\lambda_i w_i\}$ minimizes $\mathrm{E}[\sum w_i h(C_i)]$. As $\mathrm{E}[p_i] = 1/\lambda_i$, this sequence is the WSEPT and so the result generalizes that of Pinedo (2002) to an arbitrary stochastic instead of a deterministic cost function h. In particular, the $h(t)$ in here allows a random common due date with an arbitrary distribution.

There are, of course, also examples where the condition does not hold. A simple one is given below.

EXAMPLE 2. Let $f_1(t) = 2t$ and $f_2(t) = t^2$, which are deterministic cost functions. Then $m_1(t) = 2t$ and $m_2(t) = t^2$. Hence $\mathrm{d}m_1(t) = 2\mathrm{d}t$ and $\mathrm{d}m_2(t) = 2t\mathrm{d}t$. It follows that $\lambda_1 \mathrm{d}m_1(t) \leqslant \lambda_2 \mathrm{d}m_2(t)$ when $t \geqslant \lambda_1/\lambda_2$, and $\lambda_1 \mathrm{d}m_1(t) > \lambda_2 dm_2(t)$ for $t < \lambda_1/\lambda_2$. Thus (2.1) cannot hold for jobs 1 and 2. Furthermore, suppose $w_1 = w_2$. Then it is not difficult to show that $\mathrm{ETC}(1,2) < \mathrm{ETC}(2,1)$ if and only if $\lambda_1 > \lambda_2^2$. Hence the WSEPT rule is not optimal even if the jobs have a common weight.

The applications of Theorem 1 lead to the next two theorems for the problem $1 \mid p_j \sim \exp(\lambda_j), d_j \sim F_j \mid \mathrm{E}\big[\sum w_j g(C_j - d_j) I_{\{C_j > d_j\}}\big]$. We first give a result with identically distributed due dates.

THEOREM 2. *If d_i have a common distribution, then a sequence in non-increasing order of $\{\lambda_i w_i\}$, or equivalently, in non-decreasing order of $\{\mathrm{E}(p_i)/w_i\}$, minimizes the* $\mathrm{ETC}(\lambda)$ *in (1.2).*

Proof. Let $f_i(t) = w_i g(t - d_i) I_{\{t > d_i\}}$, $i = 1, 2, \ldots, n$, and $F(x)$ be the common distribution function of d_i. Then

$$m_i(t) = \mathrm{E}[f_i(t)] = w_i \mathrm{E}[g(t - d_i) I_{\{t > d_i\}}] = w_i \int_{0 \leqslant x < t} g(t - x)\ \mathrm{d}F(x). \quad (2.9)$$

Let

$$\tilde{g}(t) = \int_{0 \leqslant x < t} g(t - x) \mathrm{d}F(x)$$

so that $\lambda_i m_i(t) = \lambda_i w_i \tilde{g}(t)$. Since $g(t)$ is non-negative and non-decreasing, so is $\tilde{g}(t)$. It follows that $\lambda_i w_i \geqslant \lambda_j w_j$ implies

$$\begin{aligned} \lambda_i[m_i(b) - m_i(a)] &= \lambda_i w_i[\tilde{g}(b) - \tilde{g}(a)] \geqslant \lambda_j w_j[\tilde{g}(b) - \tilde{g}(a)] \\ &= \lambda_j[m_j(b) - m_j(a)] \end{aligned}$$

for all $a < b$. Thus if $\lambda_1 w_1 \geqslant \cdots \geqslant \lambda_n w_n$, then $\{1, ..., n\}$ is optimal by Theorem 1, in other words, a sequence in non-increasing order of $\{\lambda_i w_i\}$ minimizes ETC(λ). □

EXAMPLE 3.

(i) Let $g(x) \equiv 1$. Then Theorem 2 says that a sequence in non-increasing order of $\{\lambda_i w_i\}$ minimizes the expected weighted number of tardy jobs, or equivalently, the weighted lateness probability, when the due dates have a common distribution (not necessarily a common due date).

(ii) Let $g(x) = x$. Then by Theorem 2, a sequence in non-increasing order of $\{\lambda_i w_i\}$ minimizes the expected weighted sum of job tardinesses.

Note that the above results do not require any compatibility conditions between the weights and processing times.

The next theorem allows the due dates to have different distributions.

THEOREM 3. *If* $g(\cdot)$ *is convex with* $g(0) = 0$, *and if*

$$\lambda_1 w_1 F_1(x) \geqslant \lambda_2 w_2 F_2(x) \geqslant \cdots \geqslant \lambda_n w_n F_n(x) \quad \text{for } x \geqslant 0, \quad (2.10)$$

then the sequence $(1, 2, ..., n)$ *minimizes the* ETC(λ) *in (1.2). In other words, a sequence in the non-increasing order of* $\{\lambda_j w_j F_j(x)\}$ *is optimal.*

Notice that, if (2.10) holds, the sequence in the non-increasing order of $\{\lambda_j w_j F_j(x)\}$ exactly corresponds to the WSEPT sequence. The proof for Theorem 3 is given below.

Proof. Since $g(x)$ is non-decreasing with $g(0) = 0$, we have $g(t - x) = \int_0^{t-x} \mathrm{d}g(y)$. By Fubini's Theorem, and recall that $F_i(x)$ is the distribution function of d_i, we obtain

$$
\begin{aligned}
m_i(t) &= w_i \mathrm{E}[g(t-d_i)I_{\{t>d_i\}}] = w_i \int_{0\leqslant x<t} g(t-x)\mathrm{d}F_i(x) \\
&= w_i \int_{0\leqslant x<t}\int_0^{t-x} \mathrm{d}g(y)\mathrm{d}F_i(x) = w_i \int_{0\leqslant y<t}\int_{0\leqslant x\leqslant t-y} \mathrm{d}F_i(x)\mathrm{d}g(y) \\
&= w_i \int_{0\leqslant y<t} F_i(t-y)\mathrm{d}g(y) = w_i \int_0^t F_i(x)\mathrm{d}g_t(x),
\end{aligned}
$$

where $g_t(x) = -g(t-x)$, which is a non-decreasing function on $[0,t]$ for any $t \geqslant 0$. Hence

$$
\begin{aligned}
\lambda_i[m_i(b) - m_i(a)] &= \lambda_i w_i \left\{ \int_0^b F_i(x)\mathrm{d}g_b(x) - \int_0^a F_i(x)\mathrm{d}g_a(x) \right\} \\
&= \int_a^b \lambda_i w_i F_i(x)\mathrm{d}g_b(x)\mathrm{d}x + \\
&\quad + \int_0^a \lambda_i w_i F_i(x)[\mathrm{d}g_b(x) - \mathrm{d}g_a(x)]. \qquad (2.11)
\end{aligned}
$$

Because $g(t)$ is convex, its increment $g(t+\Delta) - g(t)$ is non-decreasing in t for $\Delta > 0$, which implies

$$
\begin{aligned}
g_b(y) - g_b(x) &= g(b-x) - g(b-y) = g(b-y+\Delta) - g(b-y) \\
&\geqslant g(a-y+\Delta) - g(a-y) = g(a-x) - g(a-y) \\
&= g_a(y) - g_a(x)
\end{aligned}
$$

for $0 \leqslant x < y \leqslant a < b$, where $\Delta = y - x$. Thus, for $a < b$, g_b has increments greater than or equal to those of g_a. As a result,

$$
\int_0^a \phi(x)\mathrm{d}g_a(x) \leqslant \int_0^a \phi(x)\mathrm{d}g_b(x), \quad \text{or} \quad \int_0^a \phi(x)[\mathrm{d}g_b(x) - \mathrm{d}g_a(x)] \geqslant 0,
$$

for any non-negative measurable function $\phi(x)$ on $[0,a]$. Consequently, if $\lambda_i w_i F_i(x) \leqslant \lambda_j w_j F_j(x)$ for $x \geqslant 0$, then

$$
\int_0^a \lambda_i w_i F_i(x)\ [\mathrm{d}g_b(x) - \mathrm{d}g_a(x)] \leqslant \int_0^a \lambda_j w_j F_j(x)[\mathrm{d}g_b(x) - \mathrm{d}g_a(x)]. \qquad (2.12)
$$

Moreover, as $g_b(x)$ is non-decreasing on $[0,t]$,

$$
\int_a^b \lambda_i w_i F_i(x)\ \mathrm{d}g_b(x) \leqslant \int_a^b \lambda_j w_j F_j(x)\ \mathrm{d}g_b(x). \qquad (2.13)
$$

Now, if $i > j$, then $\lambda_i w_i F_i(x) \leqslant \lambda_j w_j F_j(x)$ for $x \geqslant 0$ by the condition of the theorem. It then follows from (2.11)–(2.13) that

$$\lambda_i[m_i(b)-m_i(a)]\leqslant\lambda_j[m_j(b)-m_j(a)]\quad \forall a<b,\ i>j.$$

Thus by Theorem 1, the sequence $(1,2,\ldots,n)$ minimizes the ETC(λ) in (1.2).

Finally, if $\lambda_i w_i F_i(x)\geqslant\lambda_j w_j F_j(x)$ for $x\geqslant 0$, then $\lambda_i w_i\geqslant\lambda_j w_j$ as $x\to\infty$. Consequently, given the existence of an order between $\{\lambda_j w_j F_j(x)\}$, a sequence in the non-increasing order of $\{\lambda_j w_j\}$, i.e., the WSEPT, is optimal. □

In order to state a corollary of Theorem 3, we first give the definition for the 'stochastic order' as follows.

DEFINITION. *For two random variables X and Y, we say that 'X is stochastically less than or equal to Y', and write $X\leqslant_{st}Y$ if* $\Pr(X>a)\leqslant\Pr(Y>a)$ *for all real values a.*

COROLLARY TO THEOREM 3 *Let $g(\cdot)$ satisfy the conditions in Theorem 3. If $\lambda_i w_i\geqslant\lambda_j w_j$* implies *$d_i\leqslant_{st}d_j$, then a sequence in non-increasing order of $\{\lambda_j w_j\}$ is optimal.*

Proof. By the condition of the corollary and the definition for the stochastic order, we have that $\lambda_i w_i\geqslant\lambda_j w_j$ implies $F_i(x)\geqslant F_j(x)$ for all $x\geqslant 0$. As a result, a non-increasing order exists between $\{\lambda_j w_j F_j(x)\}$ and is equivalent to the non-increasing order of $\{\lambda_j w_j\}$, so the corollary follows immediately from Theorem 3. □

EXAMPLE 4. Both $g(x)=x$ and $g(x)=x^2$ satisfies the conditions of Theorem 3. Hence if the compatibility condition in the corollary to Theorem 3 holds, then a sequence in non-increasing order of $\{\lambda_i w_i\}$, or equivalently, in non-decreasing stochastic order of $\{d_i\}$, minimizes both the expected weighted sum of tardinesses $\mathrm{E}[\sum_{i:C_i>d_i}w_i(C_i-d_i)]$ and the expected weighted sum of squared tardinesses $\mathrm{E}[\sum_{i:C_i>d_i}w_i(C_i-d_i)^2]$. (This is not true for the expected weighted number of tardy jobs. Note that $g(x)=1$ does not satisfy the conditions of Theorem 3 because $g(0)\neq 0$.)

Remark 2. The assumption that the weights w_i are deterministic in Theorems 2 and 3 can be relaxed. If w_i are random variables independent of $\{p_i\}$ and $\{f_i(t)\}$, then the two theorems still hold with w_i being replaced by $\mathrm{E}[w_i]$ in the results.

Condition (2.10) is weaker than the agreeable condition between $\{\lambda_j w_j\}$ and $\{d_i\}$. If for some $i\neq j$, $\lambda_i w_i>\lambda_j w_j$ but $d_i\leqslant_{st}d_j$ fails, a sequence in non-increasing order of $\{\lambda_j w_j\}$ could still be optimal. We illustrate this in the following example:

EXAMPLE 5. Suppose $d_i \sim \exp(\delta_i)$ so that $F_i(x) = 1 - e^{-\delta_i x}$. We show below that an order exists between $\{\lambda_j w_j F_j(x)\}$ if and only if $\{\lambda_j w_j\}$ have the same order as $\{\lambda_j w_j \delta_j\}$. To see this, let $\lambda_i w_i \geqslant \lambda_j w_j$ and $\lambda_i w_i \delta_i \geqslant \lambda_j w_j \delta_j$. We show that $\lambda_i w_i F_i(x) \geqslant \lambda_j w_j F_j(x)$ for $x > 0$ below. Consider the following two cases:

Case 1: $\delta_i < \delta_j$. It is easy to see that $(1 - e^{-x})/x$ is a decreasing function of x on $(0, \infty)$. Hence $\delta_i < \delta_j$ and $\lambda_i w_i \delta_i \geqslant \lambda_j w_j \delta_j$ imply, for $x > 0$,

$$\frac{F_i(x)}{F_j(x)} = \frac{1 - e^{-\delta_i x}}{1 - e^{-\delta_j x}} > \frac{\delta_i x}{\delta_j x} = \frac{\delta_i}{\delta_j} \geqslant \frac{\lambda_j w_j}{\lambda_i w_i},$$

or equivalently, $\lambda_i w_i F_i(x) > \lambda_j w_j F_j(x)$.

Case 2: $\delta_i \geqslant \delta_j$. Then $F_i(x) \geqslant F_j(x)$ for $x \geqslant 0$, which together with $\lambda_i w_i \geqslant \lambda_j w_j$ leads immediately to $\lambda_i w_i F_i(x) \geqslant \lambda_j w_j F_j(x)$.

Conversely, if $\lambda_i w_i F_i(x) \geqslant \lambda_j w_j F_j(x)$ for $x \geqslant 0$, then letting $x \to \infty$ yields $\lambda_i w_i \geqslant \lambda_j w_j$. Furthermore,

$$1 \leqslant \frac{\lambda_i w_i F_i(x)}{\lambda_j w_j F_j(x)} = \frac{\lambda_i w_i (1 - e^{-\delta_i x})}{\lambda_j w_j (1 - e^{-\delta_j x})} \longrightarrow \frac{\lambda_i w_i \delta_i}{\lambda_j w_j \delta_j} \quad \text{as } x \downarrow 0.$$

Hence $\lambda_i w_i \delta_i \geqslant \lambda_j w_j \delta_j$.

Thus, we have shown that $\lambda_i w_i F_i(x) \geqslant \lambda_j w_j F_j(x)$ for $x \geqslant 0$ if and only if $\lambda_i w_i \geqslant \lambda_j w_j$ and $\lambda_i w_i \delta_i \geqslant \lambda_j w_j \delta_j$. As a results, even if $\lambda_i w_i > \lambda_j w_j$ but $\delta_i < \delta_j$ (so that $d_i \leqslant_{st} d_j$ fails), a sequence in non-increasing order of $\{\lambda_j w_j\}$ would still be optimal if we have $\lambda_i w_i \delta_i \geqslant \lambda_j w_j \delta_j$ for such i and j.

3. Examples of applications

To highlight the generality and applicability of our results, we provide two examples below. The first example takes into account random price variations and interest accrual of capitals, while the second one allows a deadline in addition to the due dates.

EXAMPLE 6. A company produces a variety of goods for sale. While the current price of a product is known, the future price is uncertain and expected to decline over time due to fading popularity and advancement of technology. This applies particularly to fashion products (e.g., toys, clothes), entertainment products (e.g., music, video), and technology products (e.g., computers, softwares). To allow random variations in the future price, we model the price of job i at time t by $a_i h_i(t)$, where a_i is a constant representing the current price and $h_i(t)$ is a stochastic process with $h_i(0)=1$. Assume that $\mathrm{E}[h_i(t)]=u(t)$ is a non-increasing function of t, reflecting a downward trend of price over time.

At the start of production, an amount of capital is invested to produce job i, which is proportional to the current price, namely βa_i, where $0<\beta<1$. Let α denote the interest rate, which is a random variable following an arbitrary distribution. Then the value of the investment for job i at time t is given by $\beta a_i(1+\alpha)^t$. Hence if job i is sold at time t, then its net profit is $a_i h_i(t)-\beta a_i(1+\alpha)^t$. Suppose that each job is sold to a retailer upon its completion, then the total net profit from a set of n jobs is

$$\sum_{i=1}^{n}\left[a_i h_i(C_i)-\beta a_i(1+\alpha)^{C_i}\right], \tag{3.1}$$

where C_i is the completion time of job i.

If the company produces the goods in sequel, then the problem faced by the management is how to schedule the production optimally so as to maximize the expected total net profit. Define stochastic processes

$$f_i(t)=\beta a_i(1+\alpha)^t-a_i h_i(t), \quad i=1,\dots,n. \tag{3.2}$$

Then the problem of maximizing the total net profit given by (3.1) is equivalent to minimizing $\mathrm{E}[\sum_{i=1}^{n} f_i(C_i)]$. From (3.2) we can see that the mean function of $f_i(t)$ is

$$\begin{aligned} m_i(t)&=\mathrm{E}[f_i(t)]=\beta a_i\mathrm{E}[(1+\alpha)^t]-a_i\mathrm{E}[h_i(t)]+a_i(1-\beta)\\ &=a_i\{\beta\mathrm{E}[(1+\alpha)^t]-u(t)\}. \end{aligned} \tag{3.3}$$

As $\mathrm{E}[h_i(t)]=u(t)$ is a non-increasing function of t, by (3.3) $m_i(t)$ is non-decreasing in t. Write $G(t)=\beta\mathrm{E}[(1+\alpha)^t]-u(t)$ for brevity, which is non-decreasing in t. Then, assuming that the processing times are exponentially distributed with parameters $\lambda_1,\dots,\lambda_n$, it follows from (3.3) that $\lambda_i a_i\geqslant\lambda_j a_j$ implies

$$\begin{aligned} \lambda_i[m_i(t)-m_i(s)]&=\lambda_i a_i[G(t)-G(s)]\geqslant\lambda_j a_j[G(t)-G(s)]\\ &=\lambda_j[m_j(t)-m_j(s)] \end{aligned}$$

for all $t>s$. Thus by Theorem 1, a sequence in non-increasing order of $\{\lambda_j a_j\}$ minimizes $\mathrm{E}[\sum_{i=1}^{n} f_i(C_i)]$, and so is optimal to maximize the expected total net profit.

It is interesting to note in this example that the optimal sequence can be constructed based on the current available price and the rates of the processing times, regardless of future price fluctuations and the cost of interest on the capital.

EXAMPLE 7. A laboratory is contracted to perform reliability tests on n items. The test is to be performed sequentially on a particular facility, with each item tested immediately after the failure of the last item. The failure times of the items are supposed to be independently and exponentially distributed with failure rates $\lambda_1, \ldots, \lambda_n$, respectively. If the test result for item i is reported on or before a due date d_i, the laboratory will receive a payment valued v_i for the test. If it is later than d_i by time t, then the payment will be reduced proportionally to $v_i h(t)$, where $h(t)$ is a stochastic process taking values in $[0, 1]$ and is decreasing in t almost surely. The due dates are assumed to be random variables with a common distribution. In addition, if the facility to perform the tests breaks down, then the tests will not be able to continue and so no payment will be made for items not yet tested by the breakdown time. The breakdown time B is assumed to be exponentially distributed with a rate δ.

The laboratory wishes to schedule the tests optimally so as to maximize the expected total payment it can receive. This is equivalent to minimizing the following objective function (representing the expected total loss):

$$\mathrm{ETL}(\lambda) = \mathrm{E}\left[\sum_{i=1}^{n} \left\{ v_i \tilde{h}(C_i - d_i) I_{\{d_i < C_i \leqslant B\}} + v_i I_{\{C_i > B\}} \right\}\right], \tag{3.4}$$

where $\tilde{h}(t) = 1 - h(t)$ and C_i is the completion time of testing item i. Let

$$f_i(t) = v_i \tilde{h}(t - d_i) I_{\{d_i < t \leqslant B\}} + v_i I_{\{t > B\}}.$$

Then the objective function in (3.4) is equal to $\mathrm{ETL}(\lambda) = \mathrm{E}\left[\sum_{i=1}^{n} f_i(C_i)\right]$. As $h(t)$ is decreasing in t almost surely and $0 \leqslant h(t) \leqslant 1$, $\{f_i(t), t \geqslant 0\}$ is a non-decreasing stochastic process for each i. Let d denote a random variable with the same distribution as d_i. Then the mean function of $f_i(t)$ is

$$\begin{aligned} m_i(t) &= \mathrm{E}[f_i(t)] = v_i \mathrm{E}[\tilde{h}(t - d) I_{\{d < t \leqslant B\}}] + v_i P(t > B), \\ &= v_i \mathrm{E}[\tilde{h}(t - d) \mathrm{e}^{-\delta t} I_{\{t > d\}} + v_i (1 - \mathrm{e}^{-\delta t}), \\ &= v_i \left\{ \mathrm{e}^{-\delta t} \left(\mathrm{E}[\tilde{h}(t - a) I_{\{t > a\}}] - 1 \right) + 1 \right\} = v_i G(t), \end{aligned}$$

where

$$G(t) = 1 - \mathrm{e}^{-\delta t} \left(1 - \mathrm{E}[\tilde{h}(t - d) I_{\{t > d\}}] \right).$$

Since $0 \leqslant \mathrm{E}[\tilde{h}(t-d)I_{\{t>d\}}] \leqslant 1$ and by the assumptions of the problem $\mathrm{E}[\tilde{h}(t-d)I_{\{t>d\}}]$ is non-decreasing in t, $\mathrm{e}^{-\delta t}\left(1-\mathrm{E}[\tilde{h}(t-d)I_{\{t>d\}}]\right)$ is non-increasing in t and so $G(t)$ is non-decreasing in t. Hence, similar to the arguments in Example 6, it follows from Theorem 1 that a sequence in non-increasing order of $\{\lambda_j v_j\}$ is optimal. That is, items with higher ratios of value over mean testing time should be tested earlier.

4. Concluding remarks

The paper generalizes previous studies on stochastic single-machine scheduling to more general situations. The model $1 \mid p_j \sim \exp(\lambda_j) \mid \mathrm{E}\left[\sum f_j(C_j)\right]$ with general stochastic cost functions $f_i(t)$ has been examined. This is a general model, which subsumes many performance measures commonly studied in the literature. We have also illustrated the possible applications of the results with two examples; one addresses the situation where the prices for the products of a company fluctuate randomly and the other is a problem in laboratory testing operations which are subject to not only random due dates, but also an additional random deadline related to the breakdown of the testing facility. These examples show that the models considered here can address some interesting aspects of decision making in practical environments.

We have found that a sequence in the order based on the increments of $\lambda_j \mathrm{E}[f_j(t)]$ is optimal for $1 \mid p_j \sim \exp(\lambda_j) \mid \mathrm{E}\left[\sum f_j(C_j)\right]$. This extends the results of Pinedo (2002) by dropping the compatibility condition (such as $\lambda_j w_j \geqslant \lambda_k w_k \Longleftrightarrow h_j \geqslant_s h_k$), and by allowing stochastic cost functions. We have also obtained new results on the optimal solutions for $1 \mid p_j \sim \exp(\lambda_j), d_j \sim F_j \mid \mathrm{E}\left[\sum w_j g(C_j - d_j) I_{\{C_j > d_j\}}\right]$.

Interesting future research work may include the search of optimal solutions when the job processing times follow more general distributions than the exponential, jobs have different arrival times, and/or the weights are stochastic functions of time. Other important topics include the preemptive version of the problem, stochastic machine breakdowns, as well as the extension of the models to multiple machine scheduling.

5. Acknowledgements

The authors are grateful to the guest editor and the anonymous referee for their valuable suggestions and comments that have helped improve the paper.

References

Bagga, P.C. and Kalra, K.R. (1981), Single machine scheduling problem with quadratic functions of completion times - a modified approach. *Journal of Information and Optimization Sciences*, 2, 103–108.

Boxma, O.J. and Forst, F.G. (1986), Minimizing the expected weighted number of tardy jobs in stochastic flow shops. *Operations Research Letters* 5, 119–126.

Cai, X. and Zhou, X. (2000), Asymmetric earliness and tardiness scheduling with exponential processing times on an unreliable machine. *Annals of Operations Research* 98, 313–331.

Derman, C., Lieberman, G. and Ross, S. (1978), A renewal decision problem. *Management Science* 24, 554–561.

Glazebrook, K.D. (1979), Scheduling tasks with exponential service times on parallel processors. *Journal of Applied Probability* 16, 658–689.

Kampke, T. (1989), Optimal scheduling of jobs with exponential service times on identical parallel processors. *Operations Research* 37, 126–133.

Pinedo, M. (1983), Stochastic scheduling with release dates and due dates. *Operations Research* 31, 559–572.

Pinedo, M. (2002), *Scheduling: Theory, Algorithms, and Systems*, 2nd edn., Prentice Hall, Upper Saddle River, NJ.

Rothkopf, M.H. (1966), Scheduling with random service times. *Management Science* 12, 437–447.

Righter, R. (1994), Scheduling. In: Shacked, M. and Shanthikumar, J.G. (eds.), *Stochastic Orders and Their Applications*, Academic Press, Boston, pp. 381–428.

Sarin, S.C., Erel, E. and Steiner G. (1991), Sequencing jobs on a single machine with a common due date and stochastic processing times. *European Journal of Operational Research* 51, 287–302.

Weiss, G. and Pinedo, M. (1980), Scheduling tasks with exponential services times on non-identical processors to minimize various cost functions. *Journal of Applied Probability* 17, 187–202.