

Chapter 2

Regular Performance Measures

The scheduling field has undergone significant development since 1950s. While there has been a large literature on scheduling problems, the majority, however, is devoted to models characterized by the so-called *regular* performance measures, which are monotone functions of the completion times of the jobs. This is natural, because many problems in real-world applications involve the objective to complete all jobs as early as possible, which result in the requirement of minimizing regular cost functions. Scheduling models aiming to minimize the total flowtime, the makespan, or the total tardiness cost of missing the due dates, are typical examples of regular performance measures.

This chapter covers stochastic scheduling problems with regular performance measures. Section 2.1 is focused on models of minimizing the sum of expected completion time costs. In Sect. 2.2, we consider the problem of minimizing the expected makespan (the maximum completion time). Some basic models with due-date related objective functions are addressed in Sect. 2.3. More general cost functions are considered in Sect. 2.4. Optimal scheduling policies when processing times follow certain classes of distributions are described in Sects. 2.5 and 2.6, respectively. The objective functions considered in Sects. 2.1–2.3 are in fact special cases of those studied in Sects. 2.4–2.6. However, the discussions in the first three sections illustrate the basic techniques commonly employed in the field of stochastic scheduling, including the approach of adjacent job interchange, the argument of induction, and the formation of stochastic dynamic programming.

2.1 Total Completion Time Cost

2.1.1 Single Machine

Suppose that n jobs, all available at time zero, are to be processed by a single machine, with (random) processing time P_i for job i , $i = 1, \dots, n$. If the cost to complete

job i at time t is w_it , where w_i is a constant cost rate for job i , then the expected total cost to complete all jobs is:

$$EWFT(\zeta) = E \left[\sum_{i=1}^n w_i C_i(\zeta) \right] = \sum_{i=1}^n w_i E[C_i(\zeta)], \quad (2.1)$$

where $C_i = C_i(\zeta)$ is the completion time of job i , $i = 1, \dots, n$, under policy ζ . Since jobs are available at time zero, the flowtime of a job is equal to its completion time, and so the measure above is also referred to as the *expected total weighted flowtime*. Minimization of the total weighted flowtime is a basic model in scheduling (cf. Smith, 1956; Rothkopf, 1966a).

Let us first consider the case with no job preemption being allowed. Then, the completion time of job i can be expressed as

$$C_i(\zeta) = \sum_{k \in B_i(\zeta)} P_k, \quad (2.2)$$

where $B_i(\zeta)$ denotes the set of jobs scheduled no later than job i under ζ . Consequently, $E[C_i(\zeta)] = \sum_{k \in B_i(\zeta)} E[P_k]$. It is therefore clear that, if we regard $E[P_i]$ as the processing time for job i , $i = 1, \dots, n$, the problem of minimizing $EWFT(\zeta)$ reduces to one of minimizing the weighted flowtime under deterministic processing times $E[P_i]$. It is well known that the optimal policy for this problem is to sequence the jobs in non-increasing order of the ratio $w_i/E[P_i]$, or according to the so-called *weighted shortest expected processing time* (WSEPT) rule. We can show that, by a standard induction argument, this rule is in fact also optimal in the class of non-preemptive dynamic policies.

Theorem 2.1. *When a set of jobs with random processing times are to be processed by a single machine with no preemption being allowed, the WSEPT rule minimizes $EWFT$ in the class of static policies as well as in the class of dynamic policies.*

Proof. That WSEPT minimizes $EWFT$ in the class of static policies can be shown by an adjacent job interchange argument, a technique commonly adopted in the scheduling field. Denote $p_i = E[P_i]$, $i = 1, \dots, n$, which are now regarded as the (deterministic) processing times of jobs i . When there is only a single machine and no preemption is allowed, any static policy reduces to a sequence to process the n jobs. Suppose a sequence ζ^0 is optimal, which is however not WSEPT. Then, in this sequence there must exist a pair of adjacent jobs $\{j, k\}$, with job k following job j , such that $w_j/p_j < w_k/p_k$. Denote s as the starting time to process job j . Now, create a new sequence ζ' by interchanging the positions of j and k in the sequence ζ^0 . Clearly, the completion times of all jobs before and after the pair $\{j, k\}$ are not affected by the interchange operation. The weighted completion time of $\{j, k\}$ is

$Q^0 = w_j(s + p_j) + w_k(s + p_j + p_k)$ under ζ^0 , and $Q' = w_k(s + p_k) + w_j(s + p_k + p_j)$ under ζ' . It is easy to see that

$$EWFT(\zeta^0) - EWFT(\zeta') = Q^0 - Q' = w_k p_j - w_j p_k = p_j p_k (w_k / p_k - w_j / p_j) > 0,$$

which contradicts the optimality of ζ^0 . Thus an optimal sequence must be WSEPT.

Observe that WSEPT remains an optimal static policy for any subset of jobs that start their processing at any time $s \geq 0$. The claim that WSEPT minimizes $EWFT$ in the class of non-preemptive dynamic policies can be established by an induction argument on k jobs, starting at time s . The claim is true trivially for $k = 1$. Suppose that it is true for $k - 1$ jobs with starting time s' . For k jobs, any non-preemptive dynamic policy must first process a job i , and then process the next $k - 1$ jobs non-preemptively. Denote π^i to be a WSEPT policy for the k jobs excluding job i . Then, the optimal non-preemptive dynamic policy to process the k jobs must be chosen among the k static policies $\{i, \pi^i\}$, $i = 1, 2, \dots, k$, due to the inductive hypothesis. As we have shown that WSEPT is optimal among any static policies for k jobs with any starting time $s \geq 0$, the claim is true for k jobs. ■

Suppose that the *hazard rate* of the processing time P_i is $\lambda_i(x)$ (see the definition in (1.17)). A special but important case is when $\lambda_i(x)$ is a nondecreasing function. In this case, conditional on that job i has been processed for t units of time, the remaining time to complete it is stochastically no greater than the original processing time P_i , and is in fact stochastically nonincreasing in t , as should be in most practical situations. It is easy to see that, if the hazard rate $\lambda_i(x)$ is a nondecreasing function, then job i will never be preempted by any other job once it has been selected for processing. This enables us to extend the result of Theorem 2.1 to the problem with preemption allowed.

Corollary 2.1. *If the hazard rate $\lambda_i(x)$ of any job i , $i = 1, 2, \dots, n$, is a nondecreasing function, then WSEPT is optimal in the class of preemptive dynamic policies.*

The WSEPT rule, however, cannot be easily extended to the multi-machine case with general processing times, even if the weights are identical, i.e. $w_i \equiv w$. The following is an example from Pinedo and Weiss (1987).

Example 2.1. Suppose that all weights $w_i \equiv 1$, and the distributions of processing times P_j belong to one of the following classes:

- Class I: $F_j(x) = \Pr(P_j \leq x) = 1 - (1 - 2p_j)e^{-x} - p_j e^{-x/2}$;
- Class II: $F_j(x) = \Pr(P_j \leq x) = 1 - (1 - p_j)e^{-x} - p_j x e^{-x}$;
- Class III: $\Pr(P_j = 0) = p_j$, $\Pr(P_j = 1) = 1 - 2p_j$, $\Pr(P_j = 2) = p_j$.

Then it is easy to verify that $E[P_j] = 1$ for all jobs, and the variances of processing times are $1 + 4p_j$, $1 + 2p_j$ and $2p_j$ in Class I, II and III, respectively. If there is only

one machine to process all jobs, then WSEPT is optimal (in fact, any sequence is WSEPT because all jobs have $E[P_j] = 1$). However, if there are more than one machine, then by the result of Pinedo and Weiss, the optimal policy is to process the jobs in nondecreasing order of variances. Thus in this case, the WSEPT rule fails to deliver the optimal policy. ■

Some conditions on the processing time distributions are needed in the multi-machine case. This is to be studied in the next subsection, under a more general perspective.

2.1.2 Parallel Machines

Suppose that n jobs are to be processed *non-preemptively* by m identical machines which operate in parallel. The processing times are random variables that can be *stochastically ordered* as $P_1 \leq_{st} \dots \leq_{st} P_n$. Let $\tau_k \geq 0$ denote the time at which machine k becomes available, $k = 1, \dots, m$, and $\tau = (\tau_1, \tau_2, \dots, \tau_m)$. The objective is to find the scheduling policy in the class of non-preemptive policies ζ that maximizes the expected total reward of completing all jobs,

$$R(\tau, \zeta) = E \left[\sum_{i=1}^n r(C_i) \right], \quad (2.3)$$

where $r(t)$ is a convex and decreasing function of t ($0 \leq t < \infty$), and $C_i = C_i(\zeta)$ is the completion time of job i under policy ζ . Note that for $r(t) = -t$ the problem is one of minimizing the expected flowtime as we have discussed in the previous subsection. The reward function $r(t)$ being considered here is identical for all jobs. The case of job dependent reward functions will be introduced later. The exposition below is based on Weber et al. (1986).

We first consider the optimal static list policy. As defined in Chap. 1, A *static list policy* specifies a priority list of jobs to process, and the job at the top of the list will be processed every time a machine is freed. Note that a list policy does not pre-specify the allocation of jobs to the machines. This is different from the completely static policy that specifies, a priori, both the machine allocation and the processing order on each machine.

Now let $L = (k_1, \dots, k_n)$ denote a static list policy, which processes the n jobs in the order $k_1, \dots, k_n$. Let $R(\tau; L)$ denote the expected reward obtained when the jobs are processed according to L . Without loss of generality let $\tau_1 \leq \tau_2 \leq \dots \leq \tau_m$. For convenience we suppose that $r(t)$ is twice differentiable and that the processing times are continuous random variables with density functions.

A static list policy according to the *shortest expected processing time* (SEPT) rule is to process the jobs in non-decreasing order of the expected processing times $E[P_i]$. We will show, in this subsection, that the SEPT rule is optimal to maximize the total reward $R(\tau, L)$. This result will be established through a few lemmas. The first lemma states that for a list policy L , the rate of change of the expected reward with respect to the starting time of any machine is just the expected reward obtained on that machine when the reward function is altered to the derivative $\dot{r}(t)$ of $r(t)$ with respect to t . Let $\dot{R}_i(\tau; L)$ denote the expected reward obtained on machine i when the list policy L is applied and the reward function is $\dot{r}(t)$. Let $dR(\tau; L)/d\tau_i$ denote the right-derivative of $R(\tau; L)$ with respect to τ_i .

Lemma 2.1. *For any static list policy L , $dR(\tau; L)/d\tau_i$ exists and*

$$dR(\tau; L)/d\tau_i = \dot{R}_i(\tau; L), \quad i = 1, \dots, m. \quad (2.4)$$

Proof. The proof is by induction on the number of jobs n . It is clearly true for $n = 0$. Suppose it is true for fewer than n jobs. Let $L = (i_1, i_2, \dots, i_n)$, so that job i_1 is processed first, on machine 1 (due to the assumption that $\tau_1 \leq \dots \leq \tau_m$). Denote $L_1 = (i_2, i_3, \dots, i_n)$, and let $f(t)$ be the density function of P_{i_1} . Then

$$R(\tau; L) = \int_0^\infty f(t) \{r(\tau_1 + t) + R(\tau_1 + t, \tau_2, \dots, \tau_m; L_1)\} dt.$$

Differentiating and using the inductive hypothesis,

$$\frac{dR(\tau; L)}{d\tau_1} = \int_0^\infty f(t) \{\dot{r}(\tau_1 + t) + \dot{R}_1(\tau_1 + t, \tau_2, \dots, \tau_m; L_1)\} dt = \dot{R}_1(\tau; L).$$

Similarly,

$$\frac{dR(\tau; L)}{d\tau_i} = \int_0^\infty f(t) \{\dot{R}_i(\tau_1 + t, \tau_2, \dots, \tau_m; L_1)\} dt = \dot{R}_i(\tau; L), \text{ for } i \neq 1.$$

This completes the inductive step and so the proof of the lemma. ■

The next lemma states that when the reward function is $\dot{r}(t)$ and the scheduling policy is SEPT, the expected reward obtained on a given machine is not reduced if that machine is made to start later and does not increase if any other machine is made to start later. Moreover, if jobs 1 and k are interchanged on machines 1 and 2, the expected reward (with $\dot{r}(t)$ as the reward function) increases.

Lemma 2.2.

(a) *Suppose that L is the SEPT list $(1, 2, \dots, n)$. Then for $j \neq i$ and $n \geq 1$, $\dot{R}_i(\tau; L)$ is non-decreasing in τ_i and non-increasing in τ_j .*

(b) Suppose that L^k is the list $(2, 3, \dots, n)$, omitting some $k \geq 2$. Then for $n \geq 2$,

$$\mathbb{E}[R_1(\tau_1 + P_1, \tau_2 + P_k, \dots, \tau_m; L^k) - R_1(\tau_1 + P_k, \tau_2 + P_1, \dots, \tau_m; L^k)] \leq 0. \quad (2.5)$$

Proof. The proof is again by induction on the number of jobs n . Part (a) is trivial for $n = 1$; and part (b) is trivial for $n = 2$. Suppose that the lemma is true when there are fewer than n jobs to process. We show that it is true for n jobs to process. The inductive step for (b) follows from that for (a) when there are $n - 2$ jobs to process and the fact that if a function $h(x_1, x_k) = R_1(\tau_1 + x_1, \tau_2 + x_k, \dots, \tau_m; L^k)$ is non-decreasing in x_1 and non-increasing in x_k , then $\mathbb{E}[h(P_1, P_k) - h(P_k, P_1)] \leq 0$ for $P_1 \leq_{\text{st}} P_k$.

To establish the inductive step for part (a) we begin by showing that $\dot{R}_i(\tau; L)$ is non-decreasing in τ_i . Without loss of generality, consider $i = 1$ and suppose that $\tau_2 \leq \dots \leq \tau_m$. Let $L_1 = (2, 3, \dots, n)$. Then for $\tau_1 < \tau_2$,

$$\dot{R}_1(\tau; L) = \mathbb{E}[\dot{r}(\tau_1 + P_1) + \dot{R}_1(\tau_1 + P_1, \tau_2, \dots, \tau_m; L_1)].$$

Because $\dot{r}(t)$ is non-decreasing and by the inductive hypothesis, the expression to take expectation is non-decreasing in τ_1 . Thus $\dot{R}_1(\tau; L)$ is non-decreasing in τ_1 within the region $\tau_1 < \tau_2$. If $\tau_1 > \tau_2$, then

$$\dot{R}_1(\tau; L) = \mathbb{E}[\dot{R}_1(\tau_1, \tau_2 + P_1, \dots, \tau_m; L_1)],$$

which is non-decreasing in τ_1 due to the inductive hypothesis.

It remains to consider the change in $\dot{R}_1(\tau; L)$ as τ_1 passes through the value τ_2 . Suppose that $\tau_2 = \dots = \tau_k < \tau_{k+1} \leq \dots \leq \tau_m$ and let L^k be the list $(2, 3, \dots, n)$ after omitting job k . Then the change in $\dot{R}_1(\tau; L)$ as τ_1 passes through the value τ_2 may be written as

$$\dot{R}_1(\tau_2 +, \tau_2, \dots, \tau_m; L) - \dot{R}_1(\tau_2 -, \tau_2, \dots, \tau_m; L).$$

This change equals

$$\begin{aligned} & \mathbb{E}[\dot{r}(\tau_2 + P_k) + \dot{R}_1(\tau_2 + P_k, \tau_2 + P_1, \tau_3, \dots, \tau_m; L^k) \\ & - \dot{r}(\tau_2 + P_1) - \dot{R}_1(\tau_2 + P_1, \tau_2 + P_k, \tau_3, \dots, \tau_m; L^k)], \end{aligned}$$

for $k \leq n$; and $\mathbb{E}[-\dot{r}(\tau_2 + P_1)]$ for $k > n$ (in this case, there are at least n identical machines which are available at τ_2 to process the n jobs).

In both cases, a negative and non-decreasing $\dot{r}(t)$ together with the inductive hypothesis for part (b) implies that the expression for taking expectation is non-negative. This completes the inductive step showing that $\dot{R}_i(\tau; L)$ is non-decreasing in τ_i . By similar arguments we can show that $\dot{R}_i(\tau; L)$ is non-increasing in τ_j , $j \neq i$. ■

The next lemma states that when the reward function is $\dot{r}(t)$, the SEPT list does not produce a greater expected reward on machine 1 (the machine that starts first)

than a policy that schedules the shortest job first on machine 2 (the machine that starts second) and the remaining jobs according to SEPT.

Lemma 2.3. *Suppose that $L = (1, 2, \dots, n)$ is the SEPT list. Let $L_1 = (2, 3, \dots, n)$. Then for $n \geq 2$,*

$$\dot{R}_1(\tau; L) \leq E[\dot{R}_1(\tau_1, \tau_2 + P_1, \dots, \tau_m; L_1)]. \quad (2.6)$$

Proof. The proof is again by induction on n . When $n = 2$ we have

$$\begin{aligned} \dot{R}_1(\tau; L) &= \int_0^\infty f(t) \{ \dot{r}(\tau_1 + t) + \dot{R}_1(\tau_1 + t, \tau_2, \dots, \tau_m; L_1) \} dt \\ &\leq E[\dot{r}(\tau_1 + P_1)] \leq E[\dot{r}(\tau_1 + P_2)] = E[\dot{R}_1(\tau_1, \tau_2 + P_1, \dots, \tau_m; (1))]. \end{aligned}$$

Thus (2.6) holds for $n = 2$. Suppose that the lemma is true when there are fewer than n jobs to process. Let $L_2 = (3, 4, \dots, n)$. If $\tau_1 = \tau_2$, then the lemma is true with equality. If $\tau_1 < \tau_2$, then

$$\begin{aligned} \dot{R}_1(\tau; L) &= E[\dot{r}(\tau_1 + P_1) + \dot{R}_1(\tau_1 + P_1, \tau_2, \dots, \tau_m; L_1)] \\ &\leq E[\dot{r}(\tau_1 + P_1) + \dot{R}_1(\tau_1 + P_1, \tau_2 + P_2, \dots, \tau_m; L_2)] \\ &\leq E[\dot{r}(\tau_1 + P_2) + \dot{R}_1(\tau_1 + P_2, \tau_2 + P_1, \dots, \tau_m; L_2)] \\ &= E[\dot{R}_1(\tau_1, \tau_2 + P_1, \dots, \tau_m; L_1)], \end{aligned}$$

where the first inequality follows from the inductive hypothesis, and the second inequality follows from $\dot{r}(t)$ being non-decreasing and part (b) of Lemma 2.2. This completes the proof of the lemma. $\blacksquare$

The theorem below is the main result, which shows that SEPT is the optimal static list policy and the optimal non-preemptive dynamic policy.

Theorem 2.2. *Suppose that n jobs have processing times which can be stochastically ordered and job preemption is not allowed. Then SEPT maximizes the expected reward $R(\tau, \zeta)$ in the class of static list policies and in the class of non-preemptive dynamic policies.*

Proof. We first establish the optimality of SEPT in the class of static list policies. The proof is by induction on n . The result is trivial for $n = 1$. Suppose that the result is true when there are fewer than n jobs to process. Consider a static list policy, which begins by processing job k ($k > 1$) on machine 1 (the first machine to become available). By the inductive hypothesis it must be optimal to start job 1 next and then start the remaining jobs according to the SEPT list policy L^k , where L^k is $(2, 3, \dots, n)$, omitting job k . Thus amongst those policies that start processing job k first, the best one is the list policy $(k, 1, L^k)$; denoted by $L^{k,1}$. Interchanging jobs k and 1 will generate the list policy $L^{1,k} = (1, k, L^k)$. We shall show that $L^{1,k}$ is better than $L^{k,1}$ in the sense that $\Delta = R(\tau; L^{1,k}) - R(\tau; L^{k,1}) \geq 0$. Assuming this, by

the inductive hypothesis and continuing the interchanging argument, we can show that $L = (1, 2, \dots, n)$ is optimal.

Let $R(\tau; S; c)$ be the expected reward when the policy S is applied, conditional on $P_k = c$. We shall shortly show that $\Delta(c) = R(\tau; L^{1,k}; c) - R(\tau; L^{k,1}; c)$ is non-decreasing in c . If so, we have $\Delta(P_k) \geq_{\text{st}} \Delta(\bar{X}_1)$ for any random variable $\bar{X}_1$ independent of $P_1, \dots, P_n$ and identically distributed as P_1 . By taking the expectation,

$$\Delta = \mathbb{E}[\Delta(P_k)] \geq \mathbb{E}[\Delta(\bar{X}_1)] = 0,$$

where $\mathbb{E}[\Delta(\bar{X}_1)] = 0$ because $\bar{X}_1$ and P_1 are identically distributed. Therefore, the optimality of SEPT will follow once we show that $\Delta(c)$ is non-decreasing in c . It is easy to see that

$$\begin{aligned} R(\tau; L^{1,k}; c) &= \mathbb{E}[I_{\{P_1 < \tau_2 - \tau_1\}} \{r(\tau_1 + P_1) + r(\tau_1 + P_1 + c) + R(\tau_1 + P_1 + c, \tau_2, \dots, \tau_m; L^k)\} \\ &\quad + I_{\{P_1 \geq \tau_2 - \tau_1\}} \{r(\tau_1 + P_1) + r(\tau_2 + c) + R(\tau_1 + P_1, \tau_2 + c, \dots, \tau_m; L^k)\}] \end{aligned}$$

and

$$R(\tau; L^{k,1}; c) = r(\tau_1 + c) + R(\tau_1 + c, \tau_2, \dots, \tau_m; (1) + L^k),$$

where $(1) + L^k$ denotes the list policy with job 1 processed first, then followed by L^k . Differentiation of the above gives

$$\begin{aligned} \frac{dR(\tau; L^{1,k}; c)}{dc} &= \mathbb{E}[I_{\{P_1 < \tau_2 - \tau_1\}} \{\dot{r}(\tau_1 + P_1 + c) + \dot{R}_1(\tau_1 + P_1 + c, \tau_2, \dots, \tau_m; L^k)\} \\ &\quad + I_{\{P_1 \geq \tau_2 - \tau_1\}} \{\dot{r}(\tau_2 + c) + \dot{R}_2(\tau_1 + P_1, \tau_2 + c, \dots, \tau_m; L^k)\}] \\ &= \mathbb{E}[I_{\{P_1 < \tau_2 - \tau_1\}} \{\dot{r}(\tau_1 + P_1 + c) + \dot{R}_1(\tau_1 + P_1 + c, \tau_2, \dots, \tau_m; L^k)\} \\ &\quad + I_{\{P_1 \geq \tau_2 - \tau_1\}} \{\dot{r}(\tau_2 + c) + \dot{R}_1(\tau_2 + c, \tau_1 + P_1, \dots, \tau_m; L^k)\}], \end{aligned}$$

where the second equality follows by the fact that the machines are identical, and

$$\begin{aligned} \frac{dR(\tau; L^{k,1}; c)}{dc} &= \dot{r}(\tau_1 + c) + \dot{R}_1(\tau_1 + c, \tau_2, \dots, \tau_m; (1) + L^k) \\ &\leq \mathbb{E}[\dot{r}(\tau_1 + c) + \dot{R}_1(\tau_1 + c, \tau_2 + P_1, \dots, \tau_m; L^k)], \end{aligned}$$

where the inequality follows from Lemma 2.3. Thus

$$\begin{aligned} \frac{d\Delta(c)}{dc} &\geq \mathbb{E}[I_{\{P_1 < \tau_2 - \tau_1\}} \{\dot{r}(\tau_1 + P_1 + c) + \dot{R}_1(\tau_1 + P_1 + c, \tau_2, \dots, \tau_m; L^k)\} \\ &\quad + I_{\{P_1 \geq \tau_2 - \tau_1\}} \{\dot{r}(\tau_2 + c) + \dot{R}_1(\tau_2 + c, \tau_1 + P_1, \dots, \tau_m; L^k)\} \\ &\quad - \{\dot{r}(\tau_1 + c) + \dot{R}_1(\tau_1 + c, \tau_2 + P_1, \dots, \tau_m; L^k)\}]. \end{aligned}$$

Using part (a) of Lemma 2.2, $\tau_1 \leq \tau_2$ and non-decreasing $\dot{r}(t)$, we can see that the expression above to take expectation is non-negative. This shows that $\Delta(c)$ is non-decreasing in c , and completes the inductive step. Consequently, the optimality of SEPT in the class of static list policies is proven.

Observe that SEPT is an optimal static list policy for any subset of k jobs and starting times $(\tau_1, \tau_2, \dots, \tau_m)$ on the m machines. Thus, that SEPT is optimal in the class of non-preemptive dynamic policies can be proven by a similar induction argument as in the proof of Theorem 2.1. ■

Based on the observation that a job will never be preempted once it is started if its processing time has a non-decreasing hazard rate, we have

Corollary 2.2. *If all jobs have non-decreasing hazard rate functions, then SEPT is optimal to maximize $R(\tau, \zeta)$ in the class of preemptive dynamic policies.*

By examining the proof of Theorem 2.2, one can see that the result is still true for some models in which the reward obtained on completing each job differs from job to job, under a compatibility condition as specified in the theorem below.

Theorem 2.3. *If $\dot{r}_i(t)$ are job-dependent, the results of Theorem 2.2 and Corollary 2.2 remain valid under the following conditions: the processing times of the jobs can be stochastically ordered and $P_i \leq_{\text{st}} P_j \Rightarrow \dot{r}_i(t) \leq \dot{r}_j(t)$ for all t and i, j .*

Consider $r_i(t) = -w_i t$. Then the problem of maximizing $R(\tau, \zeta)$ reduces to one of minimizing the expected weighted flowtime. The compatibility condition of Theorem 2.3 reduces to $P_i \leq_{\text{st}} P_j \Rightarrow w_i \geq w_j$ for any i, j .

2.2 Makespan

In this section we focus on the problem to minimize the expected makespan (the maximum completion time amongst all jobs). We assume that the processing time for any job does not depend on the scheduling policy.¹ The makespan measure is thus trivial when there is only one machine (because the expected makespan is equal to the total expected processing time). We will consider the problem with multiple parallel machines in this section, and show that the optimal policy should process the jobs in the non-increasing order of the expected processing times $E[P_i]$, that is, according to the *longest expected processing time* (LEPT) rule. This scheduling

¹ An alternative assumption is that the setup time, if any, before processing any job does not depend on the processing sequence and so has been included in the processing time; This is an assumption we implicitly make throughout the whole book.

policy is in sharp contrast to the SEPT rule, and is reasonable for the makespan measure because it proposes, in a way, to deal with longer jobs first, so that shorter jobs can be used for allocation to different machines to reduce their idle times. The proof of the optimality of LEPT is limited to the problem where the processing times are all exponentially distributed.

More specifically, suppose that n jobs are to be processed on m identical machines operating in parallel. All jobs are available for processing at time zero. The processing times P_i , $i = 1, 2, \dots, n$, are exponentially distributed with rates $\lambda_1, \lambda_2, \dots, \lambda_n$. Preemption is allowed, but will not be needed under the LEPT rule. This is due to the memoryless property of the exponential processing times, with which a job always remains to be of the highest priority once it is selected. Without loss of generality, suppose that the jobs have been numbered so that $\lambda_1 \leq \dots \leq \lambda_n$.

The approach of dynamic programming will be used to establish the optimality of LEPT. The proof below is based on Weber (1982a).

Theorem 2.4. *The LEPT rule minimizes the expected makespan in the class of dynamic policies.*

Proof. To simplify the notation, the proof is conducted for the case of two machines. A similar approach can be used for the general case with any number of machines. The proof is by induction on n . Suppose that the claim is true for fewer than n jobs. We will show that it is true when there are n jobs to process. It is clear that the optimal policy is non-preemptive due to the memoryless property of the exponential processing times, and that if processing jobs i and j is optimal at time t , then this will remain to be the case until one of them is completed.

Let U^J denote the expected value of the remaining time needed to complete all n jobs under an optimal policy, given that the jobs in the subset $J = \{i_1, \dots, i_\ell\}$ have already been completed. If $J = \emptyset$, then U^J is denoted by U . Let V^J denote the same quantity under the LEPT policy, and $V = V^\emptyset$.

Conditioning on the first job completion and by the inductive hypothesis that LEPT is optimal when there are fewer than n jobs, we obtain

$$U = \min_{i,j} \left\{ \frac{1}{\lambda_i + \lambda_j} + \frac{\lambda_i}{\lambda_i + \lambda_j} V^{\{i\}} + \frac{\lambda_j}{\lambda_i + \lambda_j} V^{\{j\}} \right\}, \quad (2.7)$$

where the first term on the RHS is the expected time until the first job is completed, and the second (third) term is equal to the probability that job i (j) is the first job to be completed, multiplied by the expected remaining time needed to complete the $n - 1$ remaining jobs under LEPT. It is easy to see that (2.7) is equivalent to

$$0 = \min_{i,j} \left\{ 1 + \lambda_i(V^{\{i\}} - V) + \lambda_j(V^{\{j\}} - V) + (\lambda_i + \lambda_j)(V - U) \right\}. \quad (2.8)$$

Since λ_1 and λ_2 are the two smallest values of λ_k and $V \geq U$, the fourth term on the RHS of (2.8) is minimized by $i, j = 1, 2$. Hence to show that LEPT is optimal it suffices to show that $(i, j) = (1, 2)$ also minimize the sum of the second and third terms. Define

$$V_i = \lambda_i(V^{\{i\}} - V) \quad \text{and} \quad D_{ij} = V_i - V_j.$$

The result of Lemma 2.4 below shows that if $\lambda_i < \lambda_j$, then $D_{ij} \leq 0$. Hence the sum of the second and third terms on the RHS of (2.8), $V_i + V_j$, is minimized by $(i, j) = (1, 2)$ and the induction is complete. ■

In what follows we shall consider V , V_i and D_{ij} as functions of $\lambda_1, \dots, \lambda_n$. Define V_i^J and D_{ij}^J similarly to V_i and D_{ij} , except that i and j are excluded from J . For example, $V_i^J = \lambda_i(V^{Ji} - V^J)$, where J_i denotes the list J with job i appended.

Lemma 2.4. *Suppose $\lambda_i < \lambda_j$. Then*

$$D_{ij} \leq 0 \quad \text{and} \quad \frac{dD_{12}}{d\lambda_1} \geq 0. \quad (2.9)$$

Proof. The proof is by induction on n . When $n = 2$, $D_{ij} = (\lambda_i/\lambda_j) - (\lambda_j/\lambda_i)$ and the lemma is true. If i and j are the two smallest indices not in the subset J then jobs i and j will be processed first. Conditioning on the first job completion, we have

$$V^J = \frac{1}{\lambda_i + \lambda_j} + \frac{\lambda_i}{\lambda_i + \lambda_j} V^{Ji} + \frac{\lambda_j}{\lambda_i + \lambda_j} V^{Jj},$$

or $(\lambda_i + \lambda_j)V^J = 1 + \lambda_i V^{Ji} + \lambda_j V^{Jj}$. This together with the definition of V_i^J allows us to derive the following identities.

$$\begin{aligned} (\lambda_1 + \lambda_2 + \lambda_3)V_1 &= \lambda_1(\lambda_1 + \lambda_2 + \lambda_3)V^1 - \lambda_1(\lambda_1 + \lambda_2 + \lambda_3)V \\ &= \lambda_1(1 + \lambda_1 V^1 + \lambda_2 V^{12} + \lambda_3 V^{13}) - \lambda_1(1 + \lambda_1 V^1 + \lambda_2 V^2 + \lambda_3 V) \\ &= \lambda_1(\lambda_3 V^{13} - \lambda_3 V^1) + \lambda_2(\lambda_1 V^{12} - \lambda_1 V^2) + \lambda_3 V_1, \end{aligned}$$

or $(\lambda_1 + \lambda_2)V_1 = \lambda_1 V_3^1 + \lambda_2 V_1^2$. We can establish the following similarly.

$$(\lambda_1 + \lambda_2)V_2 = \lambda_1 V_2^1 + \lambda_2 V_3^2, \quad (\lambda_1 + \lambda_2)V_i = \lambda_1 V_i^1 + \lambda_2 V_i^2, \quad i = 3, \dots, n.$$

Combining these we have

$$D_{12} = \frac{\lambda_1}{\lambda_1 + \lambda_2} D_{32}^1 + \frac{\lambda_2}{\lambda_1 + \lambda_2} D_{13}^2 \quad (2.10)$$

and

$$D_{2i} = \frac{\lambda_1}{\lambda_1 + \lambda_2} D_{2i}^1 + \frac{\lambda_2}{\lambda_1 + \lambda_2} D_{3i}^2, \quad i = 3, \dots, n. \quad (2.11)$$

The inductive hypothesis states that (2.9) is true when there are fewer than n jobs to process, and this hypothesis for the first inequality in (2.9) implies that both $D_{13}^2 \leq 0$ and $D_{23}^1 \leq 0$ are true when there are n jobs to process. The hypothesis for the second inequality in (2.9) similarly implies that $dD_{13}^2/d\lambda_1 \geq 0$. By integrating this with respect to λ_1 we have $D_{13}^2 \leq D_{23}^1 = -D_{32}^1 \leq 0$. Since $\lambda_1 \leq \lambda_2$, we can see that $D_{12} \leq 0$. The inductive hypothesis also implies that D_{2i}^1 and D_{3i}^2 are nonpositive, and thus $D_{2i} \leq 0$. Combining these inequalities establishes the inductive step for the first inequality in (2.9). The inductive step for the second inequality in (2.9) is established by differentiating the RHS of (2.10) with respect to λ_1 and then using the inductive hypothesis to show that every term is nonnegative. ■

Theorem 2.4 shows that LEPT is optimal in the class of preemptive dynamic policies. The LEPT policy, however, requires no job preemption, and is thus also optimal in the class of static list policies and in the class of non-preemptive dynamic policies.

2.3 Regular Costs with Due Dates

We consider a few basic due-date related models in this section. More results on stochastic scheduling problems involving due dates can be found in Sects. 2.4–2.6, as special cases of general cost functions. The analyses in this section are based on Pinedo (1983) and Emmons and Pinedo (1990).

2.3.1 Weighted Number of Tardy Jobs

Suppose that n jobs are to be processed on a single machine. The processing time of job i is a random variable P_i , exponentially distributed with rate λ_i . Job i has a due date D_i , which is a random variable with distribution F_i . If the job is completed later than the due date, then there is a constant tardy cost w_i , which is also called the weight for job i . The objective is to determine a scheduling policy ζ to process the n jobs so as to minimize the objective function:

$$EWNT(\zeta) = E \left[\sum_{i=1}^n w_i I_{\{C_i \geq D_i\}} \right], \quad (2.12)$$

where C_i is the completion time of job i under the scheduling policy ζ and D_i is the due date of job i . The objective function $EWNT(\zeta)$ represents the expected total tardy penalty of missing due dates: if job i misses its due date D_i , it incurs a fixed penalty w_i . The model is also commonly referred to as one of minimizing the

expected weighted number of tardy jobs. If $w_i = 1$ for all jobs i , it reduces to one of minimizing the *expected total number of tardy jobs*.

Since the processing times are exponentially distributed with means $1/\lambda_j$, the WSEPT rule processes the jobs in the non-increasing order of $\lambda_j w_j$. The following theorem shows the optimality of WSEPT in the class of non-preemptive static list policies.

Theorem 2.5. *If all due dates have the same cumulative distribution function (cdf) F , then processing the jobs in non-increasing order of $\lambda_j w_j$ is optimal to minimize $EWNT(\zeta)$ in the class of non-preemptive static list policies.*

Proof. Consider first the case with two jobs only, where there are only two possible job sequences (1, 2) and (2, 1). Then

$$\begin{aligned} EWNT(1, 2) &= w_1 \Pr(X_1 > D_1) + w_2 \Pr(X_1 + X_2 > D_2) \\ &= w_1 \int_0^\infty e^{-\lambda_1 x} f(x) dx + \frac{w_2}{\lambda_1 - \lambda_2} \int_0^\infty (\lambda_1 e^{-\lambda_2 x} - \lambda_2 e^{-\lambda_1 x}) f(x) dx. \end{aligned}$$

Similarly, we can get the expression for $EWNT(2, 1)$, and show that

$$EWNT(1, 2) - EWNT(2, 1) = (\lambda_2 w_2 - \lambda_1 w_1) \int_0^\infty K(x) f(x) dx, \quad (2.13)$$

where $K(x) = (e^{-\lambda_1 x} - e^{-\lambda_2 x})/(\lambda_2 - \lambda_1) > 0$ for $x > 0$. Thus

$$EWNT(1, 2) \leq EWNT(2, 1) \quad \text{when} \quad \lambda_1 w_1 \geq \lambda_2 w_2.$$

The argument can be extended to the case with n jobs. Compare the sequence $1, \dots, i-1, i+1, i, i+2, \dots, n$ with the sequence $1, \dots, i-1, i, i+1, i+2, \dots, n$. It is clear that the expected tardy penalties of jobs $1, \dots, i-1$ and $i+2, \dots, n$ are the same in the two sequences. Therefore, we need only compare the sum of the expected tardy penalties of jobs i and $i+1$ in the two sequences. Conditional on the time that job $i-1$ is finished, the problem of comparing the sum of the expected tardy penalties of jobs i and $i+1$ in the two sequences reduces to the case of two jobs as described above. We can thus show that the total expected tardy penalty can be reduced if the jobs are not processed in the non-increasing order of $\lambda_j w_j$. ■

We now consider the optimal policy in the class of dynamic policies. We limit our consideration here to the case with a common, deterministic due date d for all jobs. Generally speaking, the argument to show that a static list policy is optimal also in the class of dynamic policies is mainly based on the observation that it is not necessary to alter the processing priority of a job once a processing policy has been applied (consequently, the optimal static sequence is also an optimal dynamic policy). If the job due dates are random and distinct, this argument may not be valid. This can be seen from the following example.

Example 2.2. Consider a case with random due dates and optimal static sequence $(1, 2, \dots, n)$. Suppose that when the machine is processing job 1, the due date of job 2 is realized. As a result, when job 1 is finished, job 2 (which would follow job 1 under the static policy) should be preempted by job 3, because job 2 has already been tardy and it will incur a fixed tardy penalty w_2 no matter when it is processed. As a result, job 2 should be re-sequenced to the end of the job sequence. ■

This example shows that the optimal static sequence may no longer be optimal even in the class of non-preemptive dynamic policies, if the due dates are random and distinct. However, if all jobs have the same deterministic due date, then the claim that an optimal static policy WSEPT is also an optimal dynamic policy can be established.

Theorem 2.6. *Suppose that all jobs have the common fixed due date d . Then, processing the jobs in the non-increasing order of $\lambda_j w_j$ is optimal to minimize $EWNT(\zeta)$ in the class of non-preemptive dynamic policies and in the class of preemptive dynamic policies.*

Proof. We first show that the static policy WSEPT is optimal in the class of non-preemptive dynamic policies. The proof is by induction on n . It follows from (2.13) that the claim is true for $n = 2$. Suppose that the claim holds for $k - 1$ jobs, which start at time t . We consider the case with k jobs, to start at time $t' < t$. According to the inductive hypothesis, a non-preemptive dynamic policy must first process one job at time t' , and after finishing this job, process the remaining $k - 1$ jobs in the non-increasing order of $\lambda_j w_j$. The optimal policy must select the job with the highest $\lambda_j w_j$ to be processed first. Otherwise, this job must be the second to be processed due to the WSEPT order for the $k - 1$ jobs. Then, following a similar analysis as in the proof of Theorem 2.5, we can show that interchanging the positions of the first two jobs will reduce the expected value of the objective function. This completes the inductive step.

The proof that the static policy WSEPT is optimal also in the class of preemptive dynamic policies can be established using the memoryless property of the processing times, which ensures that a job will never be preempted once it is selected for processing, because under the WSEPT rule, $\lambda_j w_j$ and the due date d do not depend on time t . Thus preemption is not needed even though it is allowed, and consequently, the optimal non-preemptive dynamic policy is optimal in the class of preemptive dynamic policies. ■

The WSEPT rule remains optimal for the problem where the jobs have a common due date D , which is a random variable with an arbitrary distribution F . The proof can be found in Derman et al. (1978).

We now consider the problem with m identical machines operating in parallel. Again, the objective is to minimize $EWNT(\zeta)$; that is, to minimize the expected sum of tardy costs, where the tardy cost for a job i is a fixed penalty w_i when it

misses its due date. We can show that, under some quite restrictive conditions, the optimal non-preemptive static list policy can be determined by solving a deterministic assignment problem. The idea is to optimally assign the n jobs to the n positions of a list policy, under certain assignment costs. The assignment costs can be pre-calculated if the processing times and the due dates satisfy some conditions as shown below.

First, consider the case where all processing times are deterministic and equal (thus, without loss of generality, they can be assumed to equal 1). The weights w_i are job dependent, and the due dates D_i of the jobs are random variables, following arbitrary distributions $F_i(x)$, $i = 1, 2, \dots, n$. Job preemption is not allowed. Then, under a static list policy, the first batch of m jobs start on the m machines at time zero, and complete their processing at time 1. Thus, the probability for a job j in this batch to be overdue is $F_j(1)$, and so the expected cost is $w_j F_j(1)$. Similarly, the completion time of the second batch of m jobs complete their processing at time 2, and the corresponding expected cost is $w_j F_j(2)$, etc. To summarize, we have the following theorem.

Theorem 2.7. *Suppose that m parallel identical machines are to process n jobs with processing times being deterministic and equal to 1. Then, the optimal non-preemptive static list policy to minimize $EWNT(\zeta)$ can be obtained by solving a deterministic assignment problem with the following cost matrix: If job j is assigned to position i in the static list, where $km + 1 \leq i \leq (k + 1)m$, then the cost is $w_j F_j(k + 1)$, $k = 0, 1, 2, \dots$. The optimal assignment solution that minimizes the total assignment cost specifies the optimal non-preemptive static list policy.*

We now consider random processing times that are i.i.d. exponential with mean 1. The due date of job i is also exponential with rate μ_i , $i = 1, 2, \dots, n$. The due dates do not have to be independent. Again, we can show that the optimal non-preemptive static list policy can be obtained by solving a deterministic assignment problem. Clearly, the first batch of m jobs in a static list policy start their processing on the m machines at time 0. The probability for a job j amongst this batch to miss its due date is $\mu_j / (1 + \mu_j)$, and so the expected cost is $w_j \mu_j / (1 + \mu_j)$. Job j in position i of the list policy, $i = m + 1, \dots, n$, has to wait for $i - m$ job completions before its processing starts. Given that all machines are busy, the time between successive completions is exponentially distributed with rate m . Thus, the probability that a job starts before its due date is $(m / (m + \mu_j))^{i-m}$ and so the probability that it completes before its due date is $(m / (m + \mu_j))^{i-m} / (1 + \mu_j)$. Consequently, the probability for the job to miss its due date is $1 - (m / (m + \mu_j))^{i-m} / (1 + \mu_j)$, and the expected cost is $w_j (1 - (m / (m + \mu_j))^{i-m} / (1 + \mu_j))$. To summarize, we have the following theorem.

Theorem 2.8. *Suppose that m parallel identical machines are to process n jobs, where the processing times are i.i.d. exponential with mean 1, and the due dates of the jobs are exponential random variables with rate μ_i , $i = 1, 2, \dots, n$. Then, the optimal non-preemptive static list policy to minimize $EWNT(\zeta)$ can be obtained by*

solving a deterministic assignment problem with the following cost matrix: If job j is assigned to position $i \in \{1, 2, \dots, m\}$ in the static list, then the expected cost is $w_j \mu_j / (1 + \mu_j)$; If job j is assigned to position $i \in \{m + 1, \dots, n\}$, then the expected cost is

$$w_j \left(1 - \left(\frac{m}{m + \mu_j} \right)^{i-m} \frac{1}{1 + \mu_j} \right).$$

The optimal assignment solution that minimizes the total assignment cost specifies the optimal non-preemptive static list policy.

2.3.2 Total Weighted Tardiness

Suppose that n jobs are to be processed on a single machine. The processing time of job i is a random variable P_i , exponentially distributed with rate λ_i . Job i has a due date D_i , which is a random variable with cdf F_i . If job i is completed at time $C_i > D_i$ (missing its due date), then it incurs a tardiness cost $w_i T_i$, where w_i is the unit tardiness cost (which is also called the weight of job i) and $T_i = \max\{C_i - D_i, 0\}$ is the tardiness. The objective is to determine a scheduling policy ζ to process the n jobs so as to minimize the expected total tardiness cost:

$$EWT(\zeta) = E \left[\sum_{i=1}^n w_i T_i \right]. \quad (2.14)$$

This objective function is also referred to as the *expected total weighted tardiness*.

Since the processing times are exponentially distributed with means $1/\lambda_j$, the WSEPT rule processes the jobs in the non-increasing order of $\lambda_j w_j$. We will show that WSEPT is optimal in the class of static list policies, under a compatibility condition that requires $\lambda_k w_k \geq \lambda_l w_l \Rightarrow D_k \leq_{\text{st}} D_l$, i.e., the due date of job k is stochastically smaller than the due date of job l for every pair of jobs k, l such that $\lambda_k w_k \geq \lambda_l w_l$. Note that if the jobs have a common due date distribution, the compatibility condition is satisfied automatically.

Theorem 2.9. *When $\lambda_k w_k \geq \lambda_l w_l \Rightarrow D_k \leq_{\text{st}} D_l$, sequencing the jobs in the non-increasing order of $\lambda_j w_j$ is optimal to minimize $EWT(\zeta)$ in the class of non-preemptive static list policies.*

Proof. Consider first the case with two jobs only, so there are only two possible job sequences (1, 2) and (2, 1). Then

$$\begin{aligned}
EWT(1,2) &= \frac{w_1}{\lambda_1} \Pr(X_1 > D_1) + \frac{w_2}{\lambda_1} \Pr(X_1 > D_2) + \frac{w_2}{\lambda_2} \Pr(X_1 + X_2 > D_2) \\
&= \frac{w_1}{\lambda_1} \int_0^\infty e^{-\lambda_1 x} f_1(x) dx + \frac{w_2}{\lambda_1} \int_0^\infty e^{-\lambda_1 x} f_2(x) dx \\
&\quad + \frac{w_2}{\lambda_2(\lambda_1 - \lambda_2)} \int_0^\infty (\lambda_1 e^{-\lambda_2 x} - \lambda_2 e^{-\lambda_1 x}) f_2(x) dx.
\end{aligned}$$

Similarly, we can derive the expression for $EWT(2,1)$. These lead to

$$EWT(1,2) - EWT(2,1) = -w_1 \lambda_1 \int_0^\infty H(x) f_1(x) dx + w_2 \lambda_2 \int_0^\infty H(x) f_2(x) dx,$$

where

$$H(x) = \frac{\lambda_2 e^{-\lambda_1 x} - \lambda_1 e^{-\lambda_2 x}}{\lambda_1 \lambda_2 (\lambda_2 - \lambda_1)}$$

decreases monotonically from $1/\lambda_1 \lambda_2$ to 0 on $[0, \infty]$. Hence, by the property of stochastic ordering, if $D_1 \leq_{st} D_2$, then

$$\int_0^\infty H(x) f_1(x) dx \geq \int_0^\infty H(x) f_2(x) dx.$$

So $EWT(1,2) \leq EWT(2,1)$ when $\lambda_1 w_1 \geq \lambda_2 w_2$ and $D_1 \leq_{st} D_2$.

The argument can be extended to the case with n jobs, similar to the last part of the proof of Theorem 2.5. ■

We now examine the optimal policy in the class of dynamic policies. Again, here we limit our analysis to the case with a common, deterministic due date d for all jobs. With a fixed due date d , we can employ an idea of Pinedo (2002) to convert the weighted tardiness $w_j T_j$ to a sum of weighted number of tardy jobs. Specifically, the tardiness T_j can be approximated by an infinite series of tardy indicators:

$$T_j \approx \varepsilon \sum_{k=0}^{\infty} I_{\{C_j \geq d+k\varepsilon\}},$$

It follows from Theorem 2.6 that the same WSEPT rule minimizes the tardy penalty $\sum w_j I_{\{C_j \geq d+k\varepsilon\}}$ for each k . Consequently, it also minimizes their sum over k . This together with a continuity argument gives rise to the following theorem.

Theorem 2.10. *When all jobs have a common deterministic due date d , processing the jobs in the non-increasing order of $\lambda_j w_j$ is optimal to minimize $EWT(\zeta)$ in the class of non-preemptive dynamic policies and in the class of preemptive policies.*

Pinedo (1983) shows that the optimality of WSEPT extends to the case of random due dates, under a compatibility condition that requires the distributions of the due dates be nonoverlapping and compatible with the order of WSEPT, in the sense that $\lambda_k w_k \geq \lambda_l w_l$ implies $\Pr(D_k \leq D_l) = 1$.

2.4 General Regular Costs

We now consider more general regular cost functions. We limit our consideration to non-preemptive static policies and problems with a single machine. The cost functions under consideration in this section are, nevertheless, stochastic processes, which are to be elaborated below. We will see that, with a unified treatment of such general cost functions, numerous results established in the literature, including some of those presented in the previous sections, are covered as special cases. The exposition in this Section is mainly based on Zhou and Cai (1997).

Again, note that the completion time of job i under a sequence (static policy) π can be expressed as

$$C_i = C_i(\pi) = \sum_{k \in B_i(\pi)} P_k, \quad (2.15)$$

where $B_i(\pi)$ denotes the set of jobs scheduled no later than job i under sequence π . Let $f_i(C_i)$ denote the cost of processing job i , where $f_i(\cdot)$ is a general regular (stochastic) cost function under the following assumptions:

- (i) $\{f_i(t), t \geq 0\}$ is a stochastic process independent of processing times $\{P_i\}$;
- (ii) $\{f_i(t), t \geq 0\}$ is nondecreasing in $t \geq 0$ almost surely; and
- (iii) $m_i(t) = E[f_i(t)]$ exists and is finite for every $t \geq 0$.

This cost function $f_i(\cdot)$ is general enough to cover most regular costs, deterministic or stochastic, that have been studied in the literature. Examples include:

- *Weighted flowtime*: $f_i(t) = w_i t \Rightarrow f_i(C_i) = w_i C_i$;
- *Tardiness*: $f_i(t) = \max\{0, t - D_i\} \Rightarrow f_i(C_i) = \max\{0, C_i - D_i\}$, where D_i is the due date of job i ;
- *Weighted number of tardy jobs*: $f_i(t) = w_i I_{\{t > D_i\}} \Rightarrow f_i(C_i) = w_i I_{\{C_i > D_i\}}$.

We will address the following two types of performance measures with general regular costs:

- (i) *Total Expected Cost*:

$$TEC(\pi) = \sum_{i=1}^n E[f_i(C_i)]; \quad (2.16)$$

- (ii) *Maximum Expected Cost*:

$$MEC(\pi) = \max_{1 \leq i \leq n} \{E[f_i(C_i)]\}. \quad (2.17)$$

2.4.1 Total Expected Cost

We first state some properties regarding stochastic order in a lemma, whose proof can be found in, e.g., Zhou and Cai (1997).

Lemma 2.5.

- (i) If $X \leq_{\text{st}} Y$ and U is independent of (X, Y) , then $X + U \leq_{\text{st}} Y + U$.
- (ii) If $X \leq_{\text{st}} Y$ and $f(t)$ is nondecreasing in $t \geq 0$ a.s., then $f(X) \leq_{\text{st}} f(Y)$.
- (iii) If $X \leq Y$ a.s. and $E[f(t)]$ is nondecreasing in $t \geq 0$, then $E[f(X)] \leq E[f(Y)]$.

The main result for the *TEC* problem is as follows.

Theorem 2.11.

- (i) Let $\pi = (\dots, i, j, \dots)$ and $\pi' = (\dots, j, i, \dots)$ be two job sequences with identical order except that two consecutive jobs i and j are interchanged. If $P_i \leq_{\text{st}} P_j$ and $(m_i - m_j)(t) = m_i(t) - m_j(t)$ is nondecreasing in $t \geq 0$, where $m_i(t) = E[f_i(t)]$, then $TEC(\pi) \leq TEC(\pi')$.
- (ii) If the jobs can be arranged such that $P_1 \leq_{\text{st}} P_2 \leq_{\text{st}} \dots \leq_{\text{st}} P_n$, and $(m_i - m_j)(t)$ is nondecreasing in t for any $i < j$, then the sequence $\pi^* = \{1, 2, \dots, n\}$ minimizes $TEC(\pi)$. In other words, a sequence in nondecreasing stochastic order of the processing times is optimal.

Proof. Clearly, Part (ii) of the theorem follows immediately from Part (i). Hence it suffices to prove Part (i) only.

By (2.15), it is easy to see that $B_k(\pi) = B_k(\pi')$ for $k \neq i, j$, hence

$$C_k(\pi) = C_k(\pi') \quad \text{if } k \neq i, j. \quad (2.18)$$

Moreover, let C denote the completion time of the job sequenced just before job i under π (which is the same job sequenced just before job j under π'). Then

$$C_i(\pi') = C + P_j + P_i = C + P_i + P_j = C_j(\pi). \quad (2.19)$$

It follows that

$$\begin{aligned} TEC(\pi) - TEC(\pi') &= E[f_i(C_i(\pi))] + E[f_j(C_j(\pi))] - E[f_j(C_j(\pi'))] - E[f_i(C_i(\pi'))] \\ &= E[f_i(C_i(\pi)) - f_i(C_j(\pi'))] + E[f_i(C_j(\pi')) - f_j(C_j(\pi'))] \\ &\quad - E[f_i(C_i(\pi')) - f_j(C_i(\pi'))] \\ &= E[f_i(C_i(\pi))] - E[f_i(C_j(\pi'))] \\ &\quad + E[(f_i - f_j)(C_j(\pi'))] - E[(f_i - f_j)(C_i(\pi'))]. \end{aligned} \quad (2.20)$$

By the independence between jobs, C is independent of P_i and P_j . Thus by Part (i) of Lemma 2.5,

$$P_i \leq_{\text{st}} P_j \implies C_i(\pi) = C + P_i \leq_{\text{st}} C + P_j = C_j(\pi').$$

It then follows from Part (ii) of Lemma 2.5 that $f_i(C_i(\pi)) \leq_{\text{st}} f_i(C_j(\pi'))$, which implies

$$\mathbb{E}[f_i(C_i(\pi))] \leq \mathbb{E}[f_i(C_j(\pi'))]. \quad (2.21)$$

Furthermore, since

$$C_j(\pi') = C + P_j \leq C + P_j + P_i = C_i(\pi') \quad \text{a.s.}$$

and $\mathbb{E}[(f_i - f_j)(t)] = (m_i - m_j)(t)$ is a nondecreasing function of t by the assumption of the theorem, Part (iii) of Lemma 2.5 implies

$$\mathbb{E}[(f_i - f_j)(C_j(\pi'))] \leq \mathbb{E}[(f_i - f_j)(C_i(\pi'))]. \quad (2.22)$$

Combining (2.21) and (2.22), we get $TEC(\pi) - TEC(\pi') \leq 0$ from (2.20), which proves Part (i) of the theorem. Part (ii) then follows. ■

Remark 2.1. A key assumption in Theorem 2.11 is that the processing times $\{P_i\}$ have a stochastic order. Such an order often exists and reduces to the order of the means when $\{P_i\}$ follow a certain family of distributions. Examples include:

1. *Exponential distributions:* If P_i are exponentially distributed, then $P_i \leq_{\text{st}} P_j$ if and only if $\mathbb{E}[P_i] \leq \mathbb{E}[P_j]$.
2. *Normal distributions:* If P_i are normally distributed with a common variance, then $P_i \leq_{\text{st}} P_j$ if and only if $\mathbb{E}[P_i] \leq \mathbb{E}[P_j]$.
3. *Uniform distributions:* If P_i are uniformly distributed over intervals $[0, b_i]$, then $P_i \leq_{\text{st}} P_j \iff b_i \leq b_j \iff \mathbb{E}[P_i] \leq \mathbb{E}[P_j]$.
4. *Gamma distributions:* If P_i are gamma distributed with a common shape parameter, then $P_i \leq_{\text{st}} P_j$ if and only if $\mathbb{E}[P_i] \leq \mathbb{E}[P_j]$.
5. *Poisson distributions:* If P_i are Poisson distributed, then $P_i \leq_{\text{st}} P_j$ if and only if $\mathbb{E}[P_i] \leq \mathbb{E}[P_j]$.

The next theorem addresses a class of problems involving ‘due dates’.

Theorem 2.12. Let $f_i(t) = w_i g(t - D_i) I_{\{t > D_i\}}$, $i = 1, \dots, n$, where $D_1, D_2, \dots, D_n$ are nonnegative random variables (due dates) following arbitrary distributions, w_i is a deterministic weight associated with job i , and $g(\cdot)$ is a strictly increasing, convex and absolutely continuous function defined on $[0, \infty)$ with $g(0) = 0$. Then,

- (i) If $D_i \leq_{\text{st}} D_j$ and $w_i \geq w_j$, then $(m_i - m_j)(t)$ is nondecreasing in $t \geq 0$.
- (ii) If $P_1 \leq_{\text{st}} \dots \leq_{\text{st}} P_n$, $D_1 \leq_{\text{st}} \dots \leq_{\text{st}} D_n$ and $w_1 \geq \dots \geq w_n$, then $\pi^* = \{1, 2, \dots, n\}$ minimizes $TEC(\pi)$.

Proof. By the assumptions, $g(x) \geq 0$ on $[0, \infty)$, $g^{-1}(x)$ exists on $[0, g(\infty))$, and the derivative $g'(x)$ of $g(x)$ exists almost everywhere in any closed subinterval of $[0, \infty)$ and is nondecreasing on its domain. Hence

$$\begin{aligned}
 m_i(t) &= E[f_i(t)] = \int_0^\infty \Pr(f_i(t) \geq x) dx \\
 &= \int_0^{w_i g(t)} \Pr(w_i g(t - D_i) \geq x, t > D_i) dx \\
 &= \int_0^{w_i g(t)} \Pr(D_i \leq t - g^{-1}(x/w_i)) dx.
 \end{aligned} \tag{2.23}$$

Let $y = t - g^{-1}(x/w_i)$, so that $x = w_i g(t - y)$, $dx = -w_i g'(t - y) dy$, $x = 0 \Rightarrow y = t$ and $x = w_i g(t) \Rightarrow y = t - t = 0$. Then by (2.23),

$$m_i(t) = \int_0^t \Pr(D_i \leq y) w_i g'(t - y) dy.$$

It follows that

$$\begin{aligned}
 m_i(t) - m_j(t) &= \int_0^t w_i g'(t - y) \Pr(D_i \leq y) dy - \int_0^t w_j g'(t - y) \Pr(D_j \leq y) dy \\
 &= \int_0^t w_i g'(t - y) [\Pr(D_i \leq y) - \Pr(D_j \leq y)] dy \\
 &\quad + (w_i - w_j) \int_0^t g'(t - y) \Pr(D_j \leq y) dy.
 \end{aligned} \tag{2.24}$$

For any $i < j$, the assumptions of the theorem imply $w_i \geq w_j$ and $D_i \leq_{\text{st}} D_j$ so that $\Pr(D_i \leq y) \geq \Pr(D_j \leq y)$ for all $y \geq 0$. Note also that $g'(t - y)$ is nondecreasing in t . Thus (2.24) shows that $m_i(t) - m_j(t)$ is nondecreasing in t . This proves Part (i) of the theorem, and Part (ii) follows from Part (i) together with Theorem 2.11. ■

The next theorem gives the optimal solutions when the jobs have a common cost function associated with or without job-dependent weights.

Theorem 2.13. Let $\{f(t), t \geq 0\}$ be a stochastic process which is nondecreasing in t almost surely, and suppose that a stochastic order exists between the processing times $P_1, \dots, P_n$.

- (i) A sequence in nondecreasing stochastic order of $\{P_i\}$ minimizes

$$TEC(\pi) = \sum_{i=1}^n E[f(C_i)]. \tag{2.25}$$

(ii) Let w_i denote the weights associated with job i , $i = 1, \dots, n$. If w_i are ‘agreeable’ with the processing times in the sense that the jobs can be arranged such that $P_1 \leq_{\text{st}} \dots \leq_{\text{st}} P_n$ and $w_1 \geq \dots \geq w_n$, then a sequence in nondecreasing stochastic order of $\{P_i\}$ minimizes

$$TEC(\pi) = \sum_{i=1}^n w_i E[f(C_i)]. \quad (2.26)$$

Proof. Part (i) is obviously a special case of Part (ii). To prove (ii), it suffices to verify that $m_i(t) - m_j(t)$ is nondecreasing in t if $P_i <_{\text{st}} P_j$ according to Theorem 2.11. Since $f(t)$ is nondecreasing a.s., $E[f(t)]$ is nondecreasing in t . Moreover, under the agreeability assumption, $P_i <_{\text{st}} P_j$ implies $w_i \geq w_j$, which in turn implies that

$$m_i(t) - m_j(t) = E[w_i f(t)] - E[w_j f(t)] = (w_i - w_j)E[f(t)]$$

is nondecreasing in t . ■

Special Cases of Total Expected Cost

Because of the generality of probabilistic distributions of the processing times and the cost functions, Theorems 2.11–2.13 can cover many commonly studied stochastic problems with specific performance measures (usually involving linear or squared form of cost functions). We list a number of special cases below, which extend the previous results on these cases to more general situations.

Case 1. *Expected Total Weighted Tardiness:*

$$EWT(\pi) = E \left[\sum_{i: C_i > D_i} w_i (C_i - D_i) \right] = \sum_{i=1}^n w_i E[\max\{0, C_i - D_i\}]. \quad (2.27)$$

When all variables are deterministic, this is an unary NP-hard problem and thus an analytical optimal is unlikely obtainable (cf. Lawler et al. 1982). When $\{P_i\}$ are exponentially distributed with rates $\{\tau_i\}$, Pinedo (1983) shows that a sequence in nonincreasing order of $\{\tau_i w_i\}$ minimizes $EWT(\pi)$ under an agreeable condition that $\tau_i w_i \geq \tau_j w_j \Rightarrow D_i \leq_{\text{st}} D_j$. We now generalize this result to the general case with random processing times and due dates following arbitrary distributions.

Take $f_i(t) = w_i(t - D_i)I_{\{t > D_i\}}$, $i = 1, \dots, n$. Then our $TEC(\pi)$ equals $EWT(\pi)$. Thus under the stochastic agreeable condition: $P_1 \leq_{\text{st}} \dots \leq_{\text{st}} P_n$, $D_1 \leq_{\text{st}} \dots \leq_{\text{st}} D_n$ and $w_1 \geq \dots \geq w_n$, Part (ii) of Theorem 2.12 with $g(x) = x$ shows that a sequence in nondecreasing stochastic order of $\{P_i\}$ minimizes $EWT(\pi)$.

Case 2. *Expected Weighted Number of Tardy Jobs*:

$$EWNT(\pi) = E \left[\sum_{i: C_i > D_i} w_i \right] = \sum_{i=1}^n w_i \Pr(C_i > D_i). \quad (2.28)$$

The deterministic version of this problem is NP-hard even with equal due dates $\{D_i\}$ (Karp 1972). When $\{P_i\}$ are exponentially distributed with rates $\{\tau_i\}$ and $\{D_i\}$ are identically distributed, Pinedo (1983) shows that a sequence in nonincreasing order of $\{\tau_i w_i\}$ minimizes $EWNT(\pi)$. Boxma and Forst (1986) give several results on the optimal sequences for cases under conditions such as i.i.d. or constant due dates, i.i.d. or exponential processing times; etc. These studies reveal that the EWNT problem is difficult and certain conditions are always needed to obtain an analytic solution. We now provide a result in the general case with random processing times and due dates.

Assume that $\{D_i\}$ follow a common distribution as D . Take $f_i(t) = w_i I_{\{t > D_i\}}$. Then $TEC(\pi) = EWNT(\pi)$. By Part (ii) of Theorem 2.13, a sequence in nondecreasing stochastic order of $\{P_i\}$ minimizes $EWNT(\pi)$ under the agreeable condition as specified in Theorem 2.13.

Case 3. *Weighted Lateness Probability*:

$$WLP(\pi) = \sum_{i=1}^n w_i \Pr(L_i > 0), \quad (2.29)$$

where $L_i = C_i - D_i$ is the *lateness* of job i . Sarin et al. (1991) and Erel and Sarin (1989) investigated the problem with normally distributed $\{P_i\}$ and a common deterministic due date D . Note that $\Pr(L_i > 0) = \Pr(C_i > D_i)$. Hence the same result in Case 2 above is valid as well for $WLP(\pi)$ with general distributions of $\{P_i\}$.

Case 4. *Expected Total Weighted Squared Flowtime*: Note that all cases discussed above involve linear cost functions. Theorems 2.11–2.13 can be applied to problems with nonlinear cost functions. As an example, we consider the problem to minimize the Expected Total Weighted Squared Flowtime (EWSFT):

$$EWSFT(\pi) = E \left[\sum_{i=1}^n w_i C_i^2 \right] = \sum_{i=1}^n w_i E[C_i^2]. \quad (2.30)$$

This problem is much more difficult than the EWFT problem considered earlier. Townsend (1978) and Bagga and Kalra (1981) proposed branch-and-bound methods to solve the problem when all parameters are deterministic. Furthermore, Bagga and Kalra (1981) show that in a deterministic environment, a sequence in nondecreasing order of $\{P_i\}$ minimizes EWSFT under an agreeable condition that $P_i < P_j$ implies

$w_i \geq w_j$. We now generalize the result to the stochastic version with general random processing times. Take $f(t) = t^2$. Then

$$EWSFT(\pi) = \sum_{i=1}^n w_i E[C_i^2] = \sum_{i=1}^n w_i E[f(C_i)].$$

Since $f(t)$ is an increasing deterministic function, it clearly satisfies the condition of Theorem 2.13. Hence under the agreeable condition as specified in Theorem 2.13, a sequence in nondecreasing stochastic order of $\{P_i\}$ minimizes $EWSFT(\pi)$.

2.4.2 Maximum Expected Cost

We first define an inequality relation between two functions $f(x)$ and $g(x)$ on $[0, \infty)$ in the usual sense: $f \leq g$ if and only if $f(x) \leq g(x)$ for all $x \geq 0$. Then we have the following result for $MEC(\pi)$ defined in (2.17) under general cost functions:

Theorem 2.14. *If the jobs can be arranged such that $m_1 \geq m_2 \geq \dots \geq m_n$, then the sequence $\pi^* = \{1, 2, \dots, n\}$ minimizes $MEC(\pi)$. In other words, if an inequality relation exists between the mean cost functions, then a sequence in nonincreasing order of $\{m_i\}$ is optimal.*

Proof. Let $\pi = (\dots, i, j, \dots)$ and $\pi' = (\dots, j, i, \dots)$. It suffices to show that $m_i \geq m_j$ implies $MEC(\pi) \leq MEC(\pi')$. By (2.18),

$$E[f_k(C_k(\pi))] = E[f_k(C_k(\pi'))] \leq \max_{1 \leq i \leq n} E[f_i(C_i(\pi'))] = MEC(\pi'), \quad k \neq i, j. \quad (2.31)$$

Moreover, as $C_i(\pi) \leq C_j(\pi)$ a.s. and $m_i(t) = E[f_i(t)]$ is nondecreasing in t , Part (ii) of Lemma 2.5 together with (2.19) give

$$E[f_i(C_i(\pi))] \leq E[f_i(C_j(\pi))] = E[f_i(C_i(\pi'))] \leq MEC(\pi'). \quad (2.32)$$

If $m_i \geq m_j$, then

$$E[f_j(C_j(\pi)) | C_j(\pi) = x] = E[f_j(x)] = m_j(x) \leq m_i(x) = E[f_i(C_j(\pi)) | C_j(\pi) = x],$$

for any $x \geq 0$, which implies that

$$E[f_j(C_j(\pi))] \leq E[f_i(C_j(\pi))] = E[f_i(C_i(\pi'))] \leq MEC(\pi') \quad (2.33)$$

Combining (2.31) through (2.33) we get

$$MEC(\pi) = \max_{1 \leq i \leq n} E[f_i(C_i(\pi))] \leq MEC(\pi').$$

This completes the proof. ■

Consider the cost functions of the form

$$f_i(t) = w_i g(t - D_i) I_{\{t > D_i\}}, \quad i = 1, \dots, n,$$

with a nondecreasing function $g(\cdot)$ on $[0, \infty)$. Then Theorem 2.14 can be applied to show the following result:

Theorem 2.15. *If the jobs can be arranged such that $D_1 \leq_{\text{st}} D_2 \leq_{\text{st}} \dots \leq_{\text{st}} D_n$ and $w_1 \geq w_2 \geq \dots \geq w_n$, then a sequence in nondecreasing order of $\{D_i\}$ is optimal in minimizing*

$$MEC(\pi) = \max_{1 \leq i \leq n} w_i E[g(C_i - D_i) I_{\{C_i > D_i\}}]. \quad (2.34)$$

Proof. Let $D_i \leq_{\text{st}} D_j$ and $w_i \geq w_j$. Similar to (2.24) we can show that

$$\begin{aligned} m_i(t) - m_j(t) &= \int_0^t w_i [\Pr(D_i \leq y) - \Pr(D_j \leq y)] dg_t(y) \\ &\quad + (w_i - w_j) \int_0^t \Pr(D_j \leq y) dg_t(y), \end{aligned} \quad (2.35)$$

where $g_t(y) = -g(t - y)$. As g is a nondecreasing function, $g_t(y)$ is nondecreasing in y . This together with the facts that $\Pr(D_i \leq y) \geq \Pr(D_j \leq y)$ for all $y \leq 0$ (as $D_i \leq_{\text{st}} D_j$) and $w_i \geq w_j$ show that the right hand side of (2.35) is nonnegative. Hence $m_i(t) \geq m_j(t)$. Theorem 2.14 then applies to complete the proof. ■

Special Cases of Maximum Expected Cost

We now show the applications of the general results on MEC obtained above to some special cases, which will extend the previously known results on MEC to more general situations.

Case 5. *Maximum Expected Lateness:*

$$MEL(\pi) = \max_{1 \leq i \leq n} E[C_i - D_i].$$

In deterministic environment, Jackson (1955) provided an elegant result that $MEL(\pi)$ is minimized by a sequence in nondecreasing order of $\{D_i\}$, which is referred to as the *Earliest Due Date (EDD)* rule. We now extend this result to stochastic situation with random processing times and due dates.

Take $f_i(t) = t - D_i$ for $i = 1, \dots, n$. Then $m_i(t) = t - E[D_i]$, so that $m_i(t) \geq m_j(t)$ if and only if $E[D_i] \leq E[D_j]$. It then follows from Theorem 2.14 that a sequence in nondecreasing order of $\{E[D_i]\}$ minimizes $MEL(\pi)$. In other words, the optimal sequence is the *Earliest Expected Due Date (EEDD)* rule.

A more general result is on the problem where each job i is assigned with a weight w_i . In this case, the objective is to minimize the maximum expected weighted lateness. According to Theorem 2.14, it is not hard to show that the EEDD rule is optimal under the agreeable condition that $E[D_j] \leq E[D_i]$ if and only if $w_i \geq w_j$.

It is interesting to note that Theorem 4 of Crabill and Maxwell (1969) is similar to the above result.

Case 6. *Maximum Expected Weighted Tardiness:*

$$MEWT(\pi) = \max_{1 \leq i \leq n} w_i E[(C_i - D_i) I_{\{C_i > D_i\}}].$$

It is also a well-known result that an optimal sequence for the deterministic version of this problem, when all weights are equal, should schedule the jobs in EDD order (cf. Jackson, 1955). We now generalize this result to the stochastic case.

Take $f_i(t) = w_i(t - D_i) I_{\{t > D_i\}}$ for $i = 1, \dots, n$. Then $MEWT(\pi) = MEC(\pi)$ in (2.34) with $g(t) = t$. Thus by Theorem 2.15, if $D_i \leq_{st} D_j \iff w_i \geq w_j$, then the EEDD rule is optimal in minimizing $MEWT(\pi)$.

Case 7. *Maximum Weighted Probability of Lateness:*

$$MWPL(\pi) = \max_{1 \leq i \leq n} w_i \Pr(C_i \geq D_i).$$

When $\{P_i\}$ are random variables, $\{D_i\}$ are deterministic, and weights are equal, Banerjee (1965) shows that the EDD rule minimizes $MWPL$. Crabill and Maxwell (1969) extend this result to random $\{D_i\}$. We now consider the case with random $\{P_i\}$ and $\{D_i\}$ as a special case of MEC.

Since $\Pr(C_i \geq D_i) = E[I_{\{C_i \geq D_i\}}]$, $MWPL(\pi) = \max_{1 \leq i \leq n} w_i E[I_{\{C_i \geq D_i\}}]$ is a special case of (2.34) with $g(t) \equiv 1$. Thus by Theorem 2.15, if $D_i \leq_{st} D_j \iff w_i \geq w_j$, then the EEDD rule minimizes $MWPL(\pi)$.

2.5 Exponential Processing Times

When the processing times are exponentially distributed, we can obtain more results on optimal sequences to minimize the total expected cost $TEC(\pi)$ in (2.16).

In this section, we assume that the processing times $P_1, \dots, P_n$ follow exponential distributions with rates $\lambda_1, \dots, \lambda_n$ respectively. The density and cumulative distribution functions of P_i are given by $\lambda_i e^{-\lambda_i x}$ and $\Pr(P_i \leq x) = 1 - e^{-\lambda_i x}$, respectively, for $i = 1, \dots, n$.

The cost to complete job i is $f_i(C_i)$ with the cost function $f_i(\cdot)$ as described in Sect. 2.4. That is, $f_i(t)$ is a nondecreasing random function of t a.s., independent of

processing times $\{P_i\}$, with mean function $E[f_i(t)] = m_i(t)$. In particular, we consider the case $f_i(t) = w_i g(t - D_i) I_{\{t > D_i\}}$, where w_i is the weight of job i , $g(\cdot)$ is a nonnegative and nondecreasing function on $[0, \infty)$, and D_i is the due date of job i .

Stochastic scheduling problems with exponentially distributed processing times have been studied by many authors, which have produced some elegant results. 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 following 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 under a performance measure that covers the expected sum of weighted completion times, expected makespan, and expected lifetime of a series system, and showed that a SEPT or LEPT sequence minimizes this performance measure.

Pinedo (1983) examined the minimizations of the expected weighted sum of completion times with random arrival times, the expected weighted sum of tardinesses, and the expected weighted number of tardy jobs. He showed that the WSEPT sequences are optimal under certain (compatability) conditions.

Boxma and Forst (1986) investigated the minimization of the expected weighted number of tardy jobs and derived optimal sequences for various processing time and due date distributions, including exponential and independently and identically distributed (i.i.d.) processing times and/or due dates.

Kämpke (1989) generalized the work of Weiss and Pinedo (1980) and derived sufficient conditions for optimal priority policies beyond SEPT and LEPT.

In Pinedo (2002), the WSEPT sequence was shown to minimize the performance measure $E[\sum w_i h(C_i)]$, where $h(\cdot)$ is a general function. Moreover, the performance measure $E[\sum w_i h_i(C_i)]$ was also studied with a job-dependent cost function $h_i(\cdot)$. Pinedo defined an order $h_j \geq_s h_k$ (termed as h_j is *steeper* than h_k) between the cost functions by $dh_j(t) \geq dh_k(t)$ for all $t \geq 0$ if the differentials exist; or $h_j(t + \delta) - h_j(t) \geq h_k(t + \delta) - h_k(t)$ for all $t \geq 0$ and $\delta > 0$ otherwise. It was shown that the WSEPT sequence minimizes $E[\sum w_i h_i(C_i)]$ under the agreeability condition $\lambda_j w_j \geq \lambda_i w_i \iff h_j \geq_s h_k$.

In this section, we present three more general results:

1. A sequence in the order based on the increments of $\lambda_j E[f_j(t)]$ is optimal to minimize $E[\sum f_i(C_i)]$.
2. When the due dates $\{D_i\}$ have a common distribution, the WSEPT sequence is optimal to minimize $E[\sum w_i g(C_i - D_i) I_{\{C_i > D_i\}}]$ without requiring any additional conditions.

3. When $\{D_i\}$ have different distributions, if $g(\cdot)$ is convex on $[0, \infty)$ with $g(0) = 0$, then a sequence in the nonincreasing order of $\{\lambda_i w_i \Pr(P_i \leq x)\}$ is optimal to minimize $E[\sum w_i g(C_i - D_i) I_{\{C_i > D_i\}}]$. In particular, if $\lambda_i w_i \geq \lambda_j w_j \Rightarrow D_i \leq_{rmst} D_j$, then the WSEPT sequence is optimal to minimize $E[\sum w_i g(C_i - D_i) I_{\{C_i > D_i\}}]$.

This section is mainly based on Cai and Zhou (2005).

2.5.1 Optimal Sequence for General Costs

The optimal sequence to minimize $E[\sum f_i(C_i)]$ is stated in the following theorem.

Theorem 2.16. *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 in the sense that*

$$\lambda_i [m_i(t) - m_i(s)] \leq \lambda_j [m_j(t) - m_j(s)] \quad \forall t > s, \quad (2.36)$$

or equivalently,

$$\int_0^\infty \phi(s) \lambda_i dm_i(s) \leq \int_0^\infty \phi(s) \lambda_j dm_j(s) \quad (2.37)$$

for any nonnegative measurable function $\phi(s)$ on $[0, \infty)$, where the integrals are in Lebesgue-Stieltjes sense, then the sequence $(1, 2, \dots, n)$ is optimal to minimize $E[\sum f_i(C_i)]$. In other words, a sequence in nonincreasing order of the increments of $\{\lambda_i m_i(t)\}$ is optimal to minimize $E[\sum f_i(C_i)]$.

Proof. First, by taking $\phi(s) = I_{[s, t]}$ in (2.37) we can see that (2.37) implies (2.36). Conversely, for any nonnegative measurable function $\phi(s)$, we can construct functions $\phi_1(s) \leq \phi_2(s) \leq \dots$, with each $\phi_k(s)$ a linear combination of functions of form $I_{[s, t]}$, such that $\phi_k(s) \rightarrow \phi(s)$ as $k \rightarrow \infty$. Hence an application of the monotone convergence theorem shows that (2.36) implies (2.37). This establishes the equivalence between (2.36) and (2.37).

Next, since $\{f_i(t)\}$ are independent of $\{P_i\}$,

$$\begin{aligned} E[f_i(t + P_j)] &= E\{E[f_i(t + P_j) | P_j]\} = \int_0^\infty E[f_i(t + x) | P_j = x] \lambda_j e^{-\lambda_j x} dx \\ &= \int_0^\infty E[f_i(t + x)] \lambda_j e^{-\lambda_j x} dx = \int_0^\infty m_i(t + x) \lambda_j e^{-\lambda_j x} dx \end{aligned} \quad (2.38)$$

for $i, j = 1, 2, \dots, n$ and $t \geq 0$. Furthermore, by convolution it can be shown that

$$\text{the density of } P_i + P_j = \begin{cases} \frac{\lambda_i \lambda_j}{\lambda_j - \lambda_i} (e^{-\lambda_i x} - e^{-\lambda_j x}) & \text{if } \lambda_i \neq \lambda_j \\ \lambda_i^2 x e^{-\lambda_i x} & \text{if } \lambda_i = \lambda_j. \end{cases} \quad (2.39)$$

(Note that the second part of (2.39) is equal to the limit of the first part as λ_j converges to λ_i .) Thus, when $\lambda_i \neq \lambda_j$, by (2.39) together with an argument similar to (2.38) we obtain

$$E[f_i(t + P_i + P_j)] = \frac{\lambda_i \lambda_j}{\lambda_j - \lambda_i} \int_0^\infty m_i(t + x) \left(e^{-\lambda_i x} - e^{-\lambda_j x} \right) dx. \quad (2.40)$$

Let $\pi = \{\dots, i, j, \dots\}$ be an arbitrary job sequence, $\pi' = \{\dots, j, i, \dots\}$ be the sequence by interchanging two consecutive jobs i, j in π , and C denote the completion time of the job prior to job i under π . Then, for $TEC(\pi) = E[\sum f_i(C_i(\pi))]$,

$$\begin{aligned} TEC(\pi) - TEC(\pi') &= E[f_i(C + P_i)] + E[f_j(C + P_i + P_j)] \\ &\quad - E[f_j(C + P_j)] - E[f_i(C + P_i + P_j)]. \end{aligned} \quad (2.41)$$

Since $P_1, \dots, P_n$ are mutually independent, conditional on $C = t$ we have

$$E[f_i(C + P_i)|C = t] = E[f_i(t + P_i)|C = t] = E[f_i(t + P_i)]$$

and similarly, $E[f_i(C + P_i + P_j)|C = t] = E[f_i(t + P_i + P_j)]$. Hence a combination of (2.41) with (2.38) and (2.40) yields that, conditional on $C = t$,

$$\begin{aligned} TEC(\pi) - TEC(\pi') &= E[f_i(t + P_i)] + E[f_j(t + P_i + P_j)] - E[f_j(t + P_j)] - E[f_i(t + P_i + P_j)] \\ &= \int_0^\infty m_i(t + x) \left\{ \lambda_i e^{-\lambda_i x} - \frac{\lambda_i \lambda_j}{\lambda_j - \lambda_i} \left(e^{-\lambda_i x} - e^{-\lambda_j x} \right) \right\} dx \\ &\quad - \int_0^\infty m_j(t + x) \left\{ \lambda_j e^{-\lambda_j x} - \frac{\lambda_i \lambda_j}{\lambda_j - \lambda_i} \left(e^{-\lambda_i x} - e^{-\lambda_j x} \right) \right\} dx \\ &= \int_0^\infty [\lambda_i m_i(t + x) - \lambda_j m_j(t + x)] \frac{\lambda_j e^{-\lambda_j x} - \lambda_i e^{-\lambda_i x}}{\lambda_j - \lambda_i} dx \\ &= a_{ij}(t), \quad \text{say.} \end{aligned} \quad (2.42)$$

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 nondecreasing function on $(-\infty, \infty)$. Hence we can write

$$m_i(t + x) = \int_{-\infty}^{t+x} dm_i(s), \quad i = 1, \dots, n.$$

An application of Fubini's Theorem then gives

$$a_{ij}(t) = \int_0^\infty \int_{-\infty}^{t+x} [\lambda_j dm_j(s) - \lambda_i dm_i(s)] \frac{\lambda_i e^{-\lambda_i x} - \lambda_j e^{-\lambda_j x}}{\lambda_j - \lambda_i} dx$$

$$\begin{aligned}
&= \int_{-\infty}^t \int_0^{\infty} \frac{\lambda_i e^{-\lambda_i x} - \lambda_j e^{-\lambda_j x}}{\lambda_j - \lambda_i} dx [\lambda_j dm_j(s) - \lambda_i dm_i(s)] \\
&\quad + \int_t^{\infty} \int_{s-t}^{\infty} \frac{\lambda_i e^{-\lambda_i x} - \lambda_j e^{-\lambda_j x}}{\lambda_j - \lambda_i} dx [\lambda_j dm_j(s) - \lambda_i dm_i(s)] \\
&= \int_t^{\infty} \frac{e^{-\lambda_i(s-t)} - e^{-\lambda_j(s-t)}}{\lambda_j - \lambda_i} [\lambda_j dm_j(s) - \lambda_i dm_i(s)]. \tag{2.43}
\end{aligned}$$

It is easy to see that

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

Hence by (2.42) and (2.43) together with condition (2.37), conditional on $C = t$,

$$i > j \implies TEC(\pi) - TEC(\pi') = a_{ij}(t) \geq 0 \quad \forall t \geq 0,$$

which in turn implies, unconditionally, $TEC(\pi) - TEC(\pi') \geq 0$.

Thus we have shown that $TEC(\pi) \geq TEC(\pi')$ 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 (2.39), or considering the limit as λ_j converges to λ_i . It follows that the sequence π' is better than π if $i > j$. Consequently the sequence $(1, 2, \dots, n)$ is optimal. ■

Note that condition (2.37) is what we need to prove Theorem 2.16, while condition (2.36) is usually easier to check in specific cases. Also, (2.36) does not require $m_i(t)$ to be differentiable at all $t \geq 0$. If $m_i(t)$ may be discontinuous at some points, then (2.36) assumes that

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

at any discontinuity t (which can also be written as $\lambda_i dm_i(t) \leq \lambda_j dm_j(t)$ in that sense). If $m_i(t)$ have different left and right derivatives at some t , then (2.36) requires $\lambda_i dm_i(t+) \leq \lambda_j dm_j(t+)$ and $\lambda_i dm_i(t-) \leq \lambda_j dm_j(t-)$ for $i > j$.

Remark 2.2. Theorem 2.16 extends the results of Pinedo (2002). Condition (2.36) or (2.37) is in fact equivalent to ‘ $\lambda_j m_j$ is steeper than $\lambda_i m_i$ ’ in Pinedo’s terminology. Hence Theorem 2.16 says that the sequence in a reverse steepness order of $\{\lambda_i m_i(t), i = 1, \dots, n\}$ is optimal. Note that in Pinedo (2002), which 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., $\lambda_i w_i \geq \lambda_j w_j$ implies that $f_i(t)/w_i$ is steeper than $f_j(t)/w_j$. In Theorem 2.16, such an agreeable condition can be replaced by a weaker condition (2.36). In addition, Theorem 2.16 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.

The following example shows an application of Theorem 2.16.

Example 2.3. Let $f_i(t) = w_i h(t)$, where w_i is a deterministic weight and $h(t)$ is a nondecreasing stochastic process. Then $m_i(t) = E[f_i(t)] = w_i E[h(t)]$ is nondecreasing in t . Furthermore, if $\lambda_i w_i > \lambda_j w_j$, then

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

Hence by Theorem 2.16, a sequence in nonincreasing order of $\{\lambda_i w_i\}$ minimizes $E[\sum w_i h(C_i)]$. As $E[P_i] = 1/\lambda_i$, this sequence is the WSEPT and so the result generalizes that of Pinedo (2002) to an arbitrary stochastic cost function h , which allows, for example, a random common due date with an arbitrary distribution.

There are, of course, also examples where the condition of Theorem 2.16 does not hold. A simple one is given below.

Example 2.4. 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 $dm_1(t) = 2dt$ and $dm_2(t) = 2tdt$. It follows that $\lambda_1 dm_1(t) \leq \lambda_2 dm_2(t)$ when $t \geq \lambda_1/\lambda_2$, and $\lambda_1 dm_1(t) > \lambda_2 dm_2(t)$ for $t < \lambda_1/\lambda_2$. Thus (2.36) cannot hold for jobs 1 and 2. Furthermore, suppose $w_1 = w_2$. Then it is not difficult to show that $TEC(1, 2) < TEC(2, 1)$ if and only if $\lambda_1 > \lambda_2^2$. Hence the WSEPT rule is not optimal even if the jobs have a common weight.

2.5.2 Optimal Sequences with Due Dates

The applications of Theorem 2.16 lead to the next two theorems for the case with $f_i(t) = w_i g(t - D_i) I_{\{t > D_i\}}$. The first one is for identically distributed due dates.

Theorem 2.17. *If $\{D_i\}$ have a common distribution, then a sequence in nonincreasing order of $\{\lambda_i w_i\}$, or equivalently, in nondecreasing order of $\{E[P_i]/w_i\}$, minimizes $E[\sum w_i g(C_i - D_i) I_{\{C_i > D_i\}}]$.*

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

$$m_i(t) = E[f_i(t)] = w_i E[g(t - D_i) I_{\{t > D_i\}}] = w_i \int_{0 \leq x < t} g(t - x) dF(x). \quad (2.44)$$

Let

$$\tilde{g}(t) = \int_{0 \leq x < t} g(t - x) dF(x)$$

so that $\lambda_i m_i(t) = \lambda_i w_i \tilde{g}(t)$. Since $g(t)$ is nonnegative and nondecreasing, so is $\tilde{g}(t)$. It follows that $\lambda_i w_i \geq \lambda_j w_j$ implies

$$\lambda_i [m_i(b) - m_i(a)] = \lambda_i w_i [\tilde{g}(b) - \tilde{g}(a)] \geq \lambda_j w_j [\tilde{g}(b) - \tilde{g}(a)] = \lambda_j [m_j(b) - m_j(a)]$$

for all $a < b$. Thus if $\lambda_1 w_1 \geq \dots \geq \lambda_n w_n$, then $\{1, \dots, n\}$ is optimal by Theorem 2.16. In other words, a sequence in nonincreasing order of $\{\lambda_i w_i\}$ is optimal to minimize $E[\sum w_i g(C_i - D_i) I_{\{C_i > D_i\}}]$. ■

Example 2.5.

- (i) Let $g(x) \equiv 1$. Then Theorem 2.17 says that a sequence in nonincreasing 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 2.17, a sequence in nonincreasing 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 2.18. *Let $F_i(x) = \Pr(D_i \leq x)$. If $g(\cdot)$ is convex with $g(0) = 0$ and*

$$\lambda_1 w_1 F_1(x) \geq \lambda_2 w_2 F_2(x) \geq \dots \geq \lambda_n w_n F_n(x) \quad \text{for } x \geq 0, \quad (2.45)$$

then the sequence $(1, 2, \dots, n)$ minimizes $E[\sum w_i g(C_i - D_i) I_{\{C_i > D_i\}}]$. In other words, a sequence in the nonincreasing order of $\{\lambda_j w_j F_j(x)\}$ is optimal.

Proof. Since $g(x)$ is nondecreasing with $g(0) = 0$, we have $g(t - x) = \int_0^{t-x} dg(y)$. By Fubini's Theorem,

$$\begin{aligned} m_i(t) &= w_i E[g(t - D_i) I_{\{t > D_i\}}] = w_i \int_{0 \leq x < t} g(t - x) dF_i(x) \\ &= w_i \int_{0 \leq x < t} \int_0^{t-x} dg(y) dF_i(x) = w_i \int_{0 \leq y < t} \int_{0 \leq x \leq t-y} dF_i(x) dg(y) \\ &= w_i \int_{0 \leq y < t} F_i(t - y) dg(y) = w_i \int_0^t F_i(x) dg_t(x), \end{aligned}$$

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

$$\begin{aligned} \lambda_i [m_i(b) - m_i(a)] &= \lambda_i w_i \left\{ \int_0^b F_i(x) dg_b(x) - \int_0^a F_i(x) dg_a(x) \right\} \\ &= \int_a^b \lambda_i w_i F_i(x) dg_b(x) dx + \int_0^a \lambda_i w_i F_i(x) [dg_b(x) - dg_a(x)]. \end{aligned} \quad (2.46)$$

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

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

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

$$\int_0^a \phi(x) dg_a(x) \leq \int_0^a \phi(x) dg_b(x), \quad \text{or} \quad \int_0^a \phi(x) [dg_b(x) - dg_a(x)] \geq 0,$$

for any nonnegative measurable function $\phi(x)$ on $[0, a]$. If $\lambda_i w_i F_i(x) \leq \lambda_j w_j F_j(x)$ for $x \geq 0$, then

$$\int_0^a \lambda_i w_i F_i(x) [dg_b(x) - dg_a(x)] \leq \int_0^a \lambda_j w_j F_j(x) [dg_b(x) - dg_a(x)]. \quad (2.47)$$

Moreover, as $g_b(x)$ is nondecreasing on $[0, t]$,

$$\int_a^b \lambda_i w_i F_i(x) dg_b(x) \leq \int_a^b \lambda_j w_j F_j(x) dg_b(x). \quad (2.48)$$

Now, if $i > j$, then $\lambda_i w_i F_i(x) \leq \lambda_j w_j F_j(x)$ for $x \geq 0$ by the condition of the theorem. It then follows from (2.46) to (2.48) that

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

Thus by Theorem 2.16, the sequence $(1, 2, \dots, n)$ minimizes $E[\sum w_i g(C_i - D_i) I_{\{C_i > D_i\}}]$.

Finally, if $\lambda_i w_i F_i(x) \geq \lambda_j w_j F_j(x)$ for $x \geq 0$, then $\lambda_i w_i \geq \lambda_j w_j$ as $x \rightarrow \infty$. Consequently, given the existence of an order between $\{\lambda_j w_j F_j(x)\}$, a sequence in the nonincreasing order of $\{\lambda_j w_j\}$, i.e., the WSEPT, is optimal. ■

Corollary 2.3. *If $g(\cdot)$ satisfies the conditions in Theorem 2.18 and $\lambda_i w_i \geq \lambda_j w_j$ implies $D_i \leq_{\text{st}} D_j$, then a sequence in nonincreasing order of $\{\lambda_j w_j\}$ minimizes $E[\sum w_i g(C_i - D_i) I_{\{C_i > D_i\}}]$.*

Proof. By the condition of the corollary and the definition for the stochastic order, $\lambda_i w_i \geq \lambda_j w_j$ implies $F_i(x) \geq F_j(x)$ for all $x \geq 0$. As a result, a nonincreasing order exists between $\{\lambda_j w_j F_j(x)\}$ and is equivalent to the nonincreasing order of $\{\lambda_j w_j\}$, so the corollary follows immediately from Theorem 2.18. ■

Example 2.6. Both $g(x) = x$ and $g(x) = x^2$ satisfies the conditions of Theorem 2.18. Hence if the compatibility condition in the corollary holds, then a sequence in nonincreasing order of $\{\lambda_i w_i\}$, or equivalently, in nondecreasing stochastic order of $\{D_i\}$,

minimizes both the expected weighted sum of tardinesses $E[\sum_{i:C_i > D_i} w_i(C_i - D_i)]$ and the expected weighted sum of squared tardinesses $E[\sum_{i:C_i > D_i} w_i(C_i - D_i)^2]$. (This is not true for the expected weighted number of tardy jobs. Note that $g(x) = 1$ does not satisfy the conditions of Theorem 2.18 because $g(0) \neq 0$.)

Condition (2.45) 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 $D_i \leq_{\text{st}} D_j$ fails, a sequence in nonincreasing order of $\{\lambda_j w_j\}$ could still be optimal. We illustrate this in the next example.

Example 2.7. 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 $\lambda_i w_i \geq \lambda_j w_j$ and $\lambda_i w_i \delta_i \geq \lambda_j w_j \delta_j$. We show that $\lambda_i w_i F_i(x) \geq \lambda_j w_j F_j(x)$ 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 $\lambda_i w_i \delta_i \geq \lambda_j w_j \delta_j$ imply, for $x > 0$,

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

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

Case 2: $\delta_i \geq \delta_j$. Then $F_i(x) \geq F_j(x)$ for $x \geq 0$, which together with $\lambda_i w_i \geq \lambda_j w_j$ leads immediately to $\lambda_i w_i F_i(x) \geq \lambda_j w_j F_j(x)$. Conversely, if $\lambda_i w_i F_i(x) \geq \lambda_j w_j F_j(x)$ for $x \geq 0$, then $\lambda_i w_i \geq \lambda_j w_j$ as $x \rightarrow \infty$. Furthermore,

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

Hence $\lambda_i w_i \delta_i \geq \lambda_j w_j \delta_j$. Thus we have shown that $\lambda_i w_i F_i(x) \geq \lambda_j w_j F_j(x)$ for $x \geq 0$ if and only if $\lambda_i w_i \geq \lambda_j w_j$ and $\lambda_i w_i \delta_i \geq \lambda_j w_j \delta_j$. As a result, even if $\lambda_i w_i > \lambda_j w_j$ but $\delta_i < \delta_j$ (so that $D_i \leq_{\text{st}} D_j$ fails), a sequence in nonincreasing order of $\{\lambda_j w_j\}$ would still be optimal if we have $\lambda_i w_i \delta_i \geq \lambda_j w_j \delta_j$ for such i and j .

2.5.3 Examples of Applications

Two examples of applications are provided 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 2.8. 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 $E[h_i(t)] = u(t)$ is a nonincreasing 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

$$\sum_{i=1}^n [a_i h_i(C_i) - \beta a_i(1 + \alpha)^{C_i}] \quad (2.49)$$

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

$$f_i(t) = \beta a_i(1 + \alpha)^t - a_i h_i(t), \quad i = 1, \dots, n. \quad (2.50)$$

Then the problem of maximizing the total net profit given by (2.49) is equivalent to minimizing $E[\sum f_i(C_i)]$. From (2.50) we can see that the mean function of $f_i(t)$ is

$$\begin{aligned} m_i(t) &= E[f_i(t)] = \beta a_i E[(1 + \alpha)^t] - a_i E[h_i(t)] + a_i(1 - \beta) \\ &= a_i \{ \beta E[(1 + \alpha)^t] - u(t) \}, \end{aligned} \quad (2.51)$$

As $E[h_i(t)] = u(t)$ is a nonincreasing function of t , by (2.51) $m_i(t)$ is nondecreasing in t . Write $G(t) = \beta E[(1 + \alpha)^t] - u(t)$ for brevity, which is nondecreasing in t . Then, assuming that the processing times are exponentially distributed with parameters $\lambda_1, \dots, \lambda_n$, it follows from (2.51) that $\lambda_i a_i \geq \lambda_j a_j$ implies

$$\lambda_i [m_i(t) - m_i(s)] = \lambda_i a_i [G(t) - G(s)] \geq \lambda_j a_j [G(t) - G(s)] = \lambda_j [m_j(t) - m_j(s)]$$

for all $t > s$. Thus by Theorem 2.1, a sequence in nonincreasing order of $\{\lambda_j a_j\}$ minimizes $E[\sum f_i(C_i)]$, 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 2.9. 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 $\lambda_1, \dots, \lambda_n$ 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):

$$ETL(\pi) = E \left[\sum_{i=1}^n \{ v_i \tilde{h}(C_i - D_i) I_{\{D_i < C_i \leq B\}} + v_i I_{\{C_i > B\}} \} \right] \quad (2.52)$$

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

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

Then the objective function in (2.52) is equal to $ETL(\pi) = E[\sum f_i(C_i)]$. As $h(t)$ is decreasing in t almost surely and $0 \leq h(t) \leq 1$, $\{f_i(t), t \geq 0\}$ is a nondecreasing 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

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

where $G(t) = 1 - e^{-\delta t} (1 - E[\tilde{h}(t - D) I_{\{t > D\}}])$. Since $0 \leq E[\tilde{h}(t - D) I_{\{t > D\}}] \leq 1$ and by the assumptions of the problem $E[\tilde{h}(t - D) I_{\{t > D\}}]$ is nondecreasing in t , $e^{-\delta t} (1 - E[\tilde{h}(t - D) I_{\{t > D\}}])$ is nonincreasing in t and so $G(t)$ is nondecreasing in t . Hence, similar to the arguments in Example 2.8, it follows from Theorem 2.16 that a sequence in nonincreasing order of $\{\lambda_j v_j\}$ is optimal. That is, items with higher ratios of value over mean testing time should be tested earlier.

2.6 Compound-Type Distributions

In this section, we consider a more general class of compound-type distributions than the exponential distributions for the processing times, and derive the optimal sequence to minimize the total expected cost in (2.16) and (2.17), which generalize the results presented in Sect. 2.5.

This class of distributions are characterized by a common form of their characteristic functions. More specifically, we consider a class of distributions parameterized by γ , with characteristic functions of the form

$$\phi(t) = E[e^{itX}] = \frac{1}{1 + G(t)/\gamma} \quad (2.53)$$

for random variable X , where i denotes the imaginary unit and $G(t)$ is a complex-valued function of real variable t such that $\phi(t)$ is a characteristic function of a probability distribution. We refer to this class of distributions as *compound-type distributions*. It is easy to see that the exponential distribution is a special case of (2.53) with $G(t) = -it$.

Similar to the exponential distributions, the effects of a job interchange procedure under compound-type distributions can be computed such that the conditions for a sequence to minimize the $TEC(\pi)$ in (2.16) or (2.17) can be readily verified.

Generally speaking, the processing times on jobs, $\{P_i, 1 \leq i \leq n\}$, are nonnegative variables. But in some circumstances, it is convenient to approximate the distribution of a processing time with one that can take negative values, such as a normal distribution. See, for example, Boys et al. (1997) and Jang and Klein (2002) and the references therein. Thus we allow the processing times to be real-valued random variables with positive means.

We will show that the nonincreasing order of the increments of $m_i(t) = \gamma_j E[f_i(t)]$ is optimal to minimize $E[\sum f_i(C_i)]$. Furthermore, if the due dates $\{D_i\}$ have a common distribution, then the optimal policy to minimize $E[\sum w_i g(C_i - D_i) I_{\{C_i > D_i\}}]$ is to schedule the jobs according to non-increasing order of $w_j \gamma_j$. On the other hand, if the due dates D_j have different distributions, then the optimal policy depends on these distributions and relies on the convexity or concavity of $g(\cdot)$. Specifically, when jobs can be ordered by $w_j \gamma_j \Pr(D_j \leq x)$ for all x , such an ordering is optimal when $g(\cdot)$ is convex or concave. The exposition below is based on Cai et al. (2007a).

2.6.1 Classes of Compound-Type Distributions

We first give a lemma on the compound-type distributions defined by (2.53), which will play a key role in the interchange procedures to find the optimal sequences to minimize $TEC(\pi)$.

Suppose that the processing times $P_1, \dots, P_n$, are independent random variables following the compound-type distributions with cumulative distribution function (cdf) $F_i(x)$ for P_i , $i = 1, \dots, n$. Denote by $F_{ij}(x)$ the cdf of $P_i + P_j$ and $\phi_i(t)$ the characteristic function of P_i .

Lemma 2.6. *There exists a function $G(t)$ and a series of numbers $\gamma_1, \dots, \gamma_n$ such that $\phi_i(t) = (1 + G(t)/\gamma_i)^{-1}$ if and only if*

$$\frac{F_i(x) - F_{ij}(x)}{\gamma_i} = \frac{F_j(x) - F_{ij}(x)}{\gamma_j}.$$

In particular, γ_i can take the value $1/E[P_i]$.

Proof. Since the Fourier transformation is linear and the characteristic function of $F_{ij}(x)$ is the product of the characteristic functions of $F_i(x)$ and $F_j(x)$,

$$\begin{aligned} \frac{F_i(x) - F_{ij}(x)}{\gamma_i} &= \frac{F_j(x) - F_{ij}(x)}{\gamma_j} \quad \text{for } 1 \leq i, j \leq n \\ \iff \frac{\phi_i(t) - \phi_i(t)\phi_j(t)}{\gamma_i} &= \frac{\phi_j(t) - \phi_i(t)\phi_j(t)}{\gamma_j} \quad \text{for } 1 \leq i, j \leq n \\ \iff \frac{\phi_i(t)}{\gamma_i(1 - \phi_i(t))} &= \frac{\phi_j(t)}{\gamma_j(1 - \phi_j(t))} = \frac{1}{G(t)} \quad (\text{say}) \quad \text{for } 1 \leq i, j \leq n \\ \iff \phi_i(t) &= \frac{1}{1 + G(t)/\gamma_i} \quad \text{for } 1 \leq i \leq n. \end{aligned} \tag{2.54}$$

Moreover, it is clear that the equivalence in (2.54) still holds if $G(t)$ and γ_i are replaced with $aG(t)$ and $a\gamma_i$, respectively, for any fixed complex value a . As a result, we can take $G'(0) = -i$ without loss of generality. Consequently, since $\phi_i(0) = 1$ implies $G(0) = 0$,

$$iE[P_i] = \phi'_i(0) = -\frac{G'(0)/\gamma_i}{(1 + G(0)/\gamma_i)^2} = \frac{i}{\gamma_i} \implies \gamma_i = \frac{1}{E[P_i]}.$$

The proof is thus complete. ■

In the case of non-negative P_i and P_j , since

$$\frac{F_i(x) - F_{ij}(x)}{F_i(x) - F_{ij}(x)} = \frac{\Pr(P_i \leq x < P_i + P_j)}{\Pr(P_j \leq x < P_i + P_j)} = \frac{\Pr(P_i \leq x | P_i + P_j > x)}{\Pr(P_j \leq x | P_i + P_j > x)},$$

Lemma 2.6 states that this ratio is constant for all x . For real-valued P_i and P_j , however, this implication is not generally true as we no longer necessarily have the equality $F_i(x) - F_{ij}(x) = \Pr(P_i \leq x < P_i + P_j)$.

Some of the most commonly used distributions belong to this class of distributions, as shown in the following examples.

Example 2.10 (Exponential). If P follows the exponential distribution with rate γ , then the characteristic function of P is $\phi(t) = (1 - it/\gamma)^{-1}$ (i.e., $G(t) = -it$).

Example 2.11 (Laplace). If P has a Laplace distribution with scale parameter α with density $(2\alpha)^{-1}e^{-|x|/\alpha}$ for $x \geq 0$, then $\phi(t) = (1 + \alpha^2 t^2)^{-1}$ ($\gamma = 1/\alpha^2$ and $G(t) = t^2$).

Example 2.12 (Pólya-type). Let $\phi(t) = (1 + C|t|^\alpha)^{-1}$ for $-\infty < t < \infty$, with parameters $0 < \alpha \leq 2$ and $C \geq 0$, which has the form in (2.53) with $\gamma = 1/C$ and $G(t) = |t|^\alpha$. Then $\phi(t)$ is the characteristic function of Pólya-type; see for example, Bisgaard and Zoltan (2000).

Example 2.13 (Geometric). If $\Pr(P = k) = (1 - \alpha)\alpha^k$ ($0 < \alpha < 1$), $k = 0, 1, 2, \dots$, then $\phi(t) = (1 + \alpha(1 - e^{it})/(1 - \alpha))^{-1}$, which is also of the type in (2.53), with $\gamma = (1 - \alpha)/\alpha$ and $G(t) = 1 - e^{it}$.

Example 2.14 (Compound geometric). Let $\{X_n\}_{n=1}^\infty$ be i.i.d. with common distribution function F , and N a random variable independent of $\{X_n\}$. Then $X = \sum_{n=1}^N X_n$ is said to have a *compound distribution*, whose name is determined by the distribution of N . If N is geometrically distributed with $\Pr(N = n) = (1 - \theta)\theta^{n-1}$ (where $0 < \theta < 1$), $n = 1, 2, \dots$, then X is said to be *compound geometric*.

The compound geometric distributions arise in some practical situations of scheduling. For example, consider the situation where the processing of a job consists of many subtasks whose processing times are independently and identically distributed. The total processing time of the job is then compound geometrically distributed if the number of subtasks has a geometric distribution. As another example, a compound geometric distribution can arise when a task may not be done correctly, so it must be repeated until it is done correctly, where θ is the probability it is done incorrectly, and all repetitions are i.i.d.

The following proposition characterizes the compound geometric distribution by its characteristic functions. Its proof is straightforward and thus omitted.

Proposition 2.1. *If X is compound geometric, then its characteristic function has the form in (2.53).*

Example 2.15 (Lévy Process with Exponentially Distributed Horizon). Suppose that $\{X(t) : t \in I\}$ is a stochastic process with independent increments, where I is the set of its (time) horizon. Let T be a random variable taking values in I , independent of the process $\{X(t) : t \in I\}$. The random variable $X(T)$ is termed as a *process with random horizon*, and its distribution is called a *generalized compound distribution*. For simplicity, we will only discuss the Lévy process here, which is a process $\{X(t), t \geq 0\}$ with increment $X(t + s) - X(t)$ independent of the process $\{X(v), 0 \leq v \leq t\}$ and has the same distribution law as $X(s)$, for every $s, t \geq 0$.

The generalized compound distribution may arise from practical situations as well. Consider a manufacturing practice in which processing time is a *Lévy Process with Exponentially Distributed Horizon*. Suppose that the processing of a job is to repair the flaw of the job (or product). The flaw is an irregular area with length T and irregular width, which gives rise to a random processing time on repairing any fixed length of flaw area. The processing time to repair a flaw of l units of length is a random variable, related only to the length but not the position of the flaw. Then the processing time on a length of flaw is a Lévy process with the length as its time parameter. So if the length of flaw is distributed exponentially, then the processing time is a Lévy process with exponentially distributed horizon.

We now calculate the characteristic function of $X(T)$, which again has the form in (2.53). Denote the characteristic exponent of $X(1)$ by $\Psi(t)$, i.e., $E[e^{isX(1)}] = e^{-\Psi(s)}$. Then we have the following result.

Proposition 2.2. *If $\{X(t), t \geq 0\}$ is a Lévy process and T is exponentially distributed with rate γ , independent of $\{X(t), t \geq 0\}$, then the characteristic function of $X(T)$ is $\phi_{X(T)}(s) = (1 + \Psi(s)/\gamma)^{-1}$.*

Proof. First note that a Lévy process $X(t)$ is infinitely divisible. Then the characteristic function of $X(t)$ is given by $\phi_{X(t)}(s) = e^{-t\Psi(s)}$ (see, for example, Bertoin 1996). Hence if T is exponentially distributed with rate γ , independent of the Lévy process $\{X(t), t \geq 0\}$, then the characteristic function of $X(T)$ is

$$\begin{aligned}\phi_{X(T)}(s) &= E[e^{isX(T)}] = E\{E[e^{isX(T)}|T]\} = E[e^{-T\Psi(s)}] \\ &= \gamma \int_0^\infty e^{-t\Psi(s) - \gamma t} dt = (1 + \Psi(s)/\gamma)^{-1}.\end{aligned}$$

This completes the proof. ■

Remark 2.3. One question of interest is whether the distributions in the class given by (2.53) can be likelihood-ratio ordered. While some of the distributions in our examples above can indeed be likelihood-ratio ordered, such as the exponential and geometric distributions, this is not the case in general. For example, consider the Laplace distributions in Example 2.11 above. Let $f_1(x) = (2\alpha_1)^{-1}e^{-|x|/\alpha_1}$ and $f_2(x) = (2\alpha_2)^{-1}e^{-|x|/\alpha_2}$ be two Laplace densities. Then the likelihood-ratio

$$\frac{f_1(x)}{f_2(x)} = \frac{(2\alpha_1)^{-1}e^{-|x|/\alpha_1}}{(2\alpha_2)^{-1}e^{-|x|/\alpha_2}} = \frac{\alpha_2}{\alpha_1} e^{(1/\alpha_2 - 1/\alpha_1)|x|}$$

is not monotone in x . Therefore $f_1(x)$ and $f_2(x)$ cannot be likelihood-ratio ordered.

2.6.2 Optimal Sequences for Total Expected Costs

Some orders between two nondecreasing functions are defined below to shorten the notation. Suppose that $H_1(x)$ and $H_2(x)$ are two nondecreasing functions.

- $H_1(x)$ is said to be prior to $H_2(x)$ in *increment order*, denoted as $H_1 \prec_{\text{inc}} H_2$, if $H_2(x) - H_1(x)$ is nondecreasing in x , or equivalently, $H_2(x)$ has greater increments than $H_1(x)$ in the sense that

$$H_1(t) - H_1(s) \leq H_2(t) - H_2(s) \quad \text{for } t > s. \quad (2.55)$$

In this case we also say that H_2 is *steeper* than H_1 .

- H_1 is said to be prior to H_2 in *convexity order*, written $H_1 \prec_{\text{cv}} H_2$ (or $H_2 \succ_{\text{cv}} H_1$), if $H_2(x)$ has more convexity than $H_1(x)$ in the sense that

$$\begin{aligned} H_1(\alpha s + (1 - \alpha)t) - \alpha H_1(s) - (1 - \alpha)H_1(t) \\ \geq H_2(\alpha s + (1 - \alpha)t) - \alpha H_2(s) - (1 - \alpha)H_2(t) \end{aligned} \quad (2.56)$$

for all $\alpha \in (0, 1)$, which is equivalent to the convexity of $H_2(x) - H_1(x)$.

Let $H_{ij}(t) = \gamma_i m_i(t) - \gamma_j m_j(t)$. It follows from (2.55) and (2.56) that

$$\gamma_i m_i \prec_{\text{inc}} \gamma_j m_j \iff H_{ij}(t) \text{ is nonincreasing}; \quad (2.57)$$

$$\gamma_i m_i \prec_{\text{cv}} \gamma_j m_j \iff H_{ij}(t) \text{ is concave}. \quad (2.58)$$

The following theorem presents the optimal sequence to minimize $E[\sum f_j(C_j)]$.

Theorem 2.19. Let $P_i \sim \phi_i(t) = (1 + G(t)/\gamma_i)^{-1}$, $i = 1, \dots, n$.

(a) For nonnegative $P_1, \dots, P_n$, the sequence $\{1, 2, \dots, n\}$ minimizes $E[\sum f_j(C_j)]$ if

$$i > j \implies \gamma_i m_i \prec_{\text{inc}} \gamma_j m_j. \quad (2.59)$$

In other words, the non-increasing order of $\{\gamma_i m_i(t)\}$ in the $\prec_{\text{inc}}$ sense is optimal if such an order exists.

(b) When $P_1, \dots, P_n$ are real-valued random variables with nonnegative means, the sequence $\{1, 2, \dots, n\}$ minimizes $E[\sum f_j(C_j)]$ if

$$i > j \implies \gamma_i m_i \prec_{\text{inc}} \gamma_j m_j \quad \text{and} \quad \gamma_i m_i \prec_{\text{cv}} \gamma_j m_j. \quad (2.60)$$

In other words, if $\{\gamma_i m_i(t)\}$ have the same order in the $\prec_{\text{inc}}$ and $\prec_{\text{cv}}$ sense, then the nonincreasing order in either sense is optimal.

Proof. Let $\pi = \{\dots, i, j, \dots\}$ be a job sequence with $i > j$, $\pi' = \{\dots, j, i, \dots\}$ be the sequence by interchanging two consecutive jobs i, j in π , and C denote the completion time of the job prior to job i under π . Then for the objective function $TEC(\pi) = E[\sum f_j(C_j)]$, since $\{f_i(t)\}$ are independent of $\{P_i\}$,

$$\begin{aligned} TEC(\pi) - TEC(\pi') &= E[f_i(C + P_i)] + E[f_j(C + P_i + P_j)] - E[f_j(C + P_j)] - E[f_i(C + P_i + P_j)] \\ &= E[m_i(C + P_i)] + E[m_j(C + P_i + P_j)] - E[m_j(C + P_j)] - E[m_i(C + P_i + P_j)]. \end{aligned}$$

Denote the cdf's of P_i , P_j and $P_i + P_j$ by $F_i(x)$, $F_j(x)$ and $F_{ij}(x)$ respectively as in Lemma 2.6. Since P_i, P_j and C are independent, conditional on C ,

$$\begin{aligned} TEC(\pi) - TEC(\pi') &= E[m_i(C + P_i)] - E[m_i(C + P_i + P_j)] - E[m_j(C + P_j)] + E[m_j(C + P_i + P_j)] \\ &= E\left[\int_{-\infty}^{\infty} m_i(C + x)d[F_i(x) - F_{ij}(x)]\right] - E\left[\int_{-\infty}^{\infty} m_j(C + x)d[F_j(x) - F_{ij}(x)]\right]. \end{aligned}$$

By Lemma 2.6,

$$\begin{aligned} TEC(\pi) - TEC(\pi') &= \frac{1}{\gamma_i} E\left[\int_{-\infty}^{\infty} (\gamma_i m_i(C + x) - \gamma_j m_j(C + x))d[F_i(x) - F_{ij}(x)]\right] \\ &= \frac{1}{\gamma_i} \left\{E[H_{ij}(C + P_i)] - E[H_{ij}(C + P_i + P_j)]\right\}. \end{aligned} \quad (2.61)$$

Consider two cases corresponding to parts (a) and (b) of the theorem.

Case 1. $P_1, \dots, P_n$ are nonnegative variables. Under the condition that $i > j$ implies $\gamma_i m_i \prec_{\text{inc}} \gamma_j m_j$, $H_{ij}(t)$ is nonincreasing by (2.57). Hence by (2.61) it is clear that

$$TEC(\pi) - TEC(\pi') = \frac{1}{\gamma_i} E[H_{ij}(C + P_i) - H_{ij}(C + P_i + P_j)] \geq 0. \quad (2.62)$$

Case 2. $P_1, \dots, P_n$ are real-valued with $E[P_i] \geq 0$, $i = 1, \dots, n$. Then under the condition that $H_{ij}(x)$ are non-increasing and concave functions,

$$\begin{aligned} TEC(\pi) - TEC(\pi') &= \frac{1}{\gamma_i} \{E[H_{ij}(C + P_i) - H_{ij}(C + P_i + P_j)]\} \\ &\geq \frac{1}{\gamma_i} \{E[H_{ij}(C + P_i) - H_{ij}(C + P_i + E[P_j])]\} \geq 0, \end{aligned} \quad (2.63)$$

where the first inequality follows from applying Jensen's inequality to the concave $H_{ij}(x)$ conditional on $C + P_i$, and the second inequality holds because $H_{ij}(x)$ is non-increasing. ■

Remark 2.4. Pinedo and Wei (1986) obtained the optimal schedule to minimize the total expected waiting cost $\sum E[g(C_i)]$ with a general but deterministic waiting cost

function g . This is a special case of $TEC(\pi) = E[\sum f_j(C_j)]$ with all f_j equal to a common deterministic function. On the other hand, the results of Pinedo and Wei (1986) allow more general distributions of the processing times and multiple machines in a flowshop setting.

2.6.3 Optimal Sequences with Due Dates

We now discuss the cost function $E[\sum w_j g(C_j - D_j) I_{\{C_j > D_j\}}]$ with random due dates $\{D_i\}$. An application of Theorem 2.19 yields the next theorem.

Theorem 2.20. *Suppose that $\{D_i\}$ have a common distribution. Then a sequence in nonincreasing order of $\{\gamma_i w_i\}$, or equivalently, in nondecreasing order of $\{E[P_i]/w_i\}$, minimizes the $TEC(\pi) = E[\sum w_j g(C_j - D_j) I_{\{C_j > D_j\}}]$ if either*

- (1) $g(t)$ is an increasing function and $\{P_i\}$ are nonnegative, or
- (2) $g(t)$ is a convex and non-decreasing function and $\{P_i\}$ are real-valued with nonnegative means.

Proof. We first note by Lemma 2.6 that nonincreasing order of $\{\gamma_i w_i\}$ is equivalent to nondecreasing order of $\{E[P_i]/w_i\}$ since γ_i may take $1/E[P_i]$. Let

$$f_i(t) = w_i g(t - D_i), \quad i = 1, 2, \dots, n,$$

and D be a representative of $\{D_i\}$. Then $m_i(t) = E[f_i(t)] = w_i E[g(t - D)]$, which gives $H_{ij} = (w_i \gamma_i - w_j \gamma_j) E[g(t - D)]$. Since $g(t)$ is nondecreasing,

$$\gamma_i m_i \prec_{\text{inc}} \gamma_j m_j \iff H_{ij}(t) \text{ is nonincreasing} \iff w_i \gamma_i \leq w_j \gamma_j. \quad (2.64)$$

Thus if $\gamma_1 w_1 \geq \dots \geq \gamma_n w_n$, then $\{1, \dots, n\}$ is optimal by Theorem 2.19 and so a sequence in non-increasing order of $\{\gamma_i w_i\}$ minimizes $TEC(\pi)$. This proves the optimality result under condition (1).

Furthermore, when condition (2) holds, note that a convex g implies that

$$\gamma_i m_i \prec_{\text{cv}} \gamma_j m_j \iff H_{ij}(x) \text{ is concave} \iff w_i \gamma_i \leq w_j \gamma_j. \quad (2.65)$$

Combining (2.65) with (2.64), the result under condition (2) follows. ■

Remark 2.5. Theorem 2.20 reveals an interesting fact regarding the problem of minimizing the expected discounted cost function $E[\sum w_j (1 - e^{-rC_j})]$. This problem is a special case of the model in Theorem 2.20 by setting $D_i = 0$ and taking $g(t) = 1 - e^{-rt}$. It is well-known that the sequence with weighted discounted shortest expected processing time first rule (WDSEPT) is optimal for this problem.

Theorem 2.20 says that when the distributions of processing times are of compound-type, the WDSEPT rule reduces to the WSEPT (weighted shortest expected processing time first) rule.

We next consider due dates with different distributions. Let $Q_i(x) = \Pr(D_i \leq x)$ denote the cdf of due date D_i , $i = 1, \dots, n$.

Theorem 2.21. *Let $P_1, \dots, P_n$ be non-negative random variables and $g(\cdot)$ non-decreasing.*

(a) *The sequence $\{1, 2, \dots, n\}$ minimizes the $TEC(\pi)$ in (2.16) if*

$$\gamma_1 w_1 Q_1 \succ_{\text{inc}} \gamma_2 w_2 Q_2 \succ_{\text{inc}} \dots \succ_{\text{inc}} \gamma_n w_n Q_n. \quad (2.66)$$

(b) *If in addition, $g(\cdot)$ is also a convex (concave) function, then the sequence $\{1, 2, \dots, n\}$ ($\{n, n-1, \dots, 2, 1\}$) is optimal if*

$$\gamma_1 w_1 Q_1(x) \geq \gamma_2 w_2 Q_2(x) \geq \dots \geq \gamma_n w_n Q_n(x). \quad (2.67)$$

Proof. Let $f_i(t) = w_i g(t - D_i)$ then

$$m_i(t) = w_i E[g(t - D_i)] = w_i \int_0^\infty g(t - x) dQ_i(x),$$

which yields

$$H_{ij}(t) = \gamma_i m_i(t) - \gamma_j m_j(t) = \int_0^\infty g(t - x) d[\gamma_i w_i Q_i(x) - \gamma_j w_j Q_j(x)]. \quad (2.68)$$

For $i > j$, since $\gamma_i w_i Q_i \prec_{\text{inc}} \gamma_j w_j Q_j$, $\gamma_i w_i Q_i(x) - \gamma_j w_j Q_j(x)$ is nonincreasing in x . Hence $H_{ij}(t)$ is nonincreasing by (2.68) for $g(t - x)$ is nondecreasing in t , which is equivalent to $\gamma_i m_i \prec_{\text{inc}} \gamma_j m_j$. Part (a) of the theorem then follows from Theorem 2.19.

We now turn to part (b). When $g(x)$ is convex, it has finite right derivative, denoted as $g'_+(x)$, at every point x . Moreover, the convexity of g implies

$$g(t - x) - g(s - x) \geq (t - s)g'_+(s - x) \geq 0,$$

which in turn implies that $g(t - x) - g(s - x)$ is bounded from below with respect to x for arbitrary $t > s$. Hence (2.68) can be rewritten as

$$H_{ij}(t) - H_{ij}(s) = - \int_0^\infty [\gamma_i w_i Q_i(x) - \gamma_j w_j Q_j(x)] d[g(s - x) - g(t - x)].$$

As g is convex, $g(t - x) - g(s - x)$ is non-increasing in x . If $\gamma_i w_i Q_i \leq \gamma_j w_j Q_j$, then $H_{ij}(t) - H_{ij}(s) \leq 0$ for $t > s$, i.e., $\gamma_i m_i \prec_{\text{inc}} \gamma_j m_j$. Hence by Theorem 2.19, $\{1, 2, \dots, n\}$ is optimal when g is convex. Similarly, if g is concave, then $g(t - x)$

$-g(s-x)$ is non-decreasing in x , so that $H_{ij}(t) - H_{ij}(s) \geq 0$ for $t > s$. Thus by Theorem 2.19 again, $\{n, \dots, 2, 1\}$ is optimal when g is concave. ■

Remark 2.6. Part (a) of Theorem 2.21 shows that the order $\succ_{\text{inc}}$ between $\gamma_i w_i Q_i(x)$ leads to an optimal policy. If the due dates are identically distributed, then the optimality condition reduces to $\gamma_1 w_1 \geq \gamma_2 w_2 \geq \dots \geq \gamma_n w_n$. To show an example of the distributions of due dates such that Theorem 2.21 applies, let the due dates D_i be exponentially distributed with rate λ_i , $i = 1, \dots, n$. Suppose that $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ and $\gamma_1 w_1 \lambda_1 \geq \gamma_2 w_2 \lambda_2 \geq \dots \geq \gamma_n w_n \lambda_n$. Then for any $i < j$ and $t > s > 0$, there exists $\xi \in (s, t)$ such that

$$\frac{\gamma_i w_i [Q_i(t) - Q_i(s)]}{\gamma_j w_j [Q_j(t) - Q_j(s)]} = \frac{\gamma_i w_i Q'_i(\xi)}{\gamma_j w_j Q'_j(\xi)} = \frac{\gamma_i w_i \lambda_i}{\gamma_j w_j \lambda_j} e^{(\lambda_j - \lambda_i)\xi} \geq 1.$$

Therefore $i < j \implies \gamma_i w_i Q_i \succ_{\text{inc}} \gamma_j w_j Q_j$ and so part (a) of Theorem 2.21 applies.

Remark 2.7. Part (b) of Theorem 2.21 indicates that when the cost function g is convex (concave) and nondecreasing, the requirement for the order $\succ_{\text{inc}}$ reduces to the point-wise order of the functions $\{\gamma_i w_i Q_i(x)\}$, which will further reduce to the stochastic order of the due dates D_i when $\gamma_i w_i$ equal a common value for all jobs. If g is merely non-decreasing but not convex (concave), the pointwise order in (2.67) does not ensure the optimality of the sequence $\{1, 2, \dots, n\}$ ($\{n, \dots, 2, 1\}$); see the following example.

Example 2.16. Consider $n = 2$ and a discounted cost function $g(x) = (1 - e^{-x})I_{\{x \geq 0\}}$. Let $\Pr(P_1 > x) = \Pr(P_2 > x) = e^{-x}$, $x \geq 0$. Then $\gamma_1 = \gamma_2 = 1$, and the density function for $P_1 + P_2$ is given by xe^{-x} , $x > 0$. The due dates D_1 and D_2 are deterministic d_1 and d_2 respectively. Let $\pi_1 = \{1, 2\}$ and $\pi_2 = \{2, 1\}$, and assume $w_1 = w_2 = 1$. Then for $TEC(\pi) = \sum w_i E[g(P_i - d_i)I_{\{P_i > d_i\}}]$,

$$\begin{aligned} TEC(\pi_1) &= E[g(P_1 - d_1)] + E[g(P_1 + P_2 - d_2)] \\ &= \int_{d_1}^{\infty} (1 - e^{-(x-d_1)})e^{-x} dx + \int_{d_2}^{\infty} (1 - e^{-(x-d_1)})xe^{-x} dx \\ &= \frac{1}{2}e^{-d_1} + \frac{1}{4}(3 + 2d_2)e^{-d_2} \end{aligned}$$

and similarly,

$$TEC(\pi_2) = \frac{1}{2}e^{-d_2} + \frac{1}{4}(3 + 2d_1)e^{-d_1}.$$

Therefore,

$$TEC(\pi_1) - TEC(\pi_2) = \frac{1}{4}(1 + 2d_2)e^{-d_2} - \frac{1}{4}(1 + 2d_1)e^{-d_1}. \quad (2.69)$$

It is easy to check that $(1 + 2d)e^{-d}$ is increasing in $d < 0.5$ and decreasing in $d > 0.5$. Hence if $d_1 < d_2 < 0.5$, then $(1 + 2d_1)e^{-d_1} < (1 + 2d_2)e^{-d_2}$ and so by (2.69), $TEC(\pi_1) > TEC(\pi_2)$. On the other hand,

$$d_1 < d_2 \implies \gamma_1 w_1 Q_1(x) = I_{\{x \geq d_1\}} \geq I_{\{x \geq d_2\}} = \gamma_2 w_2 Q_2(x).$$

Therefore, when $d_1 < d_2 < 0.5$, condition (2.67) is satisfied but $\pi_1 = \{1, 2\}$ is not optimal. Similarly when $0.5 < d_1 < d_2$, (2.67) holds but $\pi_2 = \{2, 1\}$ is not optimal.

We present the next theorem without proof, which is similar to that of Theorem 2.21.

Theorem 2.22. *Let $P_1, \dots, P_n$ be real-valued random variables with nonnegative means and g non-decreasing. Then the sequence $\{1, 2, \dots, n\}$ minimizes the $TEC(\pi)$ if either*

- (1) *g is convex and $\gamma_1 w_1 Q_1 \succ_{\text{inc}} \gamma_2 w_2 Q_2 \succ_{\text{inc}} \dots \succ_{\text{inc}} \gamma_n w_n Q_n$, or*
- (2) *g is differentiable with convex derivative $g'(x)$ and $\gamma_1 w_1 Q_1(x) \geq \dots \geq \gamma_n w_n Q_n(x)$.*

Remark 2.8. The previous results can be easily extended to the case with precedence constraints in the form of nonpreemptive chains. Suppose that jobs $\{1, \dots, n\}$ are divided into m nonpreemptive chains $u_i = \{i_1, \dots, i_{k_i}\}$, $i = 1, \dots, m$. Each chain is subject to precedence constraints in the sense that jobs within a chain must be processed according to a specified order. The chains are nonpreemptive in the sense that once the machine starts to process a job in a chain, it cannot process any job in a different chain until all jobs in the current chain are finished. The scheduling problem then becomes one of ordering the m chains $\{u_1, \dots, u_m\}$. We can extend the results in Theorems 2.19–2.22 straightforwardly to the scheduling problems under such precedence constraints as follows.

For two jobs i and j , we define $i \prec j$ in accordance with each situation in Theorems 2.19–2.22 that leads to an optimal sequence. For example, in the case of Theorem 2.19, we define $i \prec j$ if $\gamma_i m_i \succ_{\text{inc}} \gamma_j m_j$ in part (a); or $\gamma_i m_i \succ_{\text{inc}} \gamma_j m_j$ and $\gamma_i m_i \succ_{\text{cv}} \gamma_j m_j$ in part (b), and so on. If the chains $\{u_1, \dots, u_m\}$ can be ordered such that $u_1 \prec u_2 \prec \dots \prec u_m$, where $u_i \prec u_j \iff k \prec l$ for all $k \in u_i$ and $l \in u_j$, then the optimal sequence of the chains is in the order of $(u_1, u_2, \dots, u_m)$. For example, when the processing times are nonnegative, if $i < j \implies \gamma_k m_k \succ_{\text{inc}} \gamma_l m_l$ for all $k \in u_i$ and $l \in u_j$, then the sequence $(u_1, u_2, \dots, u_m)$ is optimal. This extends the result in part (a) of Theorem 2.19. Similarly we can extend any other result in Theorems 2.19–2.22.

Optimal Stochastic Scheduling

Cai, X.; Wu, X.; Zhou, X.

2014, X, 416 p. 3 illus., Hardcover

ISBN: 978-1-4899-7404-4