Stochastic Convexity and Stochastic Majorization

Cheng-Shang Chang
IBM Research Division
T.J. Watson Research Ctr.
Yorktown Hts., NY 10598

J. George Shanthikumar*
Haas School of Business
University of California
Berkeley, CA 94720

David D. Yao†
IE/OR Department
Columbia University
New York, NY 10027

September 1992

Abstract

Convexity/concavity and supermodularity/submodularity are useful second-order properties in optimization. Majorization and the related arrangement ordering are key properties that support pairwise-interchange arguments. Together they play an important role in resource allocation, production planning, and scheduling. Here we present the essentials of the recently developed theory of stochastic convexity and stochastic majorization, emphasizing their interplay. We illustrate the many applications of the theory through examples in production systems, including a random yield model, a joint setup problem, a manufacturing process with trial runs, a production network with constant work-in-process (WIP) and WIP-dependent production rate, and scheduling in tandem production lines and parallel assembly systems.

*Research supported in part by NSF grant DDM-91-13008

†Research supported in part by NSF grant DDM-91-08540

1 Introduction

Consider a production process that involves two stages. The first stage consists of m subassemblies, each producing one type of components. A set of components, one or several units from each type, are then assembled at the second stage. For instance, the subassemblies produce electronic components — various types of transistors, resistors, capacitors, etc., which are assembled at a second stage into circuit boards. Suppose the component yield of each type i is random: given a production lot of n_i units, the yield — number of good units — is a random variable $Y_i(n_i, p_i)$, where p_i denotes the yield ratio, a deterministic parameter that measures the quality of the subassembly. Without loss of generality, suppose a single unit from each type is needed for the assembly, then the number of assembled products, i.e., the assembly yield, is

$$Y(n,p) = \min\{Y_i(n_i,p_i),\ i = 1, ..., m\},$$

where $n = (n_i)_{i=1}^m$ and $p = (p_i)_{i=1}^m$.

In studying the optimal design of the above process, we would like to know properties of the yield, at both the component and assembly levels, with respect to the parameters (n, p). For instance, is Y_i increasing in n_i and p_i in some sense? (Naturally, one would expect the yield to be higher when the lot size is larger and the quality is better — in terms of a higher yield ratio.) What about second-order properties of the yield with respect to the parameters, such as concavity or convexity, submodularity or supermodularity, in some sense? We know these are all useful properties in optimization. For instance, concavity implies decreasing marginals in yield. Supermodularity in (n_i, p_i) implies that lot size and yield ratio enhance each other in increasing the yield; that is, the marginals (in yield) with respect to the lot size is increasing in the yield ratio, and vice versa.

Obviously, in order to address these issues we need to know more about the random variable Y_i as a function of the parameters (n_i, p_i). In other words, we need to be more specific about the "yield model". Consider the simplest case, for instance,

$$Y_i(n_i,p_i) = \sum_{j=1}^{n_i} Y_i^j(p_i),$$

where $Y_i^j(p_i)$, $j = 1, ..., n_i$, are iid Bernoulli random variables with success probability p_i. In other words, this is an additive yield model: each unit j in the lot is independently good or bad modeled as a Bernoulli variate. This way, $Y_i(n_i, p_i)$ is a binomial variate with parameters (n_i, p_i). When n_i is large, Y_i can also be viewed as a normal variate with mean $n_i p_i$ and variance $n_i p_i(1 - p_i)$. So it appears as if the parametric behavior of the yield can be studied through it distribution function.

However, one quickly realizes that it is not always possible to rely on the availability of explicit distribution functions. For instance, even if Y_i's are independent, and all follow binomial or normal distributions (with different parameters), it is still difficult to derive the distribution function of $Y = \min_i\{Y_i\}$. Therefore, what is needed is a theory that does not rely on explicit distribution functions.

In another direction, one can also ask questions such as whether a subassembly with a higher yield ratio should be given a larger lot size, and if all the yield ratios are the same then shouldn't the lot sizes be made as evenly as possible among the subassemblies? These allude to properties of the yield with respect to the parameters under *arrangement* and *majorization* orderings.

Arrangement and majorization orderings often arise in pairwise interchange arguments. In addition to the assembly example above, consider the following simple scheduling problem: Two jobs are to be processed in sequence on a single processor. Their processing times are x and y units, respectively (x and y are real values). Process x first and then y, the total delay ("flow time") of both jobs is $x + (x + y) = 2x + y$. Interchange the processing sequence, i.e., y first and then x. The flow time is $y + (x + y) = x + 2y$. Suppose $x > y$, then $2x + y > x + 2y$. That is, the second sequence is better in terms of a smaller flow time. This is essentially the argument for optimality of the SPT (shortest processing time) rule in deterministic scheduling. (When there are more than two jobs, repeatedly apply the above argument to each pair, each interchange of the processing sequence — shorter job first — results in a decrease of the flow time.) The natural question to ask here is whether the result still holds in the stochastic setting, i.e., when the processing times are random variables.

In the deterministic setting, there is a rich theory of convexity and its applications in optimization (e.g., Rockafellar [17], and an equally rich theory of rearrangement and majorization inequalities (e.g., Hardy, Littlewood and Polya [10], Marshall and Olkin [14]). In contrast, studies on the stochastic counterpart of these theories have been largely a recent endeavor.

The purpose of this article is to introduce the reader to those recent developments on the stochastic counterpart of convexity and rearrangement/majorization inequalities. To facilitate exposition, we will focus on a selected set of special cases of the general theory, and motivate them through concrete examples in production and scheduling such as the ones introduced above.

In §2, we start with reviewing the definitions of several commonly used stochastic order relations: likelihood ratio ordering, hazard rate ordering, stochastic ordering, and convex ordering. In particular, we focus on the functional representation of likelihood ratio and hazard rate orderings. (Functional representations for stochastic and convex orderings are standard and well known.) We demonstrate that the class of functions that preserve the arrangement ordering, i.e., *arrangement increasing function*, form a defining class for the likelihood ratio ordering. The weaker, hazard rate ordering can then be characterized by a subclass of the arrangement increasing functions. These functional characterizations will play a crucial role in the later sections.

We next move on to the discussion of second-order stochastic properties in §3. We start with stochastic convexity. In addition to a natural, functional definition, we also present two sample-path based defintions: one is stronger, which mimicks the three-point definition of deterministic convexity (the third point being the convex combination of any two given points); the other is weaker, requiring a four-point construction. Several examples are presented to illustrate in detail the techniques involved in the four-point construction. We then introduce the definitions for stochastic supermodularity and submodularity, and illustrate them through several examples. These, when combined with stochastic convexity, are particularly useful in applications. For instance, the component yield, $Y_i(n_i, p_i)$, introduced earlier is stochastically linear (i.e., both convex and concave) in n_i and in p_i in the weaker (four-point) sense, and stochastically supermodular in (n_i, p_i). These properties are also useful in characterizing the spatial and temporal behavior of Markov processes.

Section 4 serves as a transition into stochastic majorization. Motivated by the characterization of likelihood ratio ordering via arrangement increasing functions (§2), we now know that random variables that are ordered in likelihood ratio naturally induce arrangement and majorization orderings as well. This is indeed the case in a class of production networks modeled as closed queueing networks. Majorization and arrangement orderings in this context greatly facilitate the solutions to several resource allocation problems.

Motivated by the relation between the (stochastic) likelihood ratio ordering and the (deterministic) arrangement and majorization orderings, in §5 we systematically develop the stochastic counterpart of deterministic rearrangement and majorization inequalities such as the well known Hardy-Littlewood-Polya inequality, and the more general Day's inequality (see, [10, 6]; also, [14]). We consider three different settings. The first one is non-parametric, where the random variables involved are ordered under certain stochastic order relations. We show that under likelihood ratio ordering, the stochastic version is no different from the deterministic counterpart: in both cases, functions that are increasing and Schur convex form the defining class. On the other hand, when weaker orderings such as hazard rate and stochastic orderings are involved, the defining class has to be narrowed by strengthening Schur convexity to convexity in each variable *plus* submodularity. The second and the third settings are both parametric, i.e., the random variables invovled are from parametric families (such as those involved in the yield model). There, in addition to the defining class of functions being symmetric, increasing, submodular, and convex in each variable, the families of random variables involved are also required to be either stochastically convex or stochastically supermodular (with respect to their parameters). Results in these two settings demonstrate the close interplay between stochastic convexity and stochastic majorization. And it is exactly through this close interplay of the two theories that we are able to solve a variety of problems that involve pairwise-interchange arguments. We illustrate this through several scheduling examples.

2 Stochastic Order Relations: Functional Characterizations

Throughout this section, let X and Y be two random variables; let $F(\cdot)$ and $G(\cdot)$ denote their distribution functions, $f(\cdot)$ and $g(\cdot)$ their densities, and let $\bar{F} = 1 - F$ and $\bar{G} = 1 - G$. Throughout, the existence of any expectations is implicitly assumed; $\overset{\mathrm{d}}{=}$ denotes equal in distribution; "increasing" and "decreasing" are used in the non-strict sense.

First recall the following definitions of some commonly used stochastic order relations.

Definition 2.1 (i) *Likelihood ratio ordering:* $X \geq_{\ell r} Y$, if their likelihood ratios are ordered: $f(y)/f(x) \leq g(y)/g(x)$ for all $x \geq y$.
(ii) *Hazard rate ordering:* $X \geq_{hr} Y$, if their hazard rates are ordered: $f(x)/\bar{F}(x) \leq g(x)/\bar{G}(x)$ for all x.
(iii) *Stochastic ordering:* $X \geq_{st} Y$, if $\bar{F}(x) \geq \bar{G}(x)$ for all x.
(iv) *Convex ordering:* $X \geq_{icx} Y$, if $\int_a^\infty \bar{F}(x)dx \geq \int_a^\infty \bar{G}(x)dx$ for all a.

For discrete random variables, likelihood ratio and hazard rate orderings are similarly defined: replace the density functions in Definition 2.1(i,ii) by probability mass functions (pmf's). The following results are easily verified from Definition 2.1.

Theorem 2.2 (i) $X \geq_{\ell r} Y \Rightarrow X \geq_{hr} Y \Rightarrow X \geq_{st} Y \Rightarrow X \geq_{icx} Y$.
(ii) $X \geq_{st} Y$ if and only if $E\phi(X) \geq E\phi(Y)$ for any function $\phi(x)$ that is increasing in x.
(iii) $X \geq_{icx} Y$ if and only if $E\phi(X) \geq E\phi(Y)$ for any function $\phi(x)$ that is increasing and convex in x.

Following Theorem 2.2(iii) above, one can also define a *concave ordering*: $X \geq_{icv} Y$ if and only if $E\phi(X) \geq E\phi(Y)$ for any function $\phi(x)$ that is increasing and concave in x. Since $\phi(x)$ is increasing and convex if and only if $-\phi(-x)$ is increasing and concave in x, we have $X \geq_{icx} Y$ if and only if $-X \leq_{icv} -Y$.

From Theorem 2.2(i,ii), the class of increasing (resp. increasing and convex) functions can be viewed as the defining class for stochastic (resp. convex) ordering. In this sense, Theorem 2.2(ii, iii) provides functional characterizations for stochastic and convex orderings. It does not seem possible, however, to come up with similar univariate functions as defining classes for likelihood ratio and hazard rate orderings. Hence, we resort to *bivariate* functions. We start from the likelihood ratio ordering.

Theorem 2.3 Define a class of bivariate functions as follows:

$$\mathbf{C}_{\ell r} = \{\phi(x,y) : \Delta\phi(x,y) = \phi(x,y) - \phi(y,x) \geq 0, \, \forall x \geq y\}.$$

Then, for two independent random variables X and Y, $X \geq_{\ell r} Y$ if and only if $E\phi(X,Y) \geq E\phi(Y,X)$, for all $\phi \in \mathbf{C}_{\ell r}$.

Proof. For the "if" part, given $u > v$, let $\phi(x,y) = \mathbf{1}[u \leq x \leq u + du, v \leq y \leq v + dv]$, where $du > 0$, $dv > 0$ and $dv < u - v$, then $\phi \in \mathbf{C}_{\ell r}$. Hence, $E\phi(X,Y) \geq E\phi(Y,X)$, or $f(u)g(v) \geq f(v)g(u)$, i.e., $X \geq_{\ell r} Y$.

For the "only if" part, taking into account $\Delta\phi(x,y) = -\Delta\phi(y,x)$, we have

$$E\Delta\phi(X,Y) = \int_y \int_x \Delta\phi(x,y) f(x) g(y) dx dy$$

$$= \int_y \int_{x \geq y} \Delta\phi(x,y)[f(x)g(y) - f(y)g(x)] dx dy \geq 0,$$

since $\phi \in \mathbf{C}_{\ell r} \Rightarrow \Delta\phi \geq 0$, and $X \geq_{\ell r} Y \Rightarrow f(x)g(y) \geq f(y)g(x)$ for all $x \geq y$. $\Box$

Note that since likelihood ratio ordering (as well as all the other orderings in Definition 2.1) is a relation concerning the *marginal* distributions of X and Y, i.e., there is no reference to any *joint* distributional relation of (X,Y), it is only natural to assume independence among the two random variables X and Y. Hence Theorem 2.3 can be rephrased as follows:

Corollary 2.4 $X \geq_{\ell r} Y$ if and only if there exist two independent random variables $\hat{X}$ and $\hat{Y}$, such that $\hat{X} \stackrel{\mathrm{d}}{=} X$, $\hat{Y} \stackrel{\mathrm{d}}{=} Y$, and $E\phi(\hat{X},\hat{Y}) \geq E\phi(\hat{Y},\hat{X})$ for all $\phi \in \mathbf{C}_{\ell r}$.

Similarly, we have the following defining class for the hazard rate ordering.

Theorem 2.5 Define $\mathbf{C}_{hr} = \{\phi(x,y) : \text{both } \phi(x,y) \text{ and } \Delta\phi(x,y) \text{ are increasing in } x, \forall x \geq y\}$. Then, $X \geq_{hr} Y$ if and only there exist two independent random variables $\hat{X}$ and $\hat{Y}$, such that $\hat{X} \stackrel{\mathrm{d}}{=} X$, $\hat{Y} \stackrel{\mathrm{d}}{=} Y$, and $E\phi(\hat{X},\hat{Y}) \geq E\phi(\hat{Y},\hat{X})$ for all $\phi \in \mathbf{C}_{hr}$.

Proof. First note that we have the following equivalent conditions to $X \geq_{hr} Y$ in Definition 2.1(ii):

$$\bar{F}(y)/\bar{F}(x) \leq \bar{G}(y)/\bar{G}(x), \quad \forall x \geq y, \tag{1}$$

and

$$f(y)/\bar{F}(x) \leq g(y)/\bar{G}(x), \quad \forall x \geq y. \tag{2}$$

For instance, Definition 2.1(ii) $\Rightarrow$ (1) follows from the following identities:

$$\bar{F}(x) = \exp\{-\int_{-\infty}^{x} \frac{f(t)}{\bar{F}(t)}\}dt, \quad \bar{G}(x) = \exp\{-\int_{-\infty}^{x} \frac{g(t)}{\bar{G}(t)}\}dt,$$

since

$$\frac{\bar{F}(y)}{\bar{F}(x)} = \exp\{\int_{y}^{x} \frac{f(t)}{\bar{F}(t)}dt\} \leq \exp\{\int_{y}^{x} \frac{g(t)}{\bar{G}(t)}dt\} = \frac{\bar{G}(y)}{\bar{G}(x)}.$$

Conversely, (1) implies, for $\delta > 0$,

$$\frac{\bar{F}(x-\delta) - \bar{F}(x)}{\delta \bar{F}(x)} \leq \frac{\bar{G}(x-\delta) - \bar{G}(x)}{\delta \bar{G}(x)}.$$

Letting $\delta \to 0$ gives Definition 2.1(ii). Next, Definition 2.1(ii) and (1) together imply (2):

$$\frac{f(y)}{\bar{F}(x)} = \frac{f(y)}{\bar{F}(y)} \cdot \frac{\bar{F}(y)}{\bar{F}(x)} \leq \frac{g(y)}{\bar{G}(y)} \cdot \frac{\bar{G}(y)}{\bar{G}(x)} = \frac{g(y)}{\bar{G}(x)}.$$

Finally, that (2) implies Definition 2.1(ii) is trivial: let $y = x$.

Now, to prove the "if" part, let $\phi(x, y) = \mathbf{1}[x \geq u, y \geq v]$, where $u \geq v$ are given. Then clearly, $\phi \in \mathbf{C}_{hr}$, in particular, $\Delta\phi(x, y) = \mathbf{1}[x \geq u, u > y \geq v]$ is increasing in x. Hence, $E\Delta\phi(\hat{X}, \hat{Y}) \geq 0$ recovers (1).

For the "only if" part, assuming differentiability of ϕ for simplicity, we have, similar to the proof of Theorem 2.3,

$$\begin{aligned} E\Delta\phi(\hat{X}, \hat{Y}) &= \int_y \int_{x \geq y} \Delta\phi(x, y)[f(x)g(y) - f(y)g(x)]dxdy \\ &= \int_y \int_{x \geq y} [\frac{\partial}{\partial x}\Delta\phi(x, y)][\bar{F}(x)g(y) - f(y)\bar{G}(x)]dxdy \geq 0, \end{aligned}$$

where the second equality follows from integration by parts, and the inequality is due to (2), as well as $\Delta\phi(x, y)$ increasing in x for $x \geq y$. $\square$

Clearly, $\mathbf{C}_{hr} \subset \mathbf{C}_{\ell r}$. It is also easy to verify that the class $\mathbf{C}_{\ell r}$ is closed under composition with increasing functions. That is, $\phi \in \mathbf{C}_{\ell r} \Rightarrow h(\phi(x, y)) \in \mathbf{C}_{\ell r}$ for any $h(z)$ that is increasing in z. Similarly, the class $\mathbf{C}_{hr}$ is closed under composition with increasing and convex functions:

$\phi \in \mathbf{C}_{hr} \Rightarrow h(\phi(x,y)) \in \mathbf{C}_{hr}$ for any $h(z)$ that is increasing and convex in z. In particular, denoting h' as the derivative, we have

$$\frac{\partial}{\partial x}[h(\phi(x,y)) - h(\phi(y,x))] = h'(\phi(x,y))\frac{\partial}{\partial x}\phi(x,y) - h'(\phi(y,x))\frac{\partial}{\partial x}\phi(y,x)$$
$$\geq h'(\phi(y,x))\frac{\partial}{\partial x}[\phi(x,y) - \phi(y,x)] \geq 0,$$

where the first inequality follows from the convexity of $h(\cdot)$, the increasingness of $\phi(x,y)$ in x for $x \geq y$ [so that $\frac{\partial}{\partial x}\phi(x,y) \geq 0$], and $\phi(x,y) \geq \phi(y,x)$, which is implied by $\Delta\phi(x,y)$ increasing in x for $x \geq y$. [Note, from the proof of (2.4), it is clear that the increasingness of $\phi(x,y)$ in x for $x \geq y$ is *not* needed there. It is only needed to make the class $\mathbf{C}_{hr}$ closed under composition with functions that are increasing and convex.] To summarize, taking into account Theorem 2.2, we have

Corollary 2.6 For two independent random variables, X and Y,
(i) $X \geq_{\ell r} Y \Leftrightarrow \phi(X,Y) \geq_{st} \phi(Y,X)$, $\forall\phi \in \mathbf{C}_{\ell r}$;
(ii) $X \geq_{hr} Y \Leftrightarrow \phi(X,Y) \geq_{icx} \phi(Y,X)$, $\forall\phi \in \mathbf{C}_{hr}$.

Example 2.7 Consider a single machine stochastic scheduling problem, where n jobs are to be scheduled with the objective of minimizing the flow time (i.e., sum of the time until completion for each job) in some stochastic sense. Let π denote a schedule such that job i is scheduled at the $\pi(i)$ th position, $i = 1, 2, ..., n$. (That is, π denotes a permutation of $\{1, 2, ..., n\}$.) Let X_i be the processing time of job i. The flow time under schedule π, denoted as T_π, can be derived as follows:

$$T_\pi = \sum_{i=1}^{n}[n+1-\pi(i)]X_i. \tag{3}$$

Consider two jobs a and b, and suppose $\pi(a) > \pi(b)$. Compare the schedule π with another schedule π' which interchanges the processing order of a and b: $\pi'(a) = \pi(b)$, $\pi'(b) = \pi(a)$, and $\pi'(i) = \pi(i), \forall i \neq a, b$. Denote $A = \sum_{i\neq a,b}[n+1-\pi(i)]X_i$. Then, we have

$$T_\pi = A + [n+1-\pi(a)]X_a + [n+1-\pi(b)]X_b, \tag{4}$$
$$T_{\pi'} = A + [n+1-\pi(a)]X_b + [n+1-\pi(b)]X_a. \tag{5}$$

Let $\phi(x,y) := [n+1-\pi(b)]x + [n+1-\pi(a)]y$. Since $\pi(a) > \pi(b)$, it is easy to verify that $\phi \in \mathbf{C}_{hr} \subset \mathbf{C}_{\ell r}$.

Proposition 2.8 For independent processing times, if $X_n \geq_{\ell r} ... \geq_{\ell r} X_1$ (resp., if $X_n \geq_{hr} ... \geq_{hr} X_1$), then $\pi = (1, ..., n)$ minimizes the flow time in the sense of stochastic (resp., convex) ordering.

Note in comparing the two flow times in (4) above, if the processing times are deterministic, then all that's needed is the following simple fact

$$a \geq b,\, x \geq y \quad \Rightarrow \quad ax + by \geq ay + bx, \tag{6}$$

where all variables are real. In the stochastic setting, however, the situation is much more complicated: different order relations among the stochastic processing times lead to different orderings in the flow times, as demonstrated in Proposition 2.8.

The inequality in (6) belongs to the class of "rearrangement inequalities", which play a key role in supporting pairwise interchange arguments that often arise in scheduling and other settings. On the other hand, the class of functions $\mathbf{C}_{\ell r}$ preserves a certain "arrangement" ordering, for instance, if $\phi(x,y) = ax + by$ with $a \geq b$, then $\phi(x,y)$ is larger than $\phi(y,x)$ when $x \geq y$. Therefore, the results in Proposition 2.8 can be viewed as stochastic versions of (6). In §4 and §5, we will pursue these ideas in a more formal and general way so as to derive stochastic versions of more general rearrangement inequalites to support more complicated applications.

3 Second-Order Stochastic Properties

3.1 Stochastic Convexity

Consider a family of random variables parameterized by a real or integer valued scalar θ, $\{X(\theta)\}$. Following (2.1 iii), $\{X(\theta)\}$ is stochastically increasing in θ, if $X(\theta_1) \leq_{st} X(\theta_2)$ for any $\theta_1 \leq \theta_2$. This is what we mean by *stochastic monotonicity*. An equivalent definition for stochastic monotonicity is:

Definition 3.1 $\{X(\theta)\}$ is stochastically increasing in θ, if for any given $\theta_1 \leq \theta_2$, there exist on a common probability space $(\Omega, \mathcal{F}, P)$ two random variables $\hat{X}_1$ and $\hat{X}_2$ that are equal in distribution to $X(\theta_1)$ and $X(\theta_2)$ respectively, and $\hat{X}_1(\omega) \leq \hat{X}_2(\omega)$ for all $\omega \in \Omega$.

Lemma 3.2 Suppose $F(x,\theta)$ denotes the distribution function of $X(\theta)$. Let $F^{-1}(u,\theta) = \inf\{x : F(x,\theta) > u\}$, where $u \in [0,1]$, denote the inverse distribution function. Then Definition 3.1 is also equivalent to $F^{-1}(u,\theta)$ increasing in θ, for any given $u \in [0,1]$. [The common probability space in this case is the one in which the uniform variate $U \in [0,1]$ is defined. In particular, note that $F^{-1}(U,\theta_i) \overset{\mathrm{d}}{=} X(\theta_i)$ for $i = 1,2$.]

From the above, it is easy to verify

Lemma 3.3 $X(\theta)$ is stochastically increasing in θ, if and only if for any real-valued function $\phi(x)$ that is increasing in x, $E\phi[X(\theta)]$ (as a *deterministic* function of θ) is increasing in θ. [Note this is consistent with the equivalence relation in Theorem 2.2(ii).]

Therefore, we have three equivalent definitions for stochastic monotonicity. We shall refer to the two definitions in Definition 3.1 and Lemma 3.2 as "sample-path" definitions, as versus the "functional" definition in Lemma 3.3. Roughly speaking, the sample-path definitions are easier to use in proofs (i.e., to establish stochastic monotonicity) while the functional version is more convenient in applications (for instance, when ϕ is part of an objective function that is to be optimized).

Now consider the stochastic *convexity* of $\{X(\theta)\}$. Parallel to the above discussion, we have the following definitions, which, however, are *not* equivalent, as we shall explain shortly below.

Definition 3.4 (SICX-sp). $\{X(\theta)\}$ is stochastically increasing and convex in the sample-path sense, denoted $\{X(\theta)\} \in$ SICX-sp, if for any four parameter values θ_i, $i = 1, 2, 3, 4$, that satisfy $\theta_1 + \theta_4 = \theta_2 + \theta_3$ and $\theta_4 \geq \max\{\theta_2, \theta_3\}$, there exist on a common probability space $(\Omega, \mathcal{F}, P)$ four random variables $\hat{X}_i$, $i = 1, 2, 3, 4$, such that $\hat{X}_i \stackrel{\mathrm{d}}{=} X(\theta_i)$, for all four i, and

$$\hat{X}_4(\omega) \geq \max\{\hat{X}_2(\omega), \hat{X}_3(\omega)\}, \quad \hat{X}_1(\omega) + \hat{X}_4(\omega) \geq \hat{X}_2(\omega) + \hat{X}_3(\omega),$$

for all $\omega \in \Omega$.

Definition 3.5 (SSICX). $\{X(\theta)\}$ is said to satisfy *strong* stochastic increasing convexity, denoted $\{X(\theta)\} \in$ SSICX, if $X(\theta)$ can be expressed as $X(\theta) \stackrel{\mathrm{d}}{=} \phi(\xi, \theta)$ where ϕ is an increasing and convex function with respect to θ, and ξ is a random variable whose distribution function is independent of θ. In particular, $X(\theta) \in$SSICX if $F^{-1}(u, \theta)$ is increasing and convex in θ (for any given $u \in [0, 1]$).

Definition 3.6 (SICX). $\{X(\theta)\}$ is stochastically increasing and convex in θ, denoted $\{X(\theta)\} \in$ SICX, if $E\phi[X(\theta)]$ is increasing and convex in θ for any function $\phi(x)$ that is increasing and convex in x.

Definitions for stochastic *concavity* are similarly defined as above. Also, stochastic *linearity* is defined as when both stochastic convexity and concavity are satisfied. We shall use CV and L (in place of CX) to denote concavity and linearity, respectively. Also, stochastically *decreasing* and convex (concave) can be similarly defined by replacing the condition for increasingness in the above definitions by an appropriate condition for decreasingness. For instance, in Definition 3.4, replace $\hat{X}_4 \geq \max\{\hat{X}_2, \hat{X}_3\}$ by $\hat{X}_1 \geq \max\{\hat{X}_2, \hat{X}_3\}$.

Remark 3.7 (i) As pointed out earlier, the above three definitions are not equivalent. In fact we have the following implications: SSICX $\Rightarrow$ SICX-sp $\Rightarrow$ SICX, which are easily verified from the definitions.

(ii) The strong version of stochastic convexity in Definition 3.5 can also be stated as follows: for any parameter values θ and η, and any constant $\alpha \in [0,1]$, there exist on a common probability space $(\Omega, \mathcal{F}, P)$ three random variables, $\hat{X}_1 \stackrel{\mathrm{d}}{=} X(\theta)$, $\hat{X}_2 \stackrel{\mathrm{d}}{=} X(\eta)$, and $\hat{X}_3 \stackrel{\mathrm{d}}{=} X(\alpha\theta + (1-\alpha)\eta)$ such that

$$\hat{X}_3(\omega) \leq \alpha \hat{X}_1(\omega) + (1-\alpha)\hat{X}_2(\omega)$$

for all $\omega \in \Omega$. In this sense, the sample values of $X(\omega, \theta)$ behave like a convex function with respect to θ. This version mimicks the convexity definition of a deterministic function. In contrast, the version in Definition 3.4 relaxes the regular convexity definition to a four-point version, and hence a weaker stochastic convexity.

(iii) The strong version in Definition 3.5 is satisfied by many families of random variables. For instance, if $X(\theta)$ is an exponential random variable with mean θ, then $X(\theta) = \theta\xi$ where ξ follows the exponential distribution with unit mean. Hence, $\{X(\theta)\} \in$ SSIL. The reader will verify that this property in fact extends to the exponential families (i.e., summations and mixtures of exponentials). Similarly, if $X(\theta)$ is a normal variate with mean μ and standard variation θ, then $X(\theta) = \theta Z + \mu$, where Z denotes the standard normal variate. Hence, for fixed μ, $\{X(\theta)\}$ is also stochastically linear in θ.

(iv) When the distribution function $F(x, \theta)$ is differentiable, it is also easy to verify SSICX, following Definition 3.5, via checking the first and second derivatives of X with respect to θ. To this end, write X as $X(\xi, \theta)$, and for each sample value of ξ, view X as a deterministic function of θ. Omitting reference to ξ, and taking derivatives from $F(X(\theta), \theta) = U$, we have

$$F_x X_\theta + F_\theta = 0, \quad F_x X_{\theta\theta} + F_{xx} X_\theta^2 + 2F_{x\theta} X_\theta + F_{\theta\theta} = 0,$$

where F_x, F_θ, X_θ, etc. denote partial derivatives with respect to x or θ. From the above, solve for X_θ and $X_{\theta\theta}$. Then $\{X(\theta)\} \in$ SSICX if both derivatives are non-negative.

Below, we illustrate that certain useful families of random variables do *not* satisfy SSICX, although they do satisfy the weaker SICX-sp property.

Example 3.8 Let $\{X(p)\}$ denote a family of Bernoulli random variables: $X(p) = \mathbf{1}[U \leq p]$, where $\mathbf{1}$ denotes the indicator function, $p \in (0,1)$ is the parameter, and $U \in [0,1]$ is the uniform variate. Since $E\phi[X(p)] = \phi(1)p$ is a linear function of p, $\{X(p)\}$ satisfies stochastic linearity in the sense of (3.2c) However, it is clear that $\{X(p)\}$ does not satisfy SSICX (or SSICV) in the sense of Definition 3.5, since the indicator function is neither convex nor concave.

The non-obvious fact here is that $\{X(p)\}$ satisfies SIL-sp in the sense of Definition 3.4. To show this, consider four parameter values: $p_1 \leq p_2 = p_3 \leq p_4$ and $p_1 + p_4 = p_2 + p_3$. Let

$$\hat{X}_i = \mathbf{1}[U \leq p_i], \quad i = 1, 2, 4;$$

and

$$\hat{X}_3 = \mathbf{1}[U \leq p_1] + \mathbf{1}[p_2 \leq U \leq p_4].$$

Then clearly $\hat{X}_i \stackrel{\mathrm{d}}{=} X(p_i)$ for all four i. In particular $\hat{X}_3 \stackrel{\mathrm{d}}{=} \mathbf{1}[U \leq p_3]$, since $p_4 - p_2 + p_1 = p_3$. Also, $\hat{X}_4 \geq \max\{\hat{X}_2, \hat{X}_3\}$, and

$$\hat{X}_1 + \hat{X}_4 = \mathbf{1}[U \leq p_1] + \mathbf{1}[U \leq p_2] + \mathbf{1}[p_2 \leq U \leq p_4] = \hat{X}_2 + \hat{X}_3.$$

Hence, $\{X(p)\}$ does satisfy SIL-sp.

Example 3.9 Now suppose $X(p)$ follows a geometric distribution with parameter p. $EX(p) = 1/p$ is decreasing and convex in p. Hence, $\{X(p)\}$ is stochastically decreasing and convex in the weak sense of Definition 3.6. We can further show that $\{X(p)\}$ in fact satisfies the stronger *sp* notion of stochastically decreasing and convex (cf. Definition 3.4). To this end, consider the four p_i values in Example 3.8. Use a common sequence of iid uniform variates $\{U_n \in [0, 1]; n = 1, 2, ...\}$ to generate the four replicas:

$$\hat{X}_i = \inf\{n : U_n \in [0, p_i]\}, \quad i = 1, 2, 4;$$

$$\hat{X}_3 = \inf\{n : U_n \in [0, p_1] \cup [p_2, p_4]\}.$$

Then, it is easy to see that $\hat{X}_1 \geq \hat{X}_i$ for $i \neq 1$, and $\hat{X}_4$ equals either $\hat{X}_2$ or $\hat{X}_3$, or both, hence $\hat{X}_1 + \hat{X}_4 \geq \hat{X}_2 + \hat{X}_3$.

Example 3.10 Let $X(n) = \sum_{i=1}^{n} Y_i$, where Y_i are iid non-negative random variables. We show $\{X(n)\}$ satisfies SIL-sp. To this end, pick four parameter values: $n, n+1, n+1$, and $n+2$. Use a sequence of independent uniform variates, $U_1, ..., U_n, U_{n+1}, U_{n+2}$ to generate, from the common distribution of Y_i, the sample values $\hat{Y}_1, ..., \hat{Y}_n, \hat{Y}_{n+1}, \hat{Y}_{n+2}$, respectively. Let

$$\hat{X}_1 = \sum_{i=1}^{n} \hat{Y}_i, \quad \hat{X}_2 = \sum_{i=1}^{n+1} \hat{Y}_i, \quad \hat{X}_4 = \sum_{i=1}^{n+2} \hat{Y}_i,$$

and

$$\hat{X}_3 = \sum_{i=1}^{n} \hat{Y}_i + \hat{Y}_{n+2}.$$

The conditions involved in Definition 3.4 are obviously satisfied by this construction. In particulr, $\hat{X}_3 \stackrel{\mathrm{d}}{=} X(n+1)$, since the $\hat{Y}_i$'s are iid; and $\hat{X}_1 + \hat{X}_4 = \hat{X}_2 + \hat{X}_3$. It is also clear that $X(n)$ is *not* convex or concave in n, in the sense of Definition 3.5.

Example 3.11 Continue with the above example. Suppose Y_i are independent, but not identically distributed. Specifically, suppose $Y_i \leq_{st} Y_{i+1}$ for all i. Then, $\{X(n)\} \in$ SICX-sp. To show this, modify the construction of $\hat{X}_3$ in Example 3.10: replace $\hat{Y}_{n+2}$ by $\hat{Y}'_{n+1}$, which is sampled from the distribution of Y_{n+1}, but using the uniform variate U_{n+2} that generates Y_{n+2}. Since $Y_{n+1} \leq_{st} Y_{n+2}$, we have $\hat{Y}'_{n+1} \leq \hat{Y}_{n+2}$. Hence, $\hat{X}_1 + \hat{X}_4 \geq \hat{X}_2 + \hat{X}_3$. Similarly, if $Y_i \geq_{st} Y_{i+1}$ for all i, then $\{X(n)\} \in$ SICV-sp.

Combining the two examples in Example 3.8 and Example 3.10, we have a useful model for manufacturing processes that have a random yield. Let $Y_i(p)$ be a set of iid Bernoulli random variables parameterized by p. Suppose p is the yield ratio. Then Y_i represents whether the i th part that is processed is good ($Y_i = 1$) or bad ($Y_i = 0$). Let n be the lot size; then $X(n)$ gives the number of good parts (i.e., "yield") that come out of this lot. Write X explicitly as $X(n,p)$. The SICX-sp definition in Definition 3.4 also motivates the development of the notions of stochastic *supermodularity* and *submodularity*, carried out in the next subsection.

3.2 Stochastic Supermodularity and Submodularity

Definition 3.12 A family of random variables, $\{X(s,t)\}$, where s and t are real or integer valued scalars, is stochastically increasing and supermodular (resp. increasing and submodular) in (s,t) if for any given $s_1 \leq s_2$, $t_1 \leq t_2$, there exist on a common probability space $(\Omega, \mathcal{F}, P)$ four random variables $\hat{X}_i$, $i = 1,2,3,4$, that are equal in distribution to $X(s_1,t_1), X(s_1,t_2), X(s_2,t_1)$, and $X(s_2,t_2)$, respectively, and for all $\omega \in \Omega$,

$$\hat{X}_4(\omega) \geq \max\{\hat{X}_2(\omega), \hat{X}_3(\omega)\}; \quad \hat{X}_1(\omega) + \hat{X}_4(\omega) \geq \text{(resp. } \leq) \hat{X}_2(\omega) + \hat{X}_3(\omega).$$

Stochastically decreasing and supermodular (decreasing and submodular) is similarly defined: replace $\hat{X}_4$ in the first inequality above by $\hat{X}_1$.

Remark 3.13 (i) Recall, in the deterministic setting, a bivariate function $\phi(x,y)$ is supermodular (resp.) submodular, if for any $x_1 \leq x_2$, $y_1 \leq y_2$,

$$\phi(x_1,y_1) + \phi(x_2,y_2) \geq \text{(resp. } \leq) \phi(x_1,y_2) + \phi(x_2,y_1).$$

Hence, the stochastic version above simply states that the sample realization of X (i.e., for any given ω) behaves as a (deterministic) supermodular or submodular function. In particular, Definition 3.12 implies that $EX(s,t)$ is increasing and supermodular or submodular in (s,t). Also notice that supermodularity or submodularity is a four-point definition (in contrast to convexity or concavity). Hence, there would be no weak and strong versions analogous to Definitions 3.4 and 3.5 in the convexity/concavity setting.

(ii) Two useful properties of a deterministic supermodular or submodular function are worth noting. When ϕ is twice differentiable, supermodularity (resp. submodularity) is equivalent to the cross partial derivative $\phi_{xy} \geq 0$ (resp. ≤ 0). (Hence, supermodularity and submodularity are quite different properties from convexity and concavity.) To maximize a supermodular $\phi(x,y)$, suppose $y_1 < y_2$, and let x_1^* be the x that maximizes $\phi(x,y_1)$ given y_1. That is, $\phi(x_1^*,y_1) \geq \phi(x,y_1)$ for all x. Similarly, let x_2^* be the x that maximizes $\phi(x,y_2)$ given y_2. We then have the "isotone" property: $x_1^* \leq x_2^*$. In other words, any $x_2 < x_1^*$ cannot be a candidate to maximize $\phi(x,y_2)$. This is simply because the supermodularity implies

$$\phi(x_2,y_2) - \phi(x_1^*,y_2) \leq \phi(x_2,y_1) - \phi(x_1^*,y_1).$$

But the RHS above is ≤ 0, since x_1^* maximizes $\phi(x,y_1)$. Hence, $\phi(x_2,y_2) \leq \phi(x_1^*,y_2)$, for any $x_2 < x_1^*$. This isotone property has useful applications in maximizing a supermodular function. A similar isotone property exists in minimizing a submodular function. [Maximizing submodular functions or minimizing supermodular functions lead to "antitone" properties. For more details in a general lattice setting, refer to Topkis [33].]

Example 3.14 Continue with the random yield model described earlier. Following Definition 3.12, it is easy to verify that $X(n,p) = \sum_{i=1}^{n} Y_i(p)$, where $Y_i(p)$ are iid Bernoulli random variables with parameter p, is stochastically supermodular in (n,p). In fact, the Bernoulli distribution here is not important; supermodularity holds as long as for each i, $Y_i(p)$ is stochastically increasing in p. This supermodularity, as well as the SIL-sp property in n and in p established earlier, will become useful when we study below an assembly operation with random component yield.

We can now combine stochastic supermodularity with stochastic convexity into a new property: *directional convexity.*

Definition 3.15 A (deterministic) bivariate function $\phi(x,y)$ is termed "directionally convex " (resp. directionally concave) if it satisfies the following property:

$$\phi(z^1) + \phi(z^4) \geq \text{ (resp. } \leq) \; \phi(z^2) + \phi(z^3),$$

for any $z^i = (x_i, y_i)$, $i = 1, 2, 3, 4$, such that

$$x_1 \leq x_2, x_3 \leq x_4, \quad y_1 \leq y_2, y_3 \leq y_4,$$

and

$$x_1 + x_4 = x_2 + x_3, \quad y_1 + y_4 = y_2 + y_3.$$

Lemma 3.16 The bivariate function $\phi(x, y)$ is directionally convex (resp. directionally concave) in (x, y) if and only if $\phi(x, y)$ is convex (resp. concave) in x and in y (individually), and supermodular (resp. submodular) in (x, y).

Proof. We prove the equivalence between directional convexity and (individual) convexity plus supermodularity, the other case being completely analogous. Note the four points (z^i) in Definition 3.15 have the following relation: z^1 and z^4 are at the southwest and northeast corners of a rectangle, while z^2 and z^3 lie within the rectangle. Hence directional convexity implies supermodularity by simply letting z^2 and z^3 coincide with the other two corner points, and convexity in each variable follows by reducing (collapsing) the rectangle to one of its sides, parallel to either the horizontal or the vertical axis.

For the converse, let (x_i, y_i), $i = 1, 2, 3, 4$, be the four points in Definition 3.12. Then,

$$\begin{aligned}
\phi(x_4, y_4) - \phi(x_3, y_3) &= \phi(x_4, y_4) - \phi(x_4, y_3) + \phi(x_4, y_3) - \phi(x_3, y_3) \\
&\geq \phi(x_4, y_2) - \phi(x_4, y_1) + \phi(x_4, y_1) - \phi(x_3, y_1) \\
&\geq \phi(x_2, y_2) - \phi(x_2, y_1) + \phi(x_2, y_1) - \phi(x_2, y_1) \\
&= \phi(x_2, y_2) - \phi(x_2, y_1),
\end{aligned}$$

where the first inequality makes use of the convexity in y and the supermodularity, while the second inequality makes use of the convexity in x as well as supermodularity. □

Based on Lemma 3.16, combining Defintion 3.12 with the earlier definitions of stochastic convexity, we have:

Theorem 3.17 If the family of random variables $\{X(s, t)\}$ is stochastically increasing and supermodular (resp. increasing and submodular) in (s, t), and in addition, increasing and convex (resp. increasing and concave) in s and in t, following any one of Definitions 3.4, 3.5 and 3.6, then $EX(s, t)$ is increasing and directionally convex (resp. increasing and directionally concave) in (s, t).

Based on Definition 3.15, one can define stochastic versions of directional convexity. There can be several versions, corresponding to the different versions of stochastic convexity. Below we list one.

Definition 3.18 A family of random variables, $\{X(s,t)\}$, where s and t are real or integer valued scalars, is stochastically increasing and directionally convex in (s,t), in the *sp* (sample-path) sense, if for any four given points $z^i = (s_i, t_i)$, $i = 1,2,3,4$, specified in Defintion 3.15, there exist on a common probability space $(\Omega, \mathcal{F}, P)$ four random variables $\hat{X}_i$ that are equal in distribution to $X(z^i)$, $i = 1,2,3,4$, respectively, and for all $\omega \in \Omega$,

$$\hat{X}_4(\omega) \geq \max\{\hat{X}_2(\omega), \hat{X}_3(\omega)\}; \quad \hat{X}_1(\omega) + \hat{X}_4(\omega) \geq \text{(resp. } \leq) \hat{X}_2(\omega) + \hat{X}_3(\omega).$$

Remark 3.19 Obviously, if $\{X(s,t)\}$ is stochastically increasing and directionally convex in (s,t), following Defintion 3.18, then it is stochastically increasing and convex in s and in t (in the *sp* sense), and stochastically increasing and supermodular in (s,t). The converse need not hold in general. However, following Theorem 3.17, the converse does hold for *expectation*; and this is often has useful applications, as we shall demonstrate below.

3.3 A Joint Setup Problem

Here we study the problem of jointly setting up a set of production facilities. We will model the state (work in process) at each facility as a pure death process with state-dependent death rate (production rate). To this end, we start with applying the earlier results to Markov processes.

Let $\{X_n(s)\}$ be a discrete-time homogeneous Markov chain with state space on the non-negative integers, and initial state s, i.e., $X_0 = s$. Let $Y(x)$ denote the generic random variable that has the same distribution as the position taken by the Markov chain in one transition starting from x; that is, $Y(x)$ equals in distribution to X_{n+1} given $X_n = x$. Not surprisingly, the spatial and temporal behavior of this homogeneous Markov chain is completely characterized by $Y(x)$. For instance, clearly, if $Y(x) \geq x$, then X_n is stochastically increasing in n. Furthermore, if $\{Y(x) - x\}$ is stochastically increasing in x (note this implies $Y(x) \geq x$), then, for any $s_1 \leq s_2$ (using "hat" to denote sample values),

$$X_0(s_1) + X_1(s_2) \stackrel{\mathrm{d}}{=} s_1 + \hat{Y}(s_2) \geq s_2 + \hat{Y}(s_1) \stackrel{\mathrm{d}}{=} X_0(s_2) + X_1(s_1).$$

Inductively, we can then show that $X_n(s)$ is supermodular in (n,s). Under the same conditions, we also have

$$\hat{X}_2 - \hat{X}_1 = \hat{Y}'(\hat{Y}(s)) - \hat{Y}(s) \geq \hat{Y}'(s) - s = \hat{X}'_1 - \hat{X}_0,$$

where $\hat{Y}$ and $\hat{Y}'$ denote the sample values of the Y's that correspond to the two transitions, $\hat{X}_1 \stackrel{\mathrm{d}}{=} \hat{X}'_1 \stackrel{\mathrm{d}}{=} X_1$, and the inequality is due to the increasingness of $\{Y(x) - x\}$, taking into account $\hat{Y}(s) \geq s$. Hence inductively, $\{X_n(s)\}$ satisfies SICX-sp in n (for any given s). Finally,

$\{Y(x)\}$ satisfying SICX-sp implies $\{X_1(s)\} \in$ SICX-sp. Hence, for any four initial states, s_i, $i = 1, 2, 3, 4$, satifying

$$s_4 \geq \max\{s_2, s_3\}, \quad s_1 + s_4 = s_2 + s_3,$$

we can generate, on a common probability space, $\hat{X}^i$, $i = 1, 2, 3, 4$, such that $\hat{X}^i \stackrel{\mathrm{d}}{=} X_1(s_i)$, and

$$\hat{X}^4 \geq \max\{\hat{X}^2, \hat{X}^3\}, \quad \hat{X}^1 + \hat{X}^4 \geq \hat{X}^2 + \hat{X}^3.$$

Making use of the SICX-sp property of $Y(x)$ again, we can generate, on a common probability space, $\hat{Y}^i$, $i = 1, 2, 3, 4$, such that $\hat{Y}^i \stackrel{\mathrm{d}}{=} Y(\hat{X}^i)$, and

$$\hat{Y}^4 \geq \max\{\hat{Y}^2, \hat{Y}^3\}, \quad \hat{Y}^1 + \hat{Y}^4 \geq \hat{Y}^2 + \hat{Y}^3.$$

But $\hat{Y}^i \stackrel{\mathrm{d}}{=} X_2(s_i)$; hence, $\{X_2(s)\} \in$ SICX-sp. This way, we can show that $\{X_n(s)\} \in$ SICX-sp in s for any given n. To summarize, we have

Theorem 3.20 If $\{Y(x) - x\}$ is stochastically increasing in x, and $\{Y(x)\}$ satisfies SICX-sp, then the Markov chain $\{X_n(s)\}$ satisfies SICX-sp in n and in s, and is stochastically increasing and supermodular in (n, s). In particular, $EX_n(s)$ is directionally increasing and convex in (n, s).

As an application of Theorem 3.20, consider a continuous-time pure death process that starts at the initial state $X_0 = s$. Later, we shall use this as a basic model for a production process.

Theorem 3.21 $\{X_t(s)\}$ is a pure death process with initial state $X_0 = s$ and death rate $\mu(x)$ given $X_t = x$. If $\mu(x)$ is increasing and concave in x, with $\mu(0) = 0$ and $\mu(x) > 0$ for all $x > 0$, then $\{X_t(s)\}$ satisfies SICX-sp in s and in $-t$, and is stochastically increasing and supermodular in $(s, -t)$. In particular, $EX_t(s)$ is directionally increasing and convex in $(s, -t)$ [or, equivalently, increasing and convex in s, decreasing and convex in t, and submodular in (s, t).]

Proof. We use uniformization to generate the sample path of the pure death process. Let $\eta = \mu(s)$ be the uniformization constant. [Note $\eta \geq \mu(x)$ for all $x \leq s$, due to the increasingness of $\mu(x)$.] Let $0 = \tau_0 < \tau_1 < \ldots < \tau_n < \ldots$ be a sequence of event epochs associated with the Poisson process with rate η. Below, we construct a discrete-time Markov chain, $\{X_n(s)\}$, with initial state s, and use it to generate the path of the death process $\hat{X}_t$ by setting, for $n = 0, 1, 2, \ldots$,

$$\hat{X}_{\tau_n} = X_n, \quad \hat{X}_t = X_n, \quad t \in [\tau_n, \tau_n + 1).$$

In order for the path generated to follow the same probability law as the original death process, the $Y(x)$ that associates with $\{X_n(s)\}$ is specified as follows:

$$Y(x) = x - \mathbf{1}[\eta U \leq \mu(x)]. \tag{7}$$

We want to show it satisfies the conditions similar to those listed in Theorem 3.20; specifically, (i) $\{x - Y(x)\}$ is stochastically increasing in x, and (ii) $\{Y(x)\}$ satisfies SICX-sp.

Since (i) follows immediately from (7), we show (ii). Pick x_i, $i = 1,2,3,4$, such that

$$x_1 \leq x_2, x_3 \leq x_4; \quad x_1 + x_4 = x_2 + x_3.$$

Let

$$\hat{Y}_i = x_i - \mathbf{1}[\eta U \leq \mu(x_i)] \quad i = 1,2,4;$$

$$\hat{Y}_3 = x_3 - \mathbf{1}[\eta U \leq \mu(x_1)] - \mathbf{1}[\mu(x_2) \leq \eta U \leq \mu(x_2) + \mu(x_3) - \mu(x_1)].$$

Then clearly $\hat{Y}_i \stackrel{\mathrm{d}}{=} Y(x_i)$ for $i = 1,2,3,4$, and $\hat{Y}_1, \hat{Y}_2, \hat{Y}_3 \leq \hat{Y}_4$. Furthermore, $\hat{Y}_1 + \hat{Y}_4 \geq \hat{Y}_2 + \hat{Y}_3$ is then equivalent to

$$\begin{aligned}
\mathbf{1}[\eta U \leq \mu(x_4)] &\leq \mathbf{1}[\eta U \leq \mu(x_2)] + \mathbf{1}[\mu(x_2) \leq \eta U \leq \mu(x_2) + \mu(x_3) - \mu(x_1)] \\
&= \mathbf{1}[\eta U \leq \mu(x_2) + \mu(x_3) - \mu(x_1)],
\end{aligned}$$

which follows immediately from the concavity of $\mu(x)$, i.e., $\mu(x_1) + \mu(x_4) \leq \mu(x_2) + \mu(x_3)$. $\square$

Corollary 3.22 When $\{X_t(s)\}$ is the death process in Theorem 3.21, $\{s - X_t(s)\}$ is a birth process that satisfies SICV-sp in s and in t, and stochastically supermodular in (s,t).

Proof. The birth process in question, denoted $Z(t)$, starts at 0, increases over time, and eventually reaches s and stays at s. The corresponding one-step transition [cf. (7)] is characterized by

$$Y(z) = z + \mathbf{1}[\eta U \leq \mu(s - z)].$$

Similar to the proof of Theorem 3.21, we can show that $\{Y(z) - z\}$ is stochastically decreasing in z and $\{Y(z)\}$ satisfies SICV-sp. The desired properties then follow, much similar to the way the conclusions in Theorem 3.20 are reached. $\square$

Example 3.23 Consider a set of production facilities indexed by $i = 1,2,...,m$. From time to time (e.g., a shift) a joint setup is performed for all m facilities: a number of jobs are prepared (pre-processed) and released into each facility for processing. The state of facility i at time t is defined as the number of jobs that have not been finished — work in process, or *WIP*; denote this state as $X_t^i(x_i)$, where $x_i := X_0^i$ is the initial state, i.e., the number of jobs immediately after the setup. Suppose for all i, $\{X_t^i\}$ is modeled as a pure death process with death rate $\mu^i(x)$ when the state is in x. Suppose for all i, $\mu^i(0) = 0$, $x > 0 \Rightarrow \mu^i(x) > 0$, and that $\mu^i(x)$ is increasing and concave in x. This last assumption appears appropriate for most production facilities, where the production rate increases as the WIP level increases, but gradually flattens out as the system capacity approaches saturation.

Suppose that the facilities are jointly set up at time zero and are then run for a given length of t units of time. The decision at time zero is to determine the number of jobs that are initially allocated to each facility, $x := (x_i)_{i=1}^m$. For each facility i, there is a profit earned from the total number of completed jobs by time t, $g_i(x_i - X_t^i(x_i))$, a penalty cost for the number of unfinished jobs (if any) at time t, $h_i(X_t^i(x_i))$, and a cost for the setup, $c_i(x_i)$; here g_i, h_i and c_i are all real-valued functions. Let

$$\phi_i(x_i) := Eg_i(x_i - X_t^i(x_i)) - Eh_i(X_t^i(x_i)) - c_i(x_i), \quad i = 1, ..., m.$$

Then $\phi_i(x_i)$ is the expected net profit (over the t time units) from facility i, given its initial allocation x_i. We want to solve the following optimization problem:

$$\max_{x \le N, |x| \le R} \phi(x) := \sum_{i=1}^{m} \phi_i(x_i), \tag{8}$$

where $N := (N_i)_{i=1}^m$; N_i is a given upper limit (integer) on x_i, reflecting the capacity at facility i; and R the (given) maximal total number of jobs to be allocated.

Proposition 3.24 If for all $i = 1, ..., m$, $g_i(x_i)$ is increasing and concave, while $h_i(x_i)$ and $c_i(x_i)$ increasing and convex, then the objective function in (3.9a), $\phi(x)$, is (jointly) concave in x.

Proof. From Theorem 3.21 and Corollary 3.22, $\{X_t^i(x_i)\}$ is stochastically increasing and convex in x_i, and $\{x_i - X_t^i(x_i)\}$ stochastically increasing and concave in x_i, both in the *sp* sense. Under the given conditions, it is then easy to see that $\phi_i(x_i)$ is concave in x_i. Joint concavity then follows from the separable form of $\phi(x)$. □

Remark 3.25 Proposition 3.24 implies that the optimal solution to (8) can be generated using a greedy type of marginal allocation algorithm (Fox [7]). The algorithm runs as follows. Start from $x^0 = \mathbf{0}$. Suppose x^k is the allocation at the k th iteration of the algorithm. Set $x^{k+1} = x^k + e_j$, where e_j is the j th unit vector, and

$$j = \arg\max_i \{\Delta_i^k := \phi_i(x_i^k + 1) - \phi_i(x_i^k) | x_i^k < N_i, \, \Delta_i^k > 0\}.$$

Once $x_i^k = N_i$ or $\Delta_i^k \leq 0$, eliminate facility i from further allocation. The algorithm stops when all facilities are eliminated or all R jobs are allocated, whichever occurs first.

Example 3.26 Continue with Example 3.23. Suppose the time length t is also a decision variable (for instance, t is the time between successive setups). For simplicity, assume that t is also discrete (e.g., number of minutes or hours), and $t \leq T$, where T is a given integer. Write the expected net profit functions as $\phi_i(x_i, t)$. The optimization problem is then

$$\max_{(x,t)\leq(N,T),|x|\leq R} \phi(x,t) := \sum_{i=1}^{m} \phi_i(x_i, t). \tag{9}$$

Proposition 3.27 Under the conditions of Proposition 3.24, if in addition assume a linear profit function g_i for facility i, then $\phi_i(x_i, t)$ is concave in t (in addition to concave in x_i), and supermodular in (x_i, t) for all i.

Proof. From Theorem 3.21, in the stochastic *sp* sense, $\{X_t^i(x_i)\}$ is increasing and convex in x_i decreasing and convex in t, and submodular in (x_i, t). Since $h_i(\cdot)$ is increasing and convex, it is easy to verify that $Eh_i(X_t^i(x_i))$ is convex in x_i and in t, and submodular in (x_i, t). Similarly, from Corollary 3.22, in the stochastic *sp* sense, $\{x_i - X_t^i(x_i)\}$ is increasing and concave in x_i and in t, and supermodular in (x_i, t). It is easy to verify that $Eg_i(x_i - X_t^i(x_i))$ is concave in x_i and in t, and supermodular in (x_i, t), taking into account that g_i is a linear function. The stated properties of ϕ_i then follow. □

Remark 3.28 To find the optimal solution to (9), for each $t \leq T$, use the marginal allocation to obtain the optimal $x^*(t)$. The final optimal solution is then obtained by comparing these $x^*(t)$, $t \leq T$. The supermodularity in Proposition 3.27 implies the isotone property [cf. Remark 3.13(ii)]: $x^*(t)$ is increasing in t. Hence, the marginal allocation for $t+1$ can start from $x^*(t)$ (instead of $\mathbf{0}$ as in Remark 3.25).

3.4 Production with Trial Runs

Example 3.29 We study a manufacturing process that involves trial runs. A job to be processed consists of J units. To process the job, a setup is required. Although the entirety of the job can be set up and processed in a single batch, it is not desirable to do so, since the setup is prone to failure, resulting in loss of units. (So, if all J units are set up in a single batch, they would all be lost if the setup fails.) On the other hand, if a setup is good, then the next setup is guaranteed to be good as well, hence the remaining units can be set up and processed in a single batch. (The remaining process is routine and reliable.) Therefore, the common practice is to conduct trial runs of the setup using small batches until a successful setup is performed. This models, for instance, the photolithographic process in semiconductor wafer fabrication. There, the manual allignment of the mask, which exposes the photoresist on the wafer to form the designed circuitry, is known to be such a failure-prone setup.

The time required for each setup is usually deterministic and independent of the number of units involved (batch size). Hence, without loss of generality, assume each setup takes one unit of time. The time for processing the batch that follows the j th setup is random and dependent on the batch size; denote this as $S(x_j)$, given the batch size is x_j, $1 \leq x_j \leq J$. To capture the economies of scale, $S(x_j)$ is assumed to be SICV-sp in x_j. That is, it increases with the batch size, but with decreasing marginals. For instance, suppose $S(x_j) = \xi_1 + \cdots + \xi_{x_j}$, where ξ_i's are independent random variables that are stochastically decreasing in i [cf. Example (3.11)].

Let $\mathbf{x} := (x_j)_{j=1}^k$ be a k-vector of positive integers, where $k \leq J-1$ and $|\mathbf{x}| := \sum_{j=1}^k x_j \leq J-1$. We are interested in studying a control policy that prescribes up to a maximum number of k trials, with the j th trial involving a batch of x_j units, $j \leq k$. If all k trials fail, the remaining $J - |\mathbf{x}|$ units are set up and processed as a single batch. Refer to this as the $k(\mathbf{x})$ policy.

Let $p_j > 0$ be the probability of success for the j th setup, let $q_j := 1 - p_j$, $j = 1, ..., J$. Let $\beta_0 := 1$ and $\beta_j := q_1 ... q_j$, $j = 1, ..., J$. Let $G(\mathbf{x})$ and $B(\mathbf{x})$ denote the number of good and bad units $[G(\mathbf{x}) = J - B(\mathbf{x})]$, let $N(\mathbf{x})$ and $T(\mathbf{x})$ denote the number of setups and the total time required to complete the whole job (all J units). We have ("w.p." below stands for "with probability"),

$$B(\mathbf{x}) = x_1 + \cdots + x_j \quad \text{w.p.} \quad \beta_j p_{j+1}, \quad j = 0, 1, ..., k; \quad B(\mathbf{x}) = J \quad \text{w.p.} \quad \beta_{k+1}; \tag{10}$$

(define $B = 0$ when $j = 0$ above)

$$N(\mathbf{x}) = j \quad \text{w.p.} \quad \beta_{j-1} p_j, \quad j = 1, ..., k; \quad N(\mathbf{x}) = k+1 \quad \text{w.p.} \quad \beta_k; \tag{11}$$

and

$$\begin{aligned} T(\mathbf{x}) &= j + \sum_{i=1}^{j} S(x_i) + S(J - x_1 - \cdots - x_j) \quad \text{w.p.} \quad \beta_{j-1} p_j, \quad j = 1, ..., k; \\ T(\mathbf{x}) &= k + 1 + \sum_{i=1}^{k} S(x_i) + S(J - x_1 - \cdots - x_k) \quad \text{w.p.} \quad \beta_k. \end{aligned} \tag{12}$$

Below, we show that given k, the maximum number of trial runs, the optimal policy is to set $x_i = 1$ for all i, i.e., to have single-unit batches for the trial runs. This will be denoted as the $k(\underline{1})$ policy.

Proposition 3.30 Given k, the $k(\underline{1})$-setup policy
(i) stochastically maximizes the number of good units (G); and
(ii) minimizes the expected completion time (ET), provided the expected processing time $ES(x)$ is concave in x.

Proof. (i) We show that $B(\mathbf{x})$ $[= J - G(\mathbf{x})]$ is stochastically minimized by letting $\mathbf{x} = \underline{1}$. Specifically, we generate on a common probability space the sample values of $B(\underline{1})$ and $B(\mathbf{x})$ (for any given $\mathbf{x}$) such that the former is dominated by the latter. To this end, divide the interval [0,1] into a set of $k+2$ segments with length $\beta_j p_{j+1}$ for the j th segment, $j = 0, 1, ..., k$, and β_{k+1} for the last segment [cf. (10)]. Generate a uniform variate $U \in$[0,1]. If U falls into the j th segment, $j \le k+1$, then $\hat{B}(\underline{1}) = j$ while $\hat{B}(\mathbf{x}) = x_1 + ... + x_j \ge j$. If U falls into the last segment, then $\hat{B}(\underline{1}) = \hat{B}(\mathbf{x}) = J$. In any case we have $\hat{B}(\underline{1}) \le \hat{B}(\mathbf{x})$, and hence the desired result.
(ii) Obivious: $ES(x)$ concave implies

$$\sum_{i=1}^{k} ES(1) + ES(J-k) \le \sum_{i=1}^{k} ES(x_i) + S(J - x_1 - \cdots - x_k),$$

for any $\mathbf{x} = (s_i)_{i=1}^k$. □

Next, we study properties of the $k(\underline{1})$ policy with respect to k.

Proposition 3.31 Under the $k(\underline{1})$-setup policy, the number of good units, $G(k)$, is stochastically increasing and concave in k, provided $p_j \ge q_j p_{j+1}$ for all $j = 1, ..., J-1$.

Proof. Equivalently, we show that the number of bad units, denoted as $B(k)$, is stochastically decreasing and convex in k. Following Definition 3.4, we want to construct on a common probability space one replica each for $\hat{B}(k-1)$ and $\hat{B}(k+1)$, and two replicas for $\hat{B}(k)$ [denoting the second one as $\hat{B}'(k)$], such that

$$\hat{B}(k-1) + \hat{B}(k+1) \ge \hat{B}(k) + \hat{B}'(k) \tag{13}$$

is satisfied.

We use a common uniform variate $U \in$[0,1] to generate all four replicas. For $\hat{B}(k-1)$, the interval [0,1] is divided into a sequence of segments of length $\beta_j p_{j+1}$ for the j th segment, $j = 0, 1, ..., k-1$, and β_k for the last segment [cf. (10)]. Suppose U falls into the jth segment, then $\hat{B}(k-1) = j$ if $j \le k-1$, and $\hat{B}(k-1) = J$ if $j = k$ (the last segment). For $\hat{B}(k)$, $\hat{B}'(k)$ and $\hat{B}(k+1)$, follow the same construction, except in the last segment (i.e., the one with length β_k). For $\hat{B}(k)$, this segment is further divided into two sub-segments: the first one has a length

$\beta_k p_{k+1}$, and the second one, β_{k+1}; and $\hat{B}(k) = k$ if U falls into the first one, while $\hat{B}(k) = J$ if U falls into the second one. Refer to Figure 1. For $\hat{B}'(k)$, interchange the two sub-segments. That is, the first one has a length β_{k+1} and corresponds to $\hat{B}'(k) = J$, while the second one has a length $\beta_k p_{k+1}$ and corresponds to $\hat{B}'(k) = k$. To construct $\hat{B}(k+1)$, the last segment is divided into three sub-segments: the first one of length $\beta_{k+1} p_{k+2}$; the second one, β_{k+2}; and the third one, $\beta_k p_{k+1}$; and $\hat{B}(K+1)$ is set at $k+1$, J, and k, respectively, when U falls into one of the three sub-segments. (Again, refer to Figure 1.)

Note that the given condition $p_{k+1} \geq q_{k+1} p_{k+2}$ implies $\beta_k p_{k+1} \geq \beta_{k+1} p_{k+2}$, i.e., the first sub-segment of $\hat{B}(k)$ is larger than that of $\hat{B}(k+1)$. This condition is crucial: it enables the above construction to rule out the possibility that both $\hat{B}(k-1)$ and $\hat{B}(k+1)$ take value J, while only one of $\hat{B}(k)$ and $\hat{B}'(k)$ equals J. Hence, the construction ensures that (13) is always satisfied.

The stochastic decreasing property of $B(k)$ is easy: just compare the replicas of $B(k-1)$ and $B(k)$ as constructed above. □

Proposition 3.32 Under the $k(\underline{1})$-setup policy, the completion time of the job, $T(k)$ is stochastically increasing and concave in k, provided the service times $S(x)$ are stochastically increasing and concave in x.

Proof. Follow the construction in the above proof. The situation here is, however, much more straightforward: all the segments just follow their natural order, no switching is needed [for instance, $\hat{T}(k)$ and $\hat{T}'(k)$ follow exactly the same construction). □

With a suitable objective function, the concavity properties with respect to k are useful in identifying the optimal number of trial runs to conduct, say k^*. It can then be shown that the overall optimal policy is the $k^*(\underline{1})$ policy.

4 Arrangement and Likelihood Ratio Orderings

The purpose of this section is two-fold: (a) to introduce the *arrangement* and *majorization* orderings, and the classes of functions that preserve these orderings, namely, *arrangement increasing* functions and *Schur convex/concave* functions; and (b) to relate these deterministic orderings to the likelihood ratio ordering of random variables. Specifically, arrangement increasing functions constitute a defining class for the likelihood ratio ordering (as well as for the arrangement ordering). The primary application example in this section is the closed Jackson network.

4.1 The Connection

Definition 4.1 Let $x = (x_i)_{i=1}^n$ and $y = (y_i)_{i=1}^n$ be two n-vectors of real-valued components. The arrangement ordering, denoted $x \geq_A y$, is defined as follows: x can be obtained from y through successive pairwise interchanges of its components, with each interchange resulting in a decreasing order of the two interchanged components. A function $\phi : \Re^n \mapsto \Re$ that preserves the arrangement ordering is termed an "arrangement increasing function"; that is, ϕ is arrangement increasing if $\phi(x) \geq \phi(y)$ for any $x \geq_A y$.

For instance, $(4, 5, 3, 1) \geq_A (4, 3, 5, 1) \geq_A (4, 1, 5, 3)$. Note $x \geq_A y$ implies that x is necessarily a permutation of the components of y.

The arrangement ordering is easy to motivate. The inequality in (6) is such an ordering, and so is the scheduling model in Example 2.7, as well as other similar pairwise interchange arguments. Further more, in §2, the defining class of functions for the likelihood ratio ordering, $\mathbf{C}_{\ell r}$ [cf. Theorem 2.3], clearly belongs to the class of arrangement increasing functions. Indeed, based on Definition 4.1, we can extend Theorem 2.3 to the multivariate case as follows.

Theorem 4.2 Let $X = (X_i)_{i=1}^n$ be a random vector of independent components. Let π and σ denote two permutations of the n integers $\{1, 2, ..., n\}$. Let X^π (X^σ) denote a random vector obtained through rearranging the components of X according to the permutation π (σ). Define $\pi \geq_A \sigma$ following the same definition for vectors. Then, $X_n \geq_{\ell r} ... \geq_{\ell r} X_1$ if and only if $E\phi(X^\pi) \geq E\phi(X^\sigma)$ for all $\pi \geq_A \sigma$ and all arrangement increasing functions ϕ.

The above in fact also suggests a likelihood ratio ordering among random variables that are not necessarily independent.

Definition 4.3 Let X be the random vector in Theorem 4.2, but the components are not necessarily independent. Let π and σ be the permutations in Theorem 4.2. Define a likelihood ratio ordering among the components of X, denoted $X_n \geq_{\ell r:j} ... \geq_{\ell r:j} X_1$, if $E\phi(X^\pi) \geq E\phi(X^\sigma)$ for all $\pi \geq_A \sigma$ and all arrangement increasing functions ϕ.

Similar to the proof of Theorem 2.3, we can establish the following

Theorem 4.4 The components of X are ordered $X_n \geq_{\ell r:j} ... \geq_{\ell r:j} X_1$, if and only if the joint density function (or pmf) of X is an arrangement increasing function.

Obviously, when the components are independent, the joint ordering $\geq_{\ell r:j}$ reduces to the regular ordering $\geq_{\ell r}$, since the joint density function factorizes under independence, recovering the original definition for $\geq_{\ell r}$. However, in the general case, $\geq_{\ell r:j}$ and $\geq_{\ell r}$ do not imply each other, as we shall demonstrate through an example below. (Note that in the dependent case, the ordering $\geq_{\ell r}$ is still well defined, although it is concerned with the marginals only. In contrast, $\geq_{\ell r:j}$ is concerned with the joint distribution.)

Similar joint orderings can also be defined for the other stochastic order relations discussed in §2. The defining functional classes obviously have to be subsets of the arrangement increasing functions.

Closely related to the arrangement ordering is majorization, and the class of Schur convex functions that preserve majorization.

Definition 4.5 Two n-vectors, x and y, are ordered under majorization, denoted as $x \leq_m y$, if

$$\sum_{j=1}^{i} x_{[j]} \leq \sum_{j=1}^{i} y_{[j]},\ i = 1, ..., n-1; \quad \sum_{j=1}^{n} x_{[j]} = \sum_{j=1}^{n} y_{[j]};$$

where $x_{[j]}$ and $y_{[j]}$ denote the j th largest components of x and y, respectively. Replacing the last equality above by $\leq$ results in "weak majorization", denoted $x \leq_{wm} y$. A function ϕ is termed Schur convex if for any x and y, $x \leq_m y \Rightarrow \phi(x) \leq \phi(y)$. In addition, $x \leq_{wm} y$ if and only if $\phi(x) \leq \phi(y)$ for any ϕ that is increasing and Schur convex. Finally, a function ϕ is Schur concave if $-\phi$ is Schur convex.

Remark 4.6 (i) From the definition of majorization, we have $x \leq_m x^\pi$ as well as $x \geq_m x^\pi$ for any permutation π. Hence, if a Schur convex/concave function ϕ has a symmetric domain (i.e., x belongs to the domain if and only if x^π belongs to the domain), then ϕ is necessarily symmetric, i.e., $\phi(x) = \phi(x^\pi)$ for any permutation π. Below, we will assume this is indeed the case throughout.
(ii) Suppose a pair of components of x is ordered: $x_i > x_j$. Then obviously, $x \geq_m x - \epsilon e^i + \epsilon e^j$, where e^i and e^j denote the i th and the j th unit vectors, and $0 < \epsilon < x_i - x_j$. In other words, x can be made smaller, in the sense of majorization, by shifting a small, positive amount (ϵ) from one of its larger component (x_i) to a smaller component (x_j) (ϵ should be small enough so that after the shift the larger component is still larger). Call such an operation on the vector x "transposition". Then clearly, $x \leq_m y$ if and only if x can be reached from y through a sequence of such transpositions, each resulting in a vector that is majorized by the vector from the previous transposition.
(iii) There are two implications from the above. One is, to establish the Schur convexity of a function f, we only need to establish the dominance relation between the function values at x

and its transposition. In other words, Schur convexity is essentially a property with respect to each *pair* of components. The other implication is that for a differentiable function f, Schur convexity is equivalent to

$$x_i > x_j \quad \Rightarrow \quad f_i(x) \geq f_j(x),$$

for any $i \neq j$ and any x, where f_i and f_j denote the partial derivatives with respect to x_i and x_j.

4.2 Queueing Network Applications

We now illustrate the arrangement and majorization orderings and the functional classes they induce through a closed queueing network of the Jackson type (Jackson [11], Gordon and Newell [9]), with m stations and N jobs. Consider the network in equilibrium. Let X_i be the number of jobs in station i, and $X = (X_i)_{i=1}^m$. Let $\mu_i(n_i)$ be the (state dependent) service rate at station i, when $X_i = n_i$. Suppose for all i, $\mu_i(0) = 0$ and $\mu_i(n_i) > 0$ for all $n_i > 0$. Let r_{ij} denote the routing probability from node i to node j. The routing matrix $[r_{ij}]$ is stochastic (i.e., each row sums to unity). In this sense, the network is "closed". Physically this can be viewed as N jobs keep circulating among the m stations, no job ever leaves the network, and no job arrives from the outside either. An alternative view is: once a job completes all its processing requirements and leaves the network, another job is immediately released into the network, so that the total job population is maintained at the constant N. Indeed, this second view is consistent with the operation of many manufacturing facilities in which the WIP (work-in-process) level is maintained at a constant.

Let v_i be the solution to

$$v_i = \sum_{j=1}^{m} v_j r_{ij}, \quad i = 1, ..., m. \tag{14}$$

Note since the routing matrix $[r_{ij}]$ is stochastic, the solution to the above linear equation system is only unique to a constant multiplier. Intuitively, v_i can be interpreted as the "visit ratio" to station i, i.e., the frequencies that the stations are visited by jobs. Formally, they are components of the invariant probability vector of the Markov chain that governs job routing among the stations.

Let $Y = (Y_i)_{i=1}^m$ be a vector of *independent* random variables with the following distributions:

$$P[Y_i = n_i] = P[Y_i = 0] v_i^{n_i} / [\mu_i(1) \cdots \mu_i(n_i)] \quad n_i = 1, ..., N; \ i = 1, ..., m. \tag{15}$$

Let $|Y| = \sum_{i=1}^m Y_i$. Let $n = (n_i)_{i=1}^m$ be an m-vector of non-negative integer components, and denote $|n| = \sum_{i=1}^m n_i$. Then, it is not difficult to derive the following relation:

$$P[X = n] = \prod_{i=1}^{m} P[Y_i = n_i]/P[|Y| = N], \tag{16}$$

for all n such that $|n| = N$.

To prove (16) amounts to verifying that it satisfies the equilibrium (or, full) balance equations:

$$\sum_i P[X = n]\mu_i(n_i) = \sum_i \sum_j P[X = n - e^i + e^j]\mu_j(n_j + 1)r_{ji}, \tag{17}$$

where e^i and e^j denote the i th and the j th unit vectors. Note the LHS and the RHS above can be interpreted as "probability flows" out of and into the state n, respectively. The LHS above can also be written as

$$\sum_i \sum_j P[X = n]\mu_i(n_i)v_j r_{ji}/v_i,$$

since, from (14) we have $1 = \sum_j v_j r_{ji}/v_i$. Therefore, it suffices to verify that (16) satisfies the following:

$$P[X = n]\mu_i(n_i)/v_i = P[X = n - e^i + e^j]\mu_j(n_j + 1)/v_j; \tag{18}$$

which is quite obvious, given the "product form" of (16).

Based on (16), it is also easy to derive the throughput at node i:

$$E\mu_i(X_i) = \sum_{n_i}^{N} \mu_i(n_i)P[X_i = n_i] = v_i \frac{P[Y_1 + \cdots + Y_m = N - 1]}{P[Y_1 + \cdots + Y_m = N]}.$$

That is, the throughput at the nodes are proportionate to the visit ratios. Hence, the common, node-independent constant, namely the probability ratio above, is often referred to as the throughput of the closed network, denoted $TH(N)$:

$$TH(N) = \frac{P[Y_1 + \cdots + Y_m = N - 1]}{P[Y_1 + \cdots + Y_m = N]} = E\mu_i(X_i)/v_i. \tag{19}$$

At this point, it would be helpful to introduce the concept of *equilibrium rate* (Shanthikumar and Yao [24, 25]).

Definition 4.7 Let Z be a non-negative, integer-valued random variable, $P[Z = k] > 0$, $k = 0, 1, ..., N$, where N can be finite or infinite. The equilibrium rate of Z, denoted r_Z, is a real-valued function defined as follows:

$$r_Z(0) = 0; \qquad r_Z(k) = P[Z = k - 1]/P[Z = k], \quad k = 1, ..., N.$$

It is easy to see that the equilibrium rate, $\{r_Z(k); k = 0, ..., N\}$, and the pmf of Z, $\{P[Z = k]; k = 0, ..., N\}$, uniquely define each other.

Recall, a discrete radom variable Z is said to have the so-called PF_2 property (Polya frequency of order 2), if its pmf is log-concave, i.e., $\log P[Z = k]$ is concave in k. Refer to, e.g., Keilson and Sumita [12]. It turns out that this PF_2 property and the likelihood ratio ordering can be conveniently represented by equilibrium rates. The lemma below is easily verified from first principle.

Lemma 4.8 (i) $Z \in PF_2 \Leftrightarrow r_Z$ is an increasing function.
(ii) Suppose the equilibrium rats of two random variables, $r_{Z^1}(k)$ and $r_{Z^2}(k)$, are defined on the same set of integers: $k \in \{0, 1, \dots, N\}$. Then, $Z^1 \geq_{\ell r} Z^2 \Leftrightarrow r_{Z^1}(k) \leq r_{Z^2}(k)$, for all k.

Hence, from (15), we can write $r_{Y_i}(n_i) = \mu_i(n_i)/v_i$ for all i. In the same spirit, the throughput function $TH(N)$ can be expressed as the equilibrium rate of the sum of random variables $Y_1 + \cdots + Y_m$ [cf. the first equation in (19)]:

$$TH(N) = r_{Y_1+\cdots+Y_m}(N). \tag{20}$$

Following Lemma 4.8, we have

$$Y_i \geq_{\ell r} Y_j \quad \Leftrightarrow \quad \mu_i(n) \leq \mu_j(n) \quad \forall n, \tag{21}$$

and

$$Y_i \in PF_2 \quad \Leftrightarrow \quad r_{Y_i}(k) \text{ increasing in } k. \tag{22}$$

Lemma 4.9 $Y_1 \geq_{\ell r} ... \geq_{\ell r} Y_m \Rightarrow X_1 \geq_{\ell r:j} ... \geq_{\ell r:j} X_m$.

Proof. Pick any pair of indices $i < j$. For any non-negative integers $x \geq y$, fixing X_k for $k \neq i, j$, while interchanging the values of X_i and X_j in the pmf, we have,

$$P[..., X_i = x, ..., X_j = y, ...]/P[..., X_i = y, ..., X_j = x, ...]$$

$$= P[Y_i = x]P[Y_j = y]/P[Y_i = y]P[Y_j = x] \geq 1,$$

since $Y_i \geq_{\ell r} Y_j$. That is, the joint pmf of X is an arrangement increasing function. The desired conclusion then follows from Theorem 4.4. □

Remark 4.10 Note that although the Y_i's are independent random variables, the X_i's are not, as evident from the joint distribution in (16). Earlier, we mentioned that the orderings

$\geq_{\ell r:j}$ and $\geq_{\ell r}$ do not imply each other in general. Lemma 4.9 now provides such an example; specifically, the result there does not hold for the ordering $\geq_{\ell r}$. To see this, from (16), we can derive the marginal distribution of X_i and then its equilibrium rate:

$$r_{X_i}(n_i) = r_{Y_i}(n_i)/r_{\Sigma_{(i)}}(N+1-n_i), \tag{23}$$

where, $\Sigma_{(i)}$ denotes the sum $Y_1 + \cdots + Y_m$ minus Y_i. A similar expression can be derived for the equilibrium rate of X_j. Hence, in order to have $r_{X_i}(k) \leq r_{X_j}(k)$ for all k, we need not only $Y_i \geq_{\ell r} Y_j$, but also $\Sigma_{(i)} \leq_{\ell r} \Sigma_{(j)}$. That is,

$$\Sigma_{(ij)} + Y_j \leq_{\ell r} \Sigma_{(ij)} + Y_i,$$

where $\Sigma_{(ij)}$ denotes the sum $Y_1 + \cdots + Y_m$ without Y_i and Y_j. Now, it is known (e.g., Keilson and Sumita [12]) that in general to preserve the likelihood-ratio ordering under summation the (independent) random variables involved have to satisfy the PF_2 property. Hence, in order for the above to be implied by $Y_i \geq_{\ell r} Y_j$, we need the PF_2 property of $\Sigma_{(ij)}$. Therefore, for Lemma 4.9 to hold for the ordering $\geq_{\ell r}$, the additional requirement that $Y_i \in PF_2$ for all i is needed. For the application below, however, the result holds without the PF_2 property, thanks to the joint ordering $\geq_{\ell r:j}$.

Example 4.11 Now consider the following allocation problem in the close Jackson network. There are m sites for allocating the m stations. The service rates, $\{\mu_i(n)\}$, and the visit ratios, $\{\nu_i\}$, are known, reflecting the processing requirements at the stations. Therefore, the congestion level (queue length) of the stations will be independent of their locations. However, the sites have different cost structure for accommodating the jobs. (For instance, suppose storage space has to be built for jobs in transition.) Suppose $\pi = (\pi(i))_{i=1}^m$ denotes an allocation with station $\pi(i)$ located at site i. For $x = (x_i)_{i=1}^m$, let $h(x)$ be the holding cost of having x_i jobs at site i, $i = 1, ..., m$. Let X^π denote the equilibrium (joint) queue lengths in the network under allocation π. The problem is to minimize the expected holding cost:

$$\min_\pi Eh(X^\pi). \tag{24}$$

Proposition 4.12 Suppose there exists a permutation π^*, such that for all $n = 1, ..., N$,

$$\mu_{\pi^*(1)}(n)/v_{\pi^*(1)} \leq ... \leq \mu_{\pi^*(m)}(n)/v_{\pi^*(m)}, \tag{25}$$

then π^* is the optimal solution to (24) for any arrangement increasing function h.

Proof. Under condition (25), we have, $Y_{\pi^*(m)} \geq_{\ell r} \dots \geq_{\ell r} Y_{\pi^*(1)}$. Hence, $X_{\pi^*(m)} \geq_{\ell r:j} \dots \geq_{\ell r:j} X_{\pi^*(1)}$ follows from Lemma 4.9. The optimality of π^* then follows from Definition 4.3, since the cost function h is arrangement increasing. $\Box$

Next, we consider another rearrangement result in the closed queueing network. Recall the visit ratios, $(v_i)_{i=1}^m$, are the solution to the equations in (14). To make the solution unique and without loss of generality, we impose a normalization condition: $\sum_{i=1}^m v_i = v$, where $v > 0$ can be any given constant. For convenience and without loss of generality, we will take $v = 1$ below. The visit ratios can also be viewed as the (normalized) amount of workload assigned to the nodes.

Now, suppose the service rate functions at the m nodes, $\mu_i(\cdot)$, $i = 1, \dots, m$ are given; we want to distribute the workload among the nodes, in terms of assigning the visit ratios to the nodes. We start with a two-node network.

Lemma 4.13 Consider a two-node closed Jackson network with N jobs. For $i = 1, 2$, let $\mu_i(\cdot)$ and v_i be, respectively, the service rate function and the visit ratio at node i. Let X_1 and $N - X_1$ be the equilibrium number of jobs at node 1 and node 2, respectively. Change the visit ratios to $v_1 - \Delta$ and $v_2 + \Delta$, where $0 \leq \Delta < v_1$, and index quantities relate to the resulting network by a "$\prime$". Then, $X_1 \geq_{\ell r} X_1'$.

Proof. From (23), we have

$$\begin{aligned} r_{X_1}(n) &= \frac{r_{Y_1}(n)}{r_{Y_2}(N+1-n)} = \frac{\mu_1(n)}{v_1} \cdot \frac{v_2}{\mu_2(N+1-n)} \\ &\leq \frac{\mu_1(n)}{v_1 - \Delta} \cdot \frac{v_2 + \Delta}{\mu_2(N+1-n)} = \frac{r_{Y_1'}(n)}{r_{Y_2'}(N+1-n)} = r_{X_1'}(n). \end{aligned}$$

Hence $X_1 \geq_{\ell r} X_1'$. $\Box$

Lemma 4.14 Consider the two networks in Lemma 4.13. Suppose $v_1 \geq v_2$.
(i) Suppose $\mu_1(n) \geq \mu_2(n)$ for all n, and let $\Delta = v_1 - v_2$ (i.e., the second network is obtained through interchanging the visit ratios at the two nodes in the first network). Then $\min(X_1, N - X_1) \geq_{st} \min(X_1', N - X_1')$.
(ii) Suppose $\mu_1(n) = \mu_2(n)$ for all n, and $0 \leq \Delta < v_1 - v_2$. Then $\min(X_1, N - X_1) \leq_{st} \min(X_1', N - X_1')$.

Proof. Write $\rho = v_1/v_2 \geq 1$, and for $i = 1, 2$, $M_i(0) = 1$, $M_i(n) = \mu_i(1) \cdots \mu_i(n)$ for $n \geq 1$. For (i), we want to show

$$P(X_1 \leq n) + P(X_1 \geq N - n) \leq P(X_1' \leq n) + P(X_1' \geq N - n) \tag{26}$$

for all $n \leq \lfloor \frac{N}{2} \rfloor$, where $\lfloor \frac{N}{2} \rfloor$ denotes the integer part of $N/2$. The above can be written explicitly as:

$$\frac{\sum_{k=0}^{n} \frac{\rho^k}{M_1(k)M_2(N-k)} + \frac{\rho^{N-k}}{M_1(N-k)M_2(k)}}{\sum_{k=0}^{\lfloor \frac{N}{2} \rfloor} \frac{\rho^k}{M_1(k)M_2(N-k)} + \frac{\rho^{N-k}}{M_1(N-k)M_2(k)}} \leq \frac{\sum_{k=0}^{n} \frac{\rho^{N-k}}{M_1(k)M_2(N-k)} + \frac{\rho^k}{M_1(N-k)M_2(k)}}{\sum_{k=0}^{\lfloor \frac{N}{2} \rfloor} \frac{\rho^{N-k}}{M_1(k)M_2(N-k)} + \frac{\rho^k}{M_1(N-k)M_2(k)}}, \tag{27}$$

which can be directly verified, taking into account $\mu_1(n) \geq \mu_2(n)$.

To prove (ii), write $M = M_1 = M_2$, since $\mu_1 = \mu_2$. We want to establish (26) with the inequality reversed. In place of (27), it suffices to show

$$\frac{\sum_{k=0}^{n} \frac{\rho^k + \rho^{N-k}}{M(k)M(N-k)}}{\sum_{k=0}^{\lfloor \frac{N}{2} \rfloor} \frac{\rho^k + \rho^{N-k}}{M(k)M(N-k)}}$$

is decreasing in ρ, since $\rho = v_1/v_2 \geq (v_1 - \Delta)/(v_2 + \Delta)$. Again, a direct verification is not difficult. □

Theorem 4.15 Consider the two networks in Lemma 4.13. Suppose $v_1 \geq v_2$.
(i) If in addition to the conditions in Lemma 4.14(i) suppose $\mu_2(n)$ is concave in n, and $\mu_1(n) - \mu_2(n)$ is increasing in n, then, $TH(N) \geq TH'(N)$ for any given N.
(ii) If in addition to the conditions in Lemma 4.14(ii) $\mu_1 = \mu_2$ is concave in n, then, $TH(N) \leq TH'(N)$ for any given N.

Proof. To show (i), following (19), we have

$$TH(N) = E\mu_1(X_1)/v_1 = E\mu_2(N - X_1)/v_2,$$

and hence, taking into account $v_1 + v_2 = 1$,

$$\begin{aligned} TH(N) &= E[\mu_1(X_1) + \mu_2(N - X_1)] \\ &= E[\mu_1(X_1) - \mu_2(X_1) + \mu_2(X_1) + \mu_2(N - X_1)]. \end{aligned} \tag{28}$$

Replacing X_1 by X_1' yields a similar expression for $TH'(N)$.

Now, since $X_1 \geq_{\ell r} X_1'$ [Lemma 4.13] implies $X_1 \geq_{st} X_1'$, and since $\mu_1(n) - \mu_2(n)$ is increasing in n as assumed, we have

$$\mu_1(X_1) - \mu_2(X_1) \geq_{st} \mu_1(X_1') - \mu_2(X_1'). \tag{29}$$

Furthermore, since $\mu_2(n)$ is concave in n, and $\min(X_1, N - X_1) \geq_{st} \min(X_1', N - X_1')$ [Lemma 4.14(i)], we have,

$$\mu_2(X_1) + \mu_2(N - X_1) \geq_{st} \mu_2(X_1') + \mu_2(N - X_1'), \tag{30}$$

following a simple coupling argument.

The desired result, $TH(N) \geq TH'(N)$ then follows from combining (28) with (29) and (30).

To prove (ii), (29) becomes trivial (both sides are zero), and (30), with the inequality reversed, follows from Lemma 4.14(ii). □

Note the required conditions in Theorem 4.15 are satisfied by an important special case, namely, when the two nodes have multiple parallel servers. Let c_1 and c_2 be the number of servers at the two nodes. Suppose $c_1 \geq c_2$, and suppose all servers have the same service rate ν, then, in addition to the increasing concavity of both μ_1 and μ_2, we also have,

$$\mu_1(n) - \mu_2(n) = \nu[\min\{n, c_1\} - c_2]^+$$

increasing in n.

In Theorem 4.15(i), the second network is obtained through interchanging the workload assignment of the first network (i.e., v_2 to node 1 and v_1 to node 2). Since $\mu_1(n) \geq \mu_2(n)$ for all n, the result simply states that the throughput is increased by assigning a higher workload to a faster node (or, a node with more servers, provided all servers have the same service rate). That is, the throughput is arrangement increasing in the visit ratios. Specifically, $v = (v_1, v_2) \geq_A v' = (v_2, v_1)$ implies $TH(v) \geq TH(v')$.

On the other hand, the result in Theorem 4.15(ii) hence indicates that the throughput is a Schur concave function of the visit ratios: $v = (v_1, v_2) \geq_m v' = (v_1 - \Delta, v_2 + \Delta)$ implies $TH(v) \leq TH(v')$. It is interesting to note how this result is obtained in the proof. We have in fact proved a stochastic majorization result:

$$X(v) = (X_1, N - X_1) \geq_m (X_1', N - X_1') = X(v'), \tag{31}$$

in the sense that

$$\min\{X_1, N - X_1\} \leq_{st} \min\{X_1', N - X_1'\}.$$

This also leads to, via coupling, $Ef(X(v)) \geq Ef(X(v'))$ for any Schur convex function. Now, $\mu(x_1) + \mu(x_2)$ is a Schur concave function when $\mu(\cdot)$ is a concave function. Hence

$$TH(v) = E[\mu(X_1) + \mu(N - X_1)] \leq E[\mu(X_1') + \mu(N - X_1')] = TH(v')$$

follows.

Since both arrangement and majorization orderings are pairwise orderings, the results in Theorem 4.15 extend immediately to an m-node network. Based on (20), in order to preserve the ordering among equilibrium rates under convolution, the only additional condition is the increasingness of the service rate functions, which translates into the PF_2 property of the Y_i's that define the joint queue-length distribution as well as the throughput.

Theorem 4.16 In a closed Jackson network with m nodes and N jobs, where the service rates at all nodes are increasing and concave functions,
(i) suppose the nodes are ordered such that $\mu_i(n) - \mu_{i+1}(n)$ is increasing in n, for all $i = 1, ..., m-1$, then the throughput is arrangement increasing in the visit ratios $(v_i)_{i=1}^m$;
(ii) suppose all nodes have the same service rate function, then the throughput is a Schur concave function of $(v_i)_{i=1}^m$.

Remark 4.17 In the special case of parallel servers, we can allow the servers at different nodes to have different service rates, say ν_i for station i, by viewing v_i/ν_i as v_i. The arrangement and majorization results then apply to the modified v_i.

5 Stochastic Rearrangement and Majorization

5.1 The Deterministic Theory

In §2, we have seen in Example 2.7 the key role played by the pairwise interchange argument through a stochastic version of a simple rearrangement inequaility (6). An immediate extension of (6) is the well known HLP inequality. Let x and y be two n-vectors of real-valued components; let x^π and y^σ denote the rearrangement of the components of x and y according to the permutations π and σ, respectively. In particular, let the superscripts $\uparrow$ and $\downarrow$ denote rearrangements of components according to, respectively, increasing and decreasing orders. Then, the HLP inequality can be expressed as a relation among the inner products:

$$x^\uparrow \cdot y^\downarrow \leq x^\pi \cdot y^\sigma \leq x^\uparrow \cdot y^\uparrow,$$

for any permutations π and σ.

A generalization of the HLP inequality is Day's inequality (Day [6]; also see Marshall and Olkin [14], Chapter 6).

Lemma 5.1 (Day [6]) Denote

$$f \circ \phi(x, y) = f(\phi(x_1, y_1), ..., \phi(x_n, y_n)).$$

Then,

$$f \circ \phi(x^\uparrow, y^\downarrow) \leq f \circ \phi(x^\pi, y^\sigma) \leq f \circ \phi(x^\uparrow, y^\uparrow),$$

for any permutations π and σ, any monotone (i.e., either increasing or decreasing) and supermodular function ϕ and any increasing Schur convex function f.

Since the increasing Schur convex functions form the defining class for weak majorization, Day's inequality can be equivalently expressed as

$$\phi(x^{\uparrow}, y^{\downarrow}) \leq_{wm} \phi(x^{\pi}, y^{\sigma}) \leq_{wm} \phi(x^{\uparrow}, y^{\uparrow}). \tag{32}$$

To prove Day's inequality, we only need to show, for $x_1 \leq x_2$ and $y_1 \leq y_2$,

$$f(\phi(x_1, y_1), \phi(x_2, y_2)) \geq f(\phi(x_1, y_2), \phi(x_2, y_1)). \tag{33}$$

Suppose ϕ is increasing. Denote the four arguments of f above as z_1, z_4, z_2 and z_3, respectively. Then,

$$z_1 \leq z_2, z_3 \leq z_4, \quad z_1 + z_4 \geq z_2 + z_3,$$

since ϕ is increasing and supermodular. In other words,

$$(z_1, z_4) \geq_{wm} (z_2, z_3).$$

Hence (33) follows, since f is increasing and Schur convex. If ϕ is decreasing, interchange z_1 and z_4, the remaining arguments still apply.

The stochastic result in (31) can be compared with the above. There, $v \geq_m v'$ led to $X(v) \geq_m X(v')$ in the sense that $Ef(X(v)) \geq Ef(X(v'))$ for any Schur convex function f. This suggests that Schur convex functions be used as a defining class for stochastic majorization as well. However, this notion of stochastic majorization often turns out to be too strong in many applications, where any interesting results might only hold if f belongs to a *subclass* of Schur convex functions. In other words, Schur convex functions as a defining class appears to be too broad. Another limitation of (31) is that the result was obtained primarily through algebraic manipulations [see the proof of Lemma 4.14], which reveal very little about the functional relationship of X with respect to v.

5.2 The Stochastic Counterpart

Below, we shall attempt to directly "translate" Day's inequality in Lemma 5.1 into a stochastic version. In doing so, we find that for Schur convex functions to be the defining class, the required stochastic order relation among the random variables involved has to be the strongest, the likelihood ratio ordering. For weaker orderings, the defining class has to be restricted to a subclass of Schur convex functions. [This class of functions have a "directional" property, analogous to (but different from) the directional convexity in Definition 3.15; see Lemma 5.4 below.] The main theorems in this section are Theorems 5.2, 5.10 and 5.12.

Theorem 5.2 Let X and Y be two independent random n-vectors, each with independent components. Suppose the components of X (Y) are ordered in the sense of $\geq_a$ ($\geq_b$) to be specified below. Suppose $\phi(\cdot,\cdot)$ is a monotone and supermodular function. Then, for any permutations π and σ of $\{1,2,...,n\}$, we have

$$Ef\circ\phi(X^{\uparrow},Y^{\downarrow})\leq Ef\circ\phi(X^{\pi},Y^{\sigma})\leq Ef\circ\phi(X^{\uparrow},Y^{\uparrow}),$$

for any function f that is
(i) increasing and Schur convex, if $(a,b)=(\ell r,\ell r)$;
(ii) symmetric, submodular, increasing and convex in each variable, if $(a,b)=(st,st)$;
(iii) expressible as $f=g\circ h$ with h satisfying the properties in (ii) and g: $\Re\mapsto\Re$, an increasing and convex function, if $(a,b)=(\ell r,hr)$ or $(hr,\ell r)$, and ϕ increasing.
Moreover, when ϕ is submodular, the above results also hold with the increasingness of f, as well as the increasingness of ϕ in (iii), replaced by decreasingness.

Remark 5.3 (i) A symmetric, convex function is Schur convex. To see this, without loss of generality, consider a bivariate function $f(x_1,x_2)$ that is symmetric and convex. Suppose $x_1>x_2$, we want to show

$$f(x_1,x_2)\geq f(x_1-\epsilon,x_2+\epsilon) \tag{34}$$

for any $0<\epsilon<x_1-x_2$. The convexity of f implies, for any $0<\theta<1$,

$$\theta f(x_1,x_2)+(1-\theta)f(x_2,x_1)\geq f[\theta x_1+(1-\theta)x_2,\theta x_2+(1-\theta)x_1].$$

Now, the symmetry of f implies the LHS above equals $f(x_1,x_2)$. Hence, pick θ such that $\epsilon=(1-\theta)(x_1-x_2)$. A Schur convex function, however, is not necessarily convex.
(ii) While (joint) convexity implies convexity in each variable, the converse is not true in general. A special case that does make the converse true is when the function f is *separable*, i.e., $f(x)=\sum_i g_i(x_i)$. When g_i is convex in x_i, for all i, f is (jointly) convex in x. [The symmetry of f would further require that the g_i functions are the same for all i.] When f is separable, it is easily verified [cf. Remark 3.13] that f is both submodular and supermodular (hence *modular*), provided the g_i's are either increasing for all i or decreasing for all i.
(iii) A symmetric function that is convex in each variable is not necessarily Schur convex either. [As an example, consider $f(a,b)=a^2+3ab+b^2$. It is symmetric, convex in a and in b. However, $f_a-f_b=-(a-b)$, making it Schur concave.] However, with the additional submodularity condition, the function does become Schur convex. See Remark 5.5, following Lemma 5.4 below. Therefore, the class of functions in Theorem 5.2(ii) is indeed a subset of the Schur convex class.

(iv) The two classes of functions in Theorem 5.2 (ii) and (iii) are simply different classes: while the class in (iii) is preserved under composition with increasing and convex functions, the class in (ii) is not (submodularity would fail). However, $\max_i(x_i)$, is a function that belongs to both classes, since for any increasing function $g(z)$,

$$g(\max_i(x_i)) = \max_i(g(x_i)).$$

In particular, when g is increasing and convex, $g(\max_i(x_i))$ belongs to both classes (max is increasing and convex).

(v) In the same spirit as expressing Day's rearrangement inequality Lemma 5.1 as a majorization result, we can also express Theorem 5.2 as a *stochastic* majorization result as follows:

$$\phi(X^{\uparrow}, Y^{\downarrow}) \leq_{wm} \phi(X^{\pi}, Y^{\sigma}) \leq_{wm} \phi(X^{\uparrow}, Y^{\uparrow}).$$

Here, the ordering $\leq_{wm}$ among the random vectors is in the precise sense stated in Theorem 5.2. For instance, in the setting of Theorem 5.2(ii), the stochastic $\leq_{wm}$ means an order that is preserved by a function that is symmetric, submodular, increasing and convex. In this sense, the stochastic version of Day's rearrangement inequality requires the additional submodularity (as well as strengthening Schur convexity to convexity) when the components of the random vectors X and Y are ordered in stochastic ordering. These additional restrictions are not needed when the components are ordered in the likelihood ratio ordering.

We now prove the three parts of Theorem 5.2 separately. Part (i) turns out to be the easiest. Since all the random variables involved are independent, no generality is lost to assume X and Y in Theorem 5.2 are 2-vectors. This is tantamount to applying a pairwise interchange argument, fixing all the other components in the original X and Y vectors.

Proof. [of Theorem 5.2(i)]. For $x_1 \geq x_2$ and $y_1 \geq y_2$, we have

$$f(\phi(x_1, y_1), \phi(x_2, y_2)) \geq f(\phi(x_1, y_2), \phi(x_2, y_1)),$$

following the given conditions on f and ϕ and Day's inequality in (33). Now, fix x_1 and x_2, the above can be written as $g(y_1, y_2) \geq g(y_2, y_1)$ for $y_1 \geq y_2$, taking into account the symmetry of f (implied by its Schur convexity). Hence, $g \in \mathbf{C}_{\ell r}$, and

$$Ef(\phi(x_1, Y_1), \phi(x_2, Y_2)) \geq Ef(\phi(x_1, Y_2), \phi(x_2, Y_1)),$$

since $Y_1 \geq_{\ell r} Y_2$. Next, repeat the arugument by writing the above as $h(x_1, x_2) \geq h(x_2, x_1)$ for $x_1 \geq x_2$. $\square$

To prove Theorem 5.2(ii), we need two lemmas. The first one reveals more about the properties of the class of functions specified there. These properties relate closely to the directional convexity in §3.

Lemma 5.4 f is submodular, and convex in each variable, if and only if

$$f(z^1) + f(z^4) \geq f(z^2) + f(z^3) \tag{35}$$

for any four vectors $z^i = (x_i, y_i)$, $i = 1, 2, 3, 4$, satisfying:

$$x_1 \geq \max\{x_2, x_3\}, \quad y_4 \geq \max\{y_2, y_3\};$$

$$x_1 + x_4 = x_2 + x_3, \quad y_1 + y_4 = y_2 + y_3.$$

If, in addition, f is increasing (resp. decreasing), the results also hold, with the last two equalities above replaced by $\geq$ (resp. $\leq$).

The proof of the lemma is completely analogous to the proof of Lemma 3.16, via the remark below, and hence omitted.

Remark 5.5 Notice the close resemblance of the above lemma to the notion of directional convexity in Definition 3.15. In fact, following Lemma 3.16, directional convexity is equivalent to *supermodularity*, instead of submodularity, plus convexity in each variable. Hence, in contrast to directional convexity, the four points (vectors), z^i (i=1,2,3,4), here have the following relation: z^1 and z^4 are, respectively the southeast and the northwest corners of the rectangle they form (on the x-y plane), z^2 and z^3 are then two points which fall inside (or at most at the boundary) of the rectangle. That is, while directional convexity is a property along the northeast-southwest diagonal, the property in Lemma 5.4 above is along the northwest-southeast diagonal.
(ii) When the funtion f in Lemma 5.4 is also symmetric, then it is Schur convex. To see this, let $x_1 > x_2$ and $0 < \epsilon < x_1 - x_2$ as in Remark 5.3, and set

$$z^1 = (x_1, x_2), \quad z^4 = (x_2, x_1);$$

$$z^2 = (x_1 - \epsilon, x_2 + \epsilon), \quad z^3 = (x_2 + \epsilon, x_1 - \epsilon).$$

Then (35) reduces to the desired inequality in (34).

Lemma 5.6 Suppose f is increasing (resp. decreasing), submodular and convex in each variable, and ϕ is monotone and supermodular (resp. submodular). Given $y_1 \geq y_2$, then

$$f(\phi(x_1, y_1), \phi(x_2, y_2)) - f(\phi(x_1, y_2), \phi(x_2, y_1))$$

is increasing in x_1.

Proof. Suppose ϕ is increasing. Let $a > 0$ be any given constant. Define the four points in Lemma 5.4 as follows:

$$z^1 = (\phi(x_1 + a, y_1), \phi(x_2, y_2)), \quad z^4 = (\phi(x_1, y_2), \phi(x_2, y_1));$$

$$z^2 = (\phi(x_1 + a, y_2), \phi(x_2, y_1)), \quad z^3 = (\phi(x_1, y_1), \phi(x_2, y_2)).$$

Then clearly z^i satisfy the requirements in Lemma 5.4. In particular,

$$\phi(x_1 + a, y_1) + \phi(x_1, y_2) \geq \phi(x_1 + a, y_2) + \phi(x_1, y_1)$$

follows from the supermodularity of ϕ. When f is increasing, submodular and convex in each variable, we have $f(z^1) - f(z^2) \geq f(z^3) - f(z^4)$, following Lemma 5.4, hence the desired increasingness in x_1.

When ϕ is decreasing, interchange the definition of z^1 and z^4. The case f is decreasing is similarly proved. □

We are now ready to prove Theorem 5.2(ii).

Proof. [of Theorem 5.2(ii)]. Consider f increasing; the decreasing case is similarly proved. We first show, under the given conditions, that

$$Ef(\phi(x_1, Y_1), \phi(x_2, Y_2)) - Ef(\phi(x_1, Y_2), \phi(x_2, Y_1)) \tag{36}$$

is increasing in x_1. Since $Y_1 \geq_{st} Y_2$, there exists a pair of random variables, $\hat{Y}_1 \geq \hat{Y}_2$, and $\hat{Y}_i \stackrel{\mathrm{d}}{=} Y_i$ for i=1,2. Generate *independently* another pair, $\hat{Y}_1' \geq \hat{Y}_2'$, that are also equal in distribution to Y_1 and Y_2, respectively. Then, for any given $a > 0$, it is easy to verify that

$$f(\phi(x_1 + a, \hat{Y}_1), \phi(x_2, \hat{Y}_2')) - f(\phi(x_1 + a, \hat{Y}_2), \phi(x_2, \hat{Y}_1'))$$

$$\geq f(\phi(x_1, \hat{Y}_1), \phi(x_2, \hat{Y}_2')) - f(\phi(x_1, \hat{Y}_2), \phi(x_2, \hat{Y}_1')),$$

since the four points (the arguments of f) above correspond to the four z^i in Lemma 5.4. Also note that involved in each z^i is a pair of *independent* replicas of Y_1 and Y_2 (which is a requirement, since Y_1 and Y_2 are independent as assumed). Taking expectations on both sides, we obtain the desired increasingness in x_1. Since f is symmetric, (36) can be written as $g(x_1, x_2) - g(x_2, x_1)$, which is increasing in x_1.

Next, since $X_1 \geq_{st} X_2$, similar as before, generate a pair of replicas of X_1 and X_2: $\hat{X}_1 \geq \hat{X}_2$, and generate *independently* another replica of X_2: $\hat{X}_2'$. Then, the increasing property (in x_1) just obtained yields

$$g(\hat{X}_1, \hat{X}_2') - g(\hat{X}_2', \hat{X}_1) \geq g(\hat{X}_2, \hat{X}_2') - g(\hat{X}_2', \hat{X}_2).$$

Taking expectations on both sides, the RHS vanishes, giving the desired conclusion. □

Proof. [of Theorem 5.2(iii)]. Suppose $(a, b) = (\ell r, hr)$. Fix $x_1 \geq x_2$, and write f as

$$f(y_1, y_2) = g \circ h(\phi(x_1, y_1), \phi(x_2, y_2)),$$

where g is an increasing and convex function and h satisfies all the properties in (5.3 ii). Write the h part above as $h(y_1, y_2)$, then it is clear that $h \in \mathbf{C}_{hr}$, since it is increasing in y_1, and $h(y_1, y_2) - h(y_2, y_1)$ is also increasing in y_1, following Lemma 5.4. It then follows that $g \circ h \in \mathbf{C}_{hr}$, since the class $\mathbf{C}_{hr}$ is closed under composition with increasing and convex functions. Therefore, $Y_1 \geq_{hr} Y_2$ leads to

$$Eg \circ h(\phi(x_1, Y_1), \phi(x_2, Y_2)) \geq Eg \circ h(\phi(x_1, Y_2), \phi(x_2, Y_1)).$$

Now, write the LHS above as $\tilde{f}(x_1, x_2)$, then the RHS is $\tilde{f}(x_2, x_1)$, taking into account the symmetry of h. Hence, $\tilde{f} \in \mathbf{C}_{\ell r}$. Since $X_1 \geq_{\ell r} X_2$, we have

$$Eg \circ h(\phi(X_1, Y_1), \phi(X_2, Y_2)) \geq Eg \circ h(\phi(X_1, Y_2), \phi(X_2, Y_1)),$$

which is the desired conclusion. □

We now illustrate the application of Theorem 5.2 through a scheduling example.

Example 5.7 Consider the problem of scheduling n jobs in a tandem queue of m servers, with infinite intermediate buffers. All jobs are present at time zero. For any job i, the processing requirement at all servers are equal, denoted P_i. Let D_i denote the due time for job i. Suppose the processing times and due times are independent among the n jobs. Let C_i, L_i and T_i denote, respectively, the completion time, lateness and tardiness of job i. Then, $L_i = C_i - D_i$, and $T_i = \max\{L_i, 0\}$. Denote by C, L and D the corresponding vectors. Of particular interest are the *departure* epochs of the jobs. These are the components of the vector C arranged in increasing order: $C^{\uparrow} = (C_{(i)})_{i=1}^{n}$, where $C_{(i)}$ denotes the ith smallest component of C. Suppose the schedule of the n jobs follows the permutation π. Then, it is not difficult to derive the following expressions (see Pinedo and Weber [15]): for $j = 1, ..., n$,

$$C^{\pi}_{(j)} = \sum_{i=1}^{j} P_{\pi(i)} + (m-1)\max\{P_{\pi(1)}, ..., P_{\pi(j)}\}.$$

In particular, $C^{\pi}_{(n)}$ is the time to complete all n jobs from the system, the "makespan".

Proposition 5.8 Suppose $P_1 \le_a P_2 \le_a \ldots \le_a P_n$. Let s and ℓ denote, respectively, the schedules (1,2,...,n) (shortest processing time first) and (n,...,2,1) (longest processing time first). Then, for any permutation schedule π,

$$Ef(C^{\uparrow s}) \le Ef(C^{\uparrow \pi}) \le Ef(C^{\uparrow \ell}),$$

(i) for any increasing function f, if $a = \ell r$;
(ii) for any increasing and supermodular function f, if $a = hr$.

Proof. To prove (i), we only need to show that for any two permutation π and σ, $\pi \le_A \sigma$ implies $f(C^{\uparrow \pi}) \le f(C^{\uparrow \sigma})$, when f is an increasing function. This is directly verified.

To prove (ii), denote $x_i = P_{\pi(i)}$ and write the completion times as $C^{\pi}_{(i)} = g_i(x_1, ..., x_i)$. Note that g_i is symmetric in its arguments. We want to show, given $x_i \ge x_{i+1}$, the difference between the schedule π and another schedule σ that interchanges the processing sequence of $\pi(i)$ and $\pi(i+1)$, expressed as

$$f(\ldots, g_i(x_1, \ldots, x_{i-1}, x_i), g_{i+1}(x_1, \ldots, x_{i-1}, x_i, x_{i+1}), \ldots) \tag{37}$$

$$-f(\ldots, g_i(x_1, \ldots, x_{i-1}, x_{i+1}), g_{i+1}(x_1, \ldots, x_{i-1}, x_{i+1}, x_i), \ldots), \tag{38}$$

is increasing in x_i, i.e., $f \in \mathbf{C}_{hr}$. Since f is supermodular and g is increasing and symmetric, for $a > 0$, the above difference is dominated by

$$f(\ldots, g_i(x_1, \ldots, x_{i-1}, x_i), g_{i+1}(x_1, \ldots, x_{i-1}, x_i + a, x_{i+1}), \ldots)$$

$$-f(\ldots, g_i(x_1, \ldots, x_{i-1}, x_{i+1}), g_{i+1}(x_1, \ldots, x_{i-1}, x_{i+1}, x_i + a), \ldots),$$

which, in turn, is dominated by

$$f(\ldots, g_i(x_1, \ldots, x_{i-1}, x_i + a), g_{i+1}(x_1, \ldots, x_{i-1}, x_i + a, x_{i+1}), \ldots)$$

$$-f(\ldots, g_i(x_1, \ldots, x_{i-1}, x_{i+1}), g_{i+1}(x_1, \ldots, x_{i-1}, x_{i+1}, x_i + a), \ldots),$$

since both f and g are increasing functions. Hence, we have established the desired increasing property (in x_i) of the difference in (37). $\Box$

Proposition 5.9 If, in addition to the given conditions in Proposition 5.8, $D_1 \le_b D_2 \le_b \ldots \le_b D_n$, then

$$Ef(L^s) \le Ef(L^\pi) \le Ef(L^\ell), \quad Ef(T^s) \le Ef(T^\pi) \le Ef(T^\ell),$$

(i) for any function f that belongs to one of the three classes in Theorem 5.2(i,ii,iii), if $(a, b) = (\ell r, \ell r)$, $(\ell r, st)$, $(\ell r, hr)$, respectively;
(ii) for any function f that belongs to one of the three classes in Theorem 5.2(i,ii,iii) with the additional property of supermodularity, if $(a, b) = (hr, \ell r)$, (hr, st), (hr, hr), respectively.

Proof. We show the results for the sequence s, results for the sequence ℓ are similarly proved. Consider lateness. Note, following the sequence s, $C^s_{(j)} = C^s_j$ for all j. For (i), since f is increasing, applying Proposition 5.8(i), we have

$$Ef(L^s) = Ef(C^s_{(1)} - D_1, \ldots, C^s_{(n)} - D_n) \leq Ef(C_{(1)} - D_1, \ldots, C_{(n)} - D_n). \tag{39}$$

Now, since $C_{(1)} \leq \ldots \leq C_{(n)}$ (a.s.), applying Theorem 5.2 and viewing $\phi = C + (-D)$, we have

$$Ef(C_{(1)} - D_1, \ldots, C_{(n)} - D_n) \leq Ef(C_1 - D_1, \ldots, C_n - D_n) = Ef(L^\pi).$$

For (ii), the proof is the same. Note, however, to claim (39), based on Proposition 5.8(ii), it is further required that f be supermodular, in addition to the properties in Theorem 5.2.

Finally, for tardiness, since $T_j = \max\{L_j, 0\}$ is an increasing and convex function of L_j, it will not change the properties (increasingness and supermodularity) of the ϕ function above. Hence, the same proof applies to tardiness. □

The additional requirement of supermodularity in Proposition 5.9(ii) above virtually restricts the function f to the separable class: When $f(x) = \sum_{i=1}^n g(x_i)$, with g increasing and convex in its argument, it is a function that satisfies all the required properties in Proposition 5.9(ii). Specifically, it is symmetric and increasing; since it is separable and convex in each variable, it is also (jointly) convex, as well as both submodular and supermodular [cf. Remark 5.3(ii)].

5.3 Connections with Stochastic Convexity/Supermodularity

We start with a theorem that combines the stochastic supermodularity and submodularity of Definition 3.12 with the stochastic rearrangement of Theorem 5.2(ii), followed by an example that illustrates its application in assembly operation with random component yield.

Theorem 5.10 $\{X(\theta)\}$ is a parameterized family of independent random variables. Suppose $\theta = (s_i, t_j)$ where s_i and t_j are real or integer valued scalars. Suppose $\{X(s_i, t_j)\}$ is stochastically monotone (i.e, either increasing or decreasing in the two parameters) and supermodular in the sense of Definition 3.12. Consider two sets of parameter values: $s = (s_i)_{i=1}^n$ and $t = (t_j)_{j=1}^n$. Let π and σ be two permutations of $\{1, \ldots, n\}$. Let $X(s^\pi, t^\sigma)$ denote the random vector $(X(s_{\pi(i)}, t_{\sigma(i)}))_{i=1}^n$. Let $f : \Re^n \mapsto \Re$ be a function that is symmetric, increasing, submodular, and convex in each variable, i.e., f belongs to the class of Theorem 5.2(ii). Then,

$$Ef(X(s^\uparrow, t^\downarrow)) \leq Ef(X(s^\pi, t^\sigma)) \leq Ef(X(s^\uparrow, t^\uparrow)).$$

If the increasingness of f is changed to decreasingness, the above also holds if $X(s_i, t_j)$ is stochastically monotone and submodular.

Proof. Similar as before, we only need to consider the two-dimensional case. Suppose $s = (s_1, s_2)$, $t = (t_1, t_2)$, and $s_1 \geq s_2$, $t_1 \geq t_2$. Consider first both f and X are increasing. Since X is stochastically increasing and supermodular, following Definition 3.12, we can generate the replicas

$$\hat{X}_1 \stackrel{\mathrm{d}}{=} X(s_1, t_1), \quad \hat{X}_2 \stackrel{\mathrm{d}}{=} X(s_1, t_2), \quad \hat{X}_3 \stackrel{\mathrm{d}}{=} X(s_2, t_1), \quad \hat{X}_4 \stackrel{\mathrm{d}}{=} X(s_2, t_2);$$

such that

$$\hat{X}_1 \geq \max\{\hat{X}_2, \hat{X}_3\}, \quad \hat{X}_1 + \hat{X}_4 \geq \hat{X}_2 + \hat{X}_3. \tag{40}$$

In addition, independently of the four $\hat{X}$, generate four $\hat{Y}$:

$$\hat{Y}_4 \stackrel{\mathrm{d}}{=} X(s_1, t_1), \quad \hat{Y}_3 \stackrel{\mathrm{d}}{=} X(s_1, t_2), \quad \hat{Y}_2 \stackrel{\mathrm{d}}{=} X(s_2, t_1), \quad \hat{Y}_1 \stackrel{\mathrm{d}}{=} X(s_2, t_2);$$

such that

$$\hat{Y}_4 \geq \max\{\hat{Y}_2, \hat{Y}_3\}, \quad \hat{Y}_1 + \hat{Y}_4 \geq \hat{Y}_2 + \hat{Y}_3. \tag{41}$$

Let $z^i = (\hat{X}_i, \hat{Y}_i)$ for $i = 1, 2, 3, 4$. Then,

$$f(z^1) + f(z^4) \geq f(z^2) + f(z^3)$$

follows directly from (39). Taking expectations on both sides, and taking into account the symmetry of f, we have

$$Ef(X(s_1, t_1), X(s_2, t_2)) \geq Ef(X(s_1, t_2), X(s_2, t_1)), \tag{42}$$

the desired rearragement result.

When X is (stochastically) decreasing, interchange the constructions of the arguments 1 and 4 in (40) and (41). When f is decreasing and X submodular, reverse the second inequality in (40) and (41); the rest still applies. □

The result in Theorem 5.10 can also be expressed as a stochastic majorization relation:

$$X(s^{\uparrow}, t^{\downarrow}) \leq_{wm} X(s^{\pi}, t^{\sigma}) \leq_{wm} X(s^{\uparrow}, t^{\uparrow}).$$

Notice the analogy between the above and its deterministic counterpart in (32). Here, X plays exactly the same role as ϕ.

Example 5.11 Consider an assembly operation over a set of components of m different types. Suppose one unit from each type is required to assemble the final product. Suppose the production of each component follows the yield model in Example 3.14. Let $n = (n_j)_{j=1}^m$ be the vector of lot sizes (of the components), and $p = (p_j)_{j=1}^m$, the vector of yield ratios. The number of final products assembled is then $\min\{X(n_j, p_j); j = 1, \ldots, m\}$, where $X(n_j, p_j) = \sum_{i=1}^{n_j} Y_i(p_j)$, and Y_i are a set of iid Bernoulli random variables with parameter p_j. Since X is stochastically increasing and supermodular, Theorem 5.10 applies:

$$E\min\{X(n^{\uparrow}, p^{\downarrow})\} \geq E\min\{X(n^{\pi}, p^{\sigma})\} \geq E\min\{X(n^{\uparrow}, p^{\uparrow})\}. \tag{43}$$

That is, the best arrangement is to assign large lot sizes to low yielding components.

To derive (43) from Theorem 5.10, make use of $\min_i\{x_i\} = -\max_i\{-x_i\}$. Also, similar to max, all the properties of min is preserved through composition with increasing and concave functions. Hence the inequalities in (43) also hold when the expectation is taken over an increasing concave (say, profit) function of the nubmer of assembled products.

Note here all that is required is the supermodularity of X in the lot size and the yield ratio. It is easy to verify that it suffices to have each component increasing in its yield ratio.

Next, following the spirit of Theorem 5.10, we develop a notion of stochastic majorization based on the SICX-sp property [cf. Definition 3.5].

Theorem 5.12 Let $\{X(\eta)\}$ be a family of independent random variables, parmeterized by a real or integer scalar η. Then, $(\theta_i^1)_{i=1}^n \leq_{wm} (\theta_i^2)_{i=1}^n$ implies

$$Ef(X(\theta_1^1), ..., X(\theta_n^1)) \leq Ef(X(\theta_1^2), ..., X(\theta_n^2))$$

for all functions f that are increasing, symmetric, submodular, and convex in each variable [i.e., f belongs to the class of (5.3 ii)], provided $\{X(\eta)\} \in$ SICX-sp.

Proof. Follow the proof of Theorem 5.10, make use of the SICX-sp property in place of the supermodularity. Without loss of generality, suppose

$$\theta_1^2 \geq \theta_1^1 \geq \theta_2^1 \geq \theta_2^2, \quad \theta_1^1 + \theta_2^1 \leq \theta_1^2 + \theta_2^2.$$

Hence, pick four parameter values:

$$\eta_1 = \theta_1^2, \quad \eta_2 = \theta_1^1, \quad \eta_3 = \theta_2^1, \quad \eta_4 = \theta_2^2,$$

and let $\hat{X}_i \stackrel{\mathrm{d}}{=} X(\eta_i)$, for $i = 1, 2, 3, 4$. Independent of the $\hat{X}_i$, generate four more replicas, $\hat{Y}_i$, in reverse order, i.e., $\hat{Y}_1 \stackrel{\mathrm{d}}{=} X(\theta_4)$, and so forth. The rest of the proof of Theorem 5.10 then follows through. In particular, (40) and (41) hold, due to the SICX-sp property; and (42) takes the form

$$Ef(X(\theta_1^2), X(\theta_2^2)) \geq Ef(X(\theta_1^1), X(\theta_2^1)),$$

which is the desired conclusion. □

Hence, the above result can be expressed as the stochastic majorization of two parameterized random vectors,

$$(X(\theta_1^1), ..., X(\theta_n^1)) \leq_{wm} (X(\theta_1^2), ..., X(\theta_n^2)),$$

given the parameter vectors are ordered under weak majorization. Here, the stochastic majorization is in the same sense as in Theorem 5.10, namely, a relation that is preserved by functions that symmetric, increasing, submodular, and convex in each variable.

Example 5.13 Continue with the assembly model in Example 5.11. Suppose all components have the same lot size, n. Since the yield of type i, $\{X(n, p_i)\} \in$ SIL-sp in p_i [cf. Example 3.8], the total number of assembled products, $\min\{X(n, p_i), i = 1, ..., m\}$, fit into the setting of Theorem 5.12. Furthermore, thanks to the min function, the required properties in Theorem 5.12 will be preserved by composition with increasing and concave functions. Hence, the number of assembles products will be increased, in the sense of concave ordering, when the vector of yield ratio, $(p_i)_{i=1}^m$ is decreased in the sense of weak majorization. In other words, it helps when the yield ratio is more uniform among the components.

Example 5.14 Consider a system of m parallel servers, with resequencing of completed jobs — a useful model for assembly systems. Jobs arrive at the system following a Poisson process with rate λ. The n th arrival requires a service time of length $S_n^{\xi_n}$ if it is assigned to server ξ_n, where $\{\xi_n\}$ is a sequence of iid multinomial random variables: $P(\xi_n = i) = p_i$, $i = 1, ..., m$, $\sum_{i=1}^m p_i = 1$. For each i, the service times, S_n^i, are iid. Once assigned to a server, the job is not allowed to switch servers. Each server serves its own queue following the FCFS discipline. After service completion, the jobs wait in a resequencing area to be released according to their arrival sequence (i.e., a job won't be released until all the jobs that originally arrived before it have been released).

Let W denote the stationary sojourn time in system of each job (including the time spent in the resequencing area). We can write W explicitly as a function of $\theta = (p_i)_{i=1}^m$:

$$W(\theta) = \max\{X_1(p_1), ..., X_m(p_m)\}, \tag{44}$$

where $X_i(p_i)$ is the stationary workload in an M/G/1 queue with arrival rate λp_i and service times $\{S_n^i; n = 1, 2, ...\}$. Without loss of generality, assume $\lambda = 1$; and to ensure stationarity, assume $E(S_n^i) < 1$ for all i.

Now, from standard queueing theory it is known that the stationary workload of an M/G/1 queue with arrival rate p can be expressed as (omitting the script i)

$$X(p) \stackrel{\mathrm{d}}{=} \sum_{k=1}^{N(p)} R_k,$$

where R_k is a set of iid random variables following the residual service time distribution [specifically, the density function $f_R(x) = P(S_n > x)/E(S_n)$]; $N(p)$ follows a geometric distribution with parameter p. From Example 3.9, we know $\{N(p)\} \in$ SIL-sp, and hence $\{X(p)\} \in$ SIL-sp as well.

When all the m servers are identical, i.e., all the service times are iid, (44) can be written as

$$W(\theta) = \max\{X(p_1), \ldots, X(p_m)\}.$$

Hence, Theorem 5.12 applies. That is, for any two job assignment schemes, indexed by the superscripts 1 and 2, if $\theta^1 \leq_m \theta^2$, then $W(\theta^1) \leq_{icx} W(\theta^2)$. [Note, for any increasing and convex function g, $g \circ \max$ still has all the required properties in Theorem 5.12, and hence the convex ordering.]

Bibliographical Notes

Section 2. Introductory materials on likelihood ratio, hazard rate, stochastic and convex orderings can be found in Ross [18] (Chapter 8). Stoyan [32] is a more detailed source of reference on stochastic comparisons, including many queueing applications. Most of the materials in this section, in particular the functional characterization of likelihood ratio and hazard rate orderings, are from Shanthikumar and Yao [29]. Theorem 2.3 was partially known before, see, e.g., Brown and Solomon [1]; also, Proposition 8.4.2 (p. 268) of [18].

Section 3. The *sp* notion of stochastic convexity was developed in Shaked and Shanthikumar [19, 20, 21], along with other notions of stochastic convexity. For instance, the *st* version, SICX-st, is defined as follows: $\{X(\theta)\}$ is stochastically increasing and convex in the *st* sense if $Ef[X(\theta)]$ is increasing and convex in θ for any increasing function f. (The convexity of f is *not* required here.) It can be shown that this notion of stochastic convexity is equivalent to the distribution function of $X(\theta)$, $F(x,\theta)$, being decreasing and concave in θ (or $\bar{F} = 1 - F$ increasing and convex). It can also be shown that SICX-st $\Rightarrow$ SICX-sp. However, SICX-st and SSICX do not imply each other. For instance, the family of degenerate random variables,

$\{X(\theta) \equiv \theta\}$ satisfies SSICX but not SICX-st. On the other hand, $F(x,\theta)$ decreasing and concave in θ need not satisfy the conditions, in Remark 3.7, for SSICX.

The notion of strong stochastic convexity is due to Shanthikumar and Yao [30]. The characterization of the spatial and temporal behavior of Markov chains and its applications in §3.3 are from Shanthikumar and Yao [28, 31]. The application to manufacturing processes with trial runs in §3.4 is from Gallego, Yao and Moon [8].

The notion of directional convexity and its properties in Definition 3.15 and Lemma 3.16 was developed in [21], including the stochastic version (see Definition 3.18), which we did not elaborate in this article. The stochastic directional convexity appears to be particularly useful in characterizing the spatial-temporal behavior of stochastic processes (e.g., strengthening the results in §3.3). Refer to [21], and Chang *et al* [3, 4] for a variety of applications along this line.

Section 4. Marshall and Olkin [14] is a standard reference for majorization and arrangement orderings, as well as the related functional classes, such as Schur convex and arrangement increasing functions. The results in Theorems 4.2 and 4.4, as well as the queueing network results in Example 4.11 and Proposition 4.12, are from [30]. The application of majorization to queueing networks in Lemmas 4.13 and 4.14, and Theorems 4.15 and 4.16 first appeared in Shanthikumar [23], and Shanthikumar and Yao [26]. The notion of equilibrium rate plays an important role in queueing networks. It was first introduced in Shanthikumar and Yao [24, 25], and applied in a variety of settings, including Shanthikumar [23] and Shanthikumar and Yao [27].

Section 5. Theorem 5.2, its proof, and the scheduling applications in Example 5.7 and Propositions 5.8 and 5.9, first appeared in Chang and Yao [5], which includes many more examples that apply Theorem 5.2 to stochastic scheduling. (Also see Righter [16].) The class of functions that are symmetric, submodular, and convex in each variable (in Theorem 5.2) first appeared in Chang [2], where it was denoted $\mathbf{C}_5$, and was shown to be a subclass of Schur convex functions (cf. Lemma 5.4 and Remark 5.5). In [2] the main motivation is to develop a new notion of stochastic majorization that is weaker (and hence more broadly applicable) than the earlier, standard notion, of which the defining class is Schur convex functions (see [14]).

In the parametric case, Theorems 5.12 and 5.10 are results of combining stochastic majorization with stochastic convexity and stochastic supermodularity. Note that Theorem 5.12 is just one way to generate stochastic majorization, in the parametric case, via stochastic convexity. Specifically, it is based on the *sp* notion of stochastic convexity (Definition 3.4). It is possible to develop different versions of stochastic majorization, based on other notions of stochastic convexity. These different versions of stochastic majorization and their applications in reliability and resource allocation have recently appeared in Liyanage and Shanthikumar [13] and in Shaked and Shanthikumar [22]. For instance, on such version in [13] is based on the *st* notion of stochastic convexity, which, as we mentioned above, is stronger than the *sp* notion. The functional class that characterizes the resulting stochastic majorization is composed of functions that are symmetric, increasing, and submodular, but not necessarily convex in each variable. In other words, it includes $\mathbf{C}_5$ as a subset. (However, this class and the class of Schur convex functions do not appear to have a subset relationship.) One advantage of this functional class is that it is preserved under composition with functions that are increasing and convex. Hence the resulting majorization can be expressed in terms of the convex ordering. [In this regard, it is similar to the functional class in Theorem 5.2(iii).] Based on the *st* notion of stochastic convexity, in [13] an example similar to Example 5.14 is shown to satisfy this stronger stochastic majorization.

References

[1] Brown, M. and Solomon, H. 1973. Optimal Issuing Policies under Stochastic Field Lives. *Journal of Applied Probability,* **10**, 761-768.

[2] Chang, C.S. 1990. A New Ordering for Stochastic Majorization: Theory and Applications. *Advances in Applied Probability,* **24**, 604-634.

[3] Chang, C.S., Chao, X.L. and Pinedo, M. 1991. Monotonicity Results for Queues with Doubly Stochastic Poisson Arrivals: Ross's Conjecture. *Advances in Applied Probability,* **23**, 210-228.

[4] Chang, C.S., Chao, X.L., Pinedo, M. and Shanthikumar, J.G. 1991. Stochastic Convexity for Multidimensional Process and Its Applications. *IEEE Transactions on Automatic Control,* **36**, 1347-1355.

[5] Chang, C.S. and Yao, D.D. 1990. Rearrangement, Majorization, and Stochastic Scheduling. *Mathematics of Operations Research,* to appear.

[6] Day, P.W. 1972. Rearrangement Inequalities. *Canadian Journal of Mathematics,* **24**, 930-943.

[7] Fox, B. 1966. Discrete Optimization via Marginal Analysis. *Management Science,* **13**, 210-216.

[8] Gallego, G., Yao, D.D. and Moon, I. 1990. Optimal Control of a Manufacturing Process with Trial Runs. Technical Report, IEOR Department, Columbia University, New York, NY 10027.

[9] Gordon, W.J. and Newell, G. F. 1967. Closed Queueing Networks with Exponential Servers. *Operations Research,* **15**, 252-267.

[10] Hardy, G.H., Littlewood, J.E. and Polya, G. 1952. *Inequalities.* Cambridge U. Press, London.

[11] Jackson, J.R. 1963. Jobshop-Like Queueing Systems. *Management Science,* **10**, 131-142.

[12] Keilson, J. and Sumita, U. 1982. Uniform Stochastic Ordering and Related Inequalities. *Canadian Journal of Statistics,* **10**, 181-198.

[13] Liyanage, L. and Shanthikumar, J.G. 1990. Allocation through Stochastic Schur Convexity and Stochastic Arrangement Increasingness. Technical Report, School of Business Administration, University of California, Berkeley, CA 94720.

[14] Marshall, A.W. and Olkin, I. 1979. *Inequalities: Theory of Majorization and Its Applications.* Academic Press, New York.

[15] Pinedo, M. and Weber, R.R. 1984. Inequalities and Bounds in Stochastic Shop Scheduling. *SIAM Journal of Applied Mathematics,* **44**, 869-879.

[16] Righter, R. 1992. Scheduling. In: *Stochastic Orders*, M. Shaked and J.G. Shanthikumar (eds.), to appear.

[17] Rockafellar, R.T. 1970. *Convex Analysis.* Princeton U. Press, Princeton, N.J.

[18] Ross, S.M. 1983. *Stochastic Processes.* Wiley, New York.

[19] Shaked, M. and Shanthikumar, J.G. 1988. Stochastic Convexity and Its Applications. *Advances in Applied Probability*, **20**, 427-446.

[20] Shaked, M. and Shanthikumar, J.G. 1990a. Convexity of a Set of Stochastically Ordered Random Variables. *Advances in Applied Probability*, **22**, 160-167.

[21] Shaked, M. and Shanthikumar, J.G. 1990b. Parametric Stochastic Convexity and Concavity of Stochastic Processes. *Ann. Inst. Statist. Math.,* **42**, 509-531.

[22] Shaked, M. and Shanthikumar, J.G. 1990c. Optimal Allocation of Resources to Nodes of Parallel and Series Systems. Technical Report, School of Business Administration, University of California, Berkeley, CA 94720.

[23] Shanthikumar, J.G. 1987. Stochastic Majorization of Random Variables with Propotional Equilibrium Rates. *Advances in Applied Probability,* **19**, 854-872.

[24] Shanthikumar, J.G. and Yao, D.D. 1986a. The Preservation of Likelihood Ratio Ordering under Convolution. *Stochastic Processes and Their Applications*, **23**, 259-267.

[25] Shanthikumar, J.G. and Yao, D.D. 1986b. The Effect of Increasing Service Rates in Closed Queueing Networks. *Journal of Applied Probability*, **23**, 474-483.

[26] Shanthikumar, J.G. and Yao, D.D. 1988a. On Server Allocation in Multiple-Center Manufacturing Systems. *Operations Research*, **36**, 333-342.

[27] Shanthikumar, J.G. and Yao, D.D. 1988b. Second-Order Properties of the Throughput in a Closed Queueing Network. *Mathematics of Operations Research,* **13**, 524-534.

[28] Shanthikumar, J.G. and Yao, D.D. 1989. Second Order Stochastic Properties in Queueing Systems. *Proceedings of IEEE,* **77**(1), 162-170.

[29] Shanthikumar, J.G. and Yao, D.D. 1991a. Bivariate Characterization of Some Stochastic Order Relations. *Advances in Applied Probability*, **23**, 642-659.

[30] Shanthikumar, J.G. and Yao, D.D. 1991b. Strong Stochastic Convexity: Closure Properties and Applications. *Journal of Applied Probability*, **28** (1991), 131-145.

[31] Shanthikumar, J.G. and Yao, D.D. 1991c. Spatiotemporal Convexity of Stochastic Processes and Applications. *Probability in the Engineering and Information Sciences,* to appear.

[32] Stoyan, D. 1983. *Comparison Methods for Queues and Other Stochastic Models.* Wiley, New York.

[33] Topkis, D.M. 1978. Minimizing a Submodular Function on a Lattice. *Operations Research*, **26**, 305-321.