1. Stochastic Orders

We all know how to compare real variables, but do we know how to compare random variables, random vectors, and functions of random vectors?

In the two ideal (trivial?) cases $\leq_o$ and $\leq_{as}$ described below, the comparison of random variables is essentially the same as that of real variables.

**Definition 1.1.** Sample-Path Dominance ($\leq_o$): $X \leq_o Y$ if $X$ and $Y$ are defined on the same sample space $\Omega$ and $X(\omega) \leq Y(\omega)$ $\forall \omega \in \Omega$.

**Definition 1.2.** Almost Sure Sample-Path Dominance $\leq_{as}$: $X \leq_{as} Y$ if $X$ and $Y$ are defined on the same $\Omega$ and $X(\omega) \leq Y(\omega)$ $\forall \omega \in A \subset \Omega$, where $P(A) = 1$.

In general, life is not as ideal as in $\leq_o$ and $\leq_{as}$, since random variables may not be defined on the same $\Omega$, and even if they are, they may not be ordered at every sample points. In this section, we will compare random variables in more realistic settings. We would like to know that if $X \geq Y$ in some (stochastic) senses, can we say anything on $P(X \leq t) < P(Y \leq t)$ for some $t$ or for all $t$? Can we predict which of $f(X)$ and $f(Y)$, or $E[f(X)]$ and $E[f(Y)]$, is “larger” for a given function $f$? Do we need other conditions, such as a precise notion of order and some restrictions on $f$ in order to draw a concrete conclusion?

1.1. Stochastic Order $\leq_{st}$

Stochastic order $\leq_{st}$ is the most common notion of comparison of random variables. Let $F$ and $G$ be the distribution function of random variables $X$ and $Y$, respectively. The following is an inviting definition of $\leq_{st}$ for $X$ and $Y$.

**Trial Definition 1.1.1.** $X \leq_{st} Y$ if $E(X) \leq E(Y)$.

Does the above definition have nice properties? For example, if $E(X) \leq E(Y)$, will $E[f(X)] \leq E[f(Y)]$ for a wild class of functions? Will $E(X) \leq E(Y)$ shed light on $P(X \leq t) \geq P(Y \leq t)$ for some $t$? For all $t$?

A more useful definition is:

**Definition 1.1.2.** $X \leq_{st} Y$ if $F^c(t) \leq G^c(t)$ $\forall t$.

**Remark 1.1.3.** It is equally valid to say that $F \leq_{st} G$ when $X \leq_{st} Y$. Wherever applicable, we can also say that the density function or the probability mass function of one random variable is stochastically less than or equal to that of another random variable.

Stochastic ordering is a partial ordering, i.e., it is reflexive ($F \leq_{st} F$), transitive ($F \leq_{st} G, G \leq_{st} H \Rightarrow F \leq_{st} H$), and antisymmetric ($F \leq_{st} G$ and $G \leq_{st} F \Rightarrow F = G$). It is not a complete
ordering since there exist random variables (distributions) which cannot be ordered through the ordering. The antisymmetry defines stochastic equivalence: \( X \sim F \equiv Y \sim G \) if \( F = G \) (\( X \leq_{st} Y \) and \( Y \leq_{st} X \)).

There are many equivalent definitions of \( \leq_{st} \).

**Equivalent Definitions 1.1.4.**

(a). \( X \leq_{st} Y \iff F(t) \geq G(t) \forall t. \)

(b). \( X \leq_{st} Y \iff P(X \geq t) \leq P(Y \geq t) \forall t. \)

(c). \( X \leq_{st} Y \iff E[f(X)] \leq E[f(Y)] \forall \) increasing (non-decreasing) function \( f. \)

(d). **Coupling (Sample-space construction):** \( X \leq_{st} Y \iff \) there exist random variables \( \hat{X} \sim X \) and \( \hat{Y} \sim Y \) such that \( \hat{X}(\omega) \leq \hat{Y}(\omega) \forall \omega \) in the constructed sample space \( \Omega. \)

**Example 1.1.5.**

(a). If \( X \sim \exp(\lambda) \) and \( Y \sim \exp(\mu) \) such that \( \lambda < \mu \), \( X \geq_{st} Y. \)

(b). If \( X \sim \text{Poisson}(\lambda) \) and \( Y \sim \text{Poisson}(\mu) \) such that \( \lambda < \mu \), \( X \leq_{st} Y. \)

**Properties of Stochastic Order 1.1.6.**

(a). Ordering of mean values: If \( X \leq_{st} Y \) and \( E(Y) \) is well defined, \( E(X) \leq E(Y). \)

(b). Closed under transformations by increasing (decreasing) functions: If \( X \leq_{st} Y, f(X) \leq_{st} f(Y) \) for any increasing function \( f. \)

(c). Closed under scaling: \( X \leq_{st} Y \Rightarrow cX \leq_{st} cY \forall c > 0. \)

(d). Closed under shifting: \( X \leq_{st} Y \Rightarrow X+c \leq_{st} Y+c \forall c, \)

(e). Closed under positive part: The positive part of a random variable \( Z, Z^+ \equiv \max\{0, Z\}, \) and its negative part is \( Z \equiv \min\{0, Z\}. \) If \( X \leq_{st} Y, \) then \( X^+ \leq_{st} Y^+ \) and \( Y^- \leq_{st} X^- \).

(f). \( X \leq_{st} Y \Rightarrow -Y \leq_{st} -X. \)

(g). Closed under mixture: If \( (X|\Theta = \theta) \leq_{st} (Y|\Theta = \theta) \forall \theta \in \text{support of} \Theta. \) then \( X \leq_{st} Y. \)
(h). Closed under convolution: Let \( \{X_i\} \) and \( \{Y_i\} \), \( X_i \) and \( Y_i \), be two \( m \)-dimensional random vectors, with independent elements within each vector. If \( X_i \leq_{st} Y_i \) for all \( i \), \( \sum_{i=1}^{m} X_i \leq_{st} \sum_{i=1}^{m} Y_i \).

(i). Closed under convergence in distribution: Suppose that \( X_n \xrightarrow{D} X, Y_n \xrightarrow{D} Y \), and \( X_n \leq_{st} Y_n \ \forall \ n \). Then \( X \leq_{st} Y \).

Example 1.1.7. Use of stochastic orderings

Let \( D_n \) be the delay of the \( n \)th customers in a \( GI/G/1 \) queue. Suppose that the queue starts empty. It is straightforward to check that \( D_{n+1} = [D_n + S_n - T_n]^+ \), where \( S_n \) and \( T_n \) are the service time of the \( n \)th customer and the inter-arrival time between the \( (n-1) \)st and the \( n \)th customers. Consider two \( GI/G/1 \) queues \( A_1/G_1/1 \) and \( A_2/G_2/1 \) with \( A_2 \leq_{st} A_1 \) and \( G_1 \leq_{st} G_2 \). Let \( D_n^{(i)} \) be \( D_n \) in the \( i \)th queue. The properties above and recursive expression \( D_{n+1} = [D_n + S_n - T_n]^+ \) imply that \( D_2^{(1)} \leq_{st} D_2^{(2)} \). By induction, \( D_n^{(1)} \leq_{st} D_n^{(2)} \). If the queues are stable and the limiting distributions of \( D_\infty^{(i)} \)'s exist, implies that \( D_\infty^{(1)} \leq_{st} D_\infty^{(2)} \).

Exercise 1.1.8.

(a). Extension of Closed Under Convolution: Let \( \{X_i\} \) be \( m \) independent random variables and \( \{Y_i\} \) be another \( m \) independent random variables such that \( X_i \leq_{st} Y_i \). For any increasing function \( f: \mathbb{R}^m \rightarrow \mathbb{R}, f(X_1, ..., X_m) \leq_{st} f(Y_1, ..., Y_m) \).

(b). Closed under Random Sum: Let \( \{X_i\} \) be a sequence of independent random variables; \( M \) be a positive integer-valued random variable independent of \( \{X_i\} \). Let \( \{Y_i\} \) be another sequence of independent, non-negative random variables; \( N \) be a positive integer-valued random variable independent of \( \{Y_i\} \). If \( M \leq_{st} N \) and \( X_i \leq_{st} Y_i \), then \( \sum_{i=1}^{M} X_i \leq_{st} \sum_{i=1}^{N} Y_i \).

(c). Closed under Order Statistics: Let \( \{X_i\} \) be a (finite) collection of \( i.i.d. \) random variables and \( \{Y_i\} \) be another collection of \( i.i.d. \) random variables. Suppose that \( X_i \leq_{st} Y_i \). Then \( X_{(i)} \leq_{st} Y_{(i)} \), where \( X_{(i)} \) and \( Y_{(i)} \) are the order statistics of \( \{X_i\} \) and \( \{Y_i\} \), respectively.

(d). Let \( f: \mathbb{R} \rightarrow \mathbb{R} \) be such that \( x \leq f(x) \ \forall \ x \). Then \( X \leq_{st} f(X) \).

(e). \( (X \mid X \leq t) \) and \( (X \mid X > t) \) are stochastically increasing in \( t \), i.e., \( (X \mid X \leq t_1) \leq_{st} (X \mid X \leq t_2) \) and \( (X \mid X > t_1) \leq_{st} (X \mid X > t_2) \) for \( t_1 \leq t_2 \).

(f).
(i). If \( X \leq_{st} Y \) and \( E(X) = E(Y) \), then \( X =_{st} Y \).
(ii). If \( X \leq_{st} Y \) and \( E[g(X)] = E[g(Y)] \) for some strictly increasing function \( g \), then \( X =_{st} Y \).

(g). Let \( \{X_i\} \) be a collection of \( i.i.d. \) random variables and \( \{Y_i\} \) be another collection of \( i.i.d. \) random variables. Let \( M \) and \( N \) be two non-negative integer-valued random variables
such that $M$ is independent of $\{X_i\}$ and $N$ is independent of $\{Y_i\}$. Let $k$ be a positive integer.

Suppose that $M \geq_{st} kN$, and $\sum_{i=1}^{k} X_i \geq_{st} Y_1$. Then $\sum_{i=1}^{M} X_i \geq_{st} \sum_{i=1}^{N} Y_i$.

(h). Let $Y_i =_{st} X(\Theta)$. $X(\Theta_i) \leq_{st} X(\Theta_2)$ if $\Theta_1 \leq \Theta_2$. Suppose $\Theta_1 \leq_{st} \Theta_2$. Then $Y_1 \leq_{st} Y_2$.

Example 1.1.9. (Motivation for Bi-variate Characterization of Stochastic Order) $n$ jobs of independent processing times, $X_1$, ..., $X_n$, are available for processing. For any schedule (policy) $\pi$, the completion time of job $i$ is $C^\pi_i$, and the total flow time of jobs is $F_{\pi} = \sum_{i=1}^{n} C^\pi_i$.

We would like to find a policy $\pi$ that minimizes $F_{\pi}$, assuming that no new job comes, and that the processing of any job cannot be pre-empted once it has been started.

Sol. If the problem is deterministic Problem, i.e., $X_i = x_i$, $i = 1$ to $n$. The policy SEPT, shortest-expected-processing-time-first, minimizes $F_{\pi}$. When the problem is stochastic, SEPT still minimizes the mean total flow time. However, what happens if the objective is to minimize $f(F_{\pi})$, where $f$ is some pre-specified function? Can we conclude $E[f(F_{\pi})] \leq E[f(F_{\pi})]$ or compare $f(F_{\pi})$ and $f(F_{\pi})$ simply by noting $F_{\pi}$, $F_{\pi}$, and property of $f$?

To compare policies $\pi_1$ and $\pi_2$, the interchange argument basically leads to comparing functions of two random variables. So it is natural to find bi-variate characterization of stochastic orders.

By a bi-variate characterization of $\leq_{st}$, we want to find (a collection of) functions $g(X, Y)$ such that we can assert whether $X \leq_{st} Y$ from properties of $g$. Think about the deterministic case. How can we tell $x > y$ if we cannot see the values of $x$ and $y$. Suppose that we need to determine which of $x$ and $y$ is bigger just by looking at the function value of $g(x,y)$. What do we require the properties of $g$ to be? A natural choice is $g(x, y) = 0$ if $x = y$; $g(x, y) > 0$ if $x < y$; $g(x, y) < 0$ if $x > y$. Then $x \leq y \iff g(x, y) \geq 0$. Any such $g$ completely characterizes the relative magnitude of $x$ and $y$.

Can we borrow such an idea to stochastic orderings? Let $G = \{g(x, y); g(x, y) = 0$ if $x = y; g(x, y) > 0$ if $x < y; g(x, y) < 0$ if $x > y\}$. Does the following work?

Trial Definition 1. Let $g \in G$. $X \leq_{st} Y \iff$ There exist independent random variables $\hat{X}$ and $\hat{Y}$ such that $\hat{X} =_{st} X$, $\hat{Y} =_{st} Y$, and $E[g(\hat{X}, \hat{Y})] \geq 0$.

Remark 1.1.10. Stochastic order is defined through the marginal distributions of two random variables. It does not care whether the two random variables are independent or not. However, in bi-variate characterizations, the operations on $g$ may depend on the dependence of the random variables. Presently, most (all?) bi-variate characterizations are defined on independent random variables. To simply notation, in the following, we will simply put $E[g(X, Y)] \geq 0$ instead of saying there exist independent $\hat{X} =_{st} X$ and $\hat{Y} =_{st} Y$ such that
$E[g(\hat{X},\hat{Y})] \geq 0$. We can either assume that the original $X$ and $Y$ are independent, or we the construction of independent $\tilde{X} =_{st} X$ and $\tilde{Y} =_{st} Y$ is done automatically.

Does this definition work? No, it does not. For example, $g(x, y) = -x+y$ satisfies the conditions of the required function. However $E[g(X, Y)] \geq 0 \iff E(X) \leq E(Y)$, which is not sufficient to conclude $X \leq_{st} Y$.

**Trial Definition 2.** $X \leq_{st} Y \iff E[g(X, Y)] \geq 0$ for any $g \in G$.

Suppose that $E[g(X, Y)] \geq 0$ for any $g \in G$. Then for any increasing function $f$, $g(x, y) = f(y) - f(x) \in G$. $E[g(X, Y)] \geq 0 \iff E[f(X)] \leq E[f(Y)] \iff X \leq_{st} Y$. One side implication has been established. Suppose that $X \leq_{st} Y$. Can we establish $E[g(X, Y)] \geq 0$ for any $g \in G$? No, $G$ is too loose without specifying the positive and negative values of $g(x, y)$, which must bear some symmetry.

Define $G_{st} = \{g(x, y): g(x, y) - g(y, x)$ is decreasing in $x$ for all $x\}$. Note that implicitly $g(x, y) - g(y, x)$ is increasing in $y$ for all $y$.

**Definition 1.1.11.** $X \leq_{st} Y \iff E[g(X, Y)] \geq E[g(Y, X)]$ for any $g \in G_{st}$.

Proof. “$\Rightarrow$”: Let $f$ be an increasing function. Then $g(x, y) = -f(x) \in G_{st}$. $E[g(X, Y)] \geq E[g(Y, X)] \iff -E[f(X)] \geq -E[f(Y)] \iff E[f(X)] \leq E[f(Y)] \iff X \leq_{st} Y$.

“$\Leftarrow$”: Suppose that $X \leq_{st} Y$. Generate two independent samples $(\hat{X}_1, \hat{Y}_1)$ and $(\hat{X}_2, \hat{Y}_2)$ such that $\hat{X}_i \leq \hat{Y}_i$, $i = 1, 2$.

\[
\begin{align*}
g(\hat{X}_1, \hat{Y}_2) - g(\hat{Y}_2, \hat{X}_1) + g(\hat{X}_2, \hat{Y}_1) - g(\hat{Y}_1, \hat{X}_2) &= g(\hat{X}_1, \hat{Y}_2) - g(\hat{Y}_1, \hat{X}_2) + g(\hat{X}_2, \hat{Y}_1) - g(\hat{Y}_1, \hat{X}_2) \\
&= g(\hat{X}_1, \hat{Y}_2) - g(\hat{X}_1, \hat{X}_2) + g(\hat{X}_2, \hat{Y}_1) - g(\hat{Y}_1, \hat{X}_2) \\
&\quad + g(\hat{X}_2, \hat{Y}_1) - g(\hat{X}_2, \hat{X}_1) + g(\hat{X}_2, \hat{X}_1) - g(\hat{Y}_2, \hat{X}_1) \geq 0
\end{align*}
\]

$\iff E[g(X, Y)] \geq E[g(X, Y)]$. $\Box$

The current form of $G_{st}$ is not convenient to check. Are there any subsets of $G_{st}$ which characterize stochastic order and whose properties are easy to check?

Consider a function $g(x, y)$ such that $g$ is decreasing in $x$ and increasing in $y$. Let $X$ and $Y$ be independent random variables such that $X \leq_{st} Y$. Can we say something about $g(X, Y)$ and $g(Y, X)$?

Construct two independent pairs $(\hat{X}_1, \hat{Y}_1)$ and $(\hat{X}_2, \hat{Y}_2)$ such that $\hat{X}_i \leq \hat{Y}_i$, $i = 1, 2$. By construction, $g(\hat{X}_1, \hat{Y}_2) + g(\hat{X}_2, \hat{Y}_1) \geq g(\hat{Y}_1, \hat{X}_2) + g(\hat{Y}_2, \hat{X}_1)$, which implies $E[g(X, Y)] \geq E[g(Y, X)]$. Now define
Let \( G^2_{st} = \{ g(x, y) : g(x, y) \text{ is decreasing in } x \text{ and increasing in } y \} \).

Suppose that for all \( g \in G^2_{st} \), \( E[g(X, Y)] \geq E[g(Y, X)] \). Then let \( f \colon \mathbb{R} \rightarrow \mathbb{R} \) be an arbitrary increasing function. \(-f \in G^2_{st}\) and hence \(-E[f(X)] \geq -E[f(Y)] \Leftrightarrow E[f(X)] \leq E[f(Y)] \Leftrightarrow X \leq_{st} Y\). Hence, we have shown

**Definition 1.1.12.** \( X \leq_{st} Y \Leftrightarrow E[g(X, Y)] \geq E[g(Y, X)] \) for any \( g \in G^2_{st} \).

**Exercise 1.1.13.** Let \( X \) and \( Y \) be independent random variables. \( X \leq_{st} Y \Leftrightarrow g(X, Y) \geq_{st} g(Y, X) \) for any \( g \in G^2_{st} \).

**Remark 1.1.14.** We deliberately use the notation \( G^2_{st} \), because there are other bi-variate characterizations using related but different sets of functions \( G^i_{st} \). In some of these bi-variate characterizations, weaker orders between bi-variate functions, such as the increasing convex order, may imply the stochastic order between two random variables.

### 1.2. Hazard Rate Order

Hazard rate is a concept from reliability. Let \( X \) be a lifetime (i.e., nonnegative random variable) with an absolutely continuous distribution function \( F \). The hazard rate of \( X \) (or \( F \)) when \( X \) has survived \( t \) units, \( r(t) \), is defined as

\[
r(t) = \lim_{\Delta t \to 0} \frac{P(X \text{ fails within next } \Delta t \text{ units} \mid X \text{ survives } t \text{ units})}{\Delta t} = \lim_{\Delta t \to 0} \frac{P(t < X < t + \Delta t \mid X > t)}{\Delta t} = \frac{f(t)}{F^c(t)}, \quad t \geq 0.
\]

\( X \) is of increasing failure rate (IFR) if \( r(t) \) is an increasing function of \( t \). Note that \( r(t) = \frac{d[-\log F^c(t)]}{dt} \). Hence, \( X \) is of IFR \( \Leftrightarrow -\log F^c(t) \) is convex on \( \{t : F^c(t) > 0\} \).

Let \( \{r(t)\} \) and \( \{q(t)\} \) be the hazard rate function of (nonnegative) \( X \) and \( Y \), respectively.

**Definition 1.2.1.** \( X \) is smaller than \( Y \) in the hazard rate order, i.e., \( X \leq_{hr} Y \), if \( r(t) \geq q(t), t \geq 0 \).

This definition can be generalized to any two random variables with absolutely continuous distributions. In that case, \( X \leq_{hr} Y \) iff \( r(t) \geq q(t) \) for \(-\infty < t < \infty\).
Equivalent Definitions 1.2.2. (Continuous random variables) 

\( X \leq_{hr} Y \iff \)

(a). \( \frac{F^c(t)}{G^c(t)} \) is decreasing in \( t \).

(b). \( F^c(u)G^c(v) \geq F^c(v)G^c(u) \) for all \( u \leq v \).

(c). \( \frac{f(u)}{F^c(v)} \geq \frac{g(u)}{G^c(v)} \) for all \( u \leq v \).

(d). \( \frac{F^c(t + s)}{F^c(t)} \leq \frac{G^c(t + s)}{G^c(t)} \) for all \( s \geq 0 \), for all \( t \).

(e). \( P(X - t > s|X > t) \leq P(Y - t > s|Y > t) \) for all \( s \), for all \( t \).

(f). \( (X|X > t) \leq_{st} (Y|Y > t) \) for all \( t \).

(g). \( \frac{1 - FG^{-1}(1-u)}{u} \leq \frac{1 - FG^{-1}(1-v)}{v} \) for all \( 0 < u \leq v < 1 \).

Hazard rate order can be applied to non-negative discrete random variables \( X \) and \( Y \) defined on \( \{0, 1, 2, 3, \ldots\} \).

Definition 1.2.3. (Discrete random variables) \( X \leq_{hr} Y \) if

\[ \frac{P(X = n)}{P(X \geq n)} \geq \frac{P(Y = n)}{P(Y \geq n)}, \quad n \in \{0, 1, 2, \ldots\}. \]

Equivalent definitions can easily be deduced from those for continuous random variables. For example, \( X \leq_{hr} Y \iff P(X \geq n_1) P(Y \geq n_2) \geq P(X \geq n_2) P(Y \geq n_1) \) for all \( n_1 \leq n_2 \).

Properties of Hazard Rate Order 1.2.4.

(a). \( \leq_{hr} \Rightarrow \leq_{st} \).

(b). Closed under increasing functions: If \( X \leq_{hr} Y \) and \( f \) is an increasing function, then \( f(X) \leq_{hr} f(Y) \).

(c). Closed under minimization: Let \((X_i, Y_i), i = 1, \ldots, m\) be independent pairs of random variables such that \( X_i \leq_{hr} Y_i, i = 1, \ldots, m \). Then \( \min\{X_1, \ldots, X_m\} \leq_{hr} \min\{Y_1, \ldots, Y_m\} \).

(d). Closed under order statistics: Let \( X_i, i = 1, \ldots, m \) be i.i.d. random variables, and \( Y_i, i = 1, \ldots, m \), be another set of i.i.d. random variables such that \( \{X_i\} \) is independent of \( \{Y_i\} \). Suppose that \( X_i \leq_{hr} Y_i, i = 1, \ldots, m \). Then \( X_{(k)} \leq_{hr} Y_{(k)}, k = 1, \ldots, m \).
(e) Closed under convolution: Let \((X_i, Y_i), i = 1, \ldots, m,\) be independent pairs of random variables such that \(X_i \leq_{hr} Y_i, i = 1, \ldots, m.\) If all \(X_i\)'s and \(Y_i\)'s are of IFR, then \(\sum_{i=1}^{m} X_i \leq_{hr} \sum_{i=1}^{m} Y_i.\)

(f) Closed under random convolution: Let \(\{X_i\}\) be a sequence of nonnegative IFR independent random variables. Let \(M\) and \(N\) be two discrete positive integer-valued random variable such that \(M \leq_{hr} N.\) Then \(\sum_{i=1}^{M} X_i \leq_{hr} \sum_{i=1}^{N} Y_i.\)

(g) Closed under mixture: Let \(X, Y,\) and \(\Theta\) be random variables such that \([X|\Theta = \theta_1] \leq_{hr} [Y|\Theta = \theta_2]\) for all \(\theta_1\) and \(\theta_2\) in the support of \(\Theta.\) Then \(X \leq_{hr} Y.\)

How can we find out a bi-variate characterization of \(\leq_{hr}\)? It is simple to check that \(X \leq_{hr} Y \Rightarrow X \leq_{st} Y.\) So if \(G_{hr}\) is a set of functions that characterizes \(\leq_{hr},\) \(G_{st} = \{g(x, y): \Delta g(x, y) = g(x, y) - g(y, x)\) is decreasing in \(x\) for all \(x\}\) \(\subset G_{hr}.\)

A definition of \(X \leq_{hr} Y\) is that \(F'(u)G'(v) \geq F'(v)G'(u)\) for all \(u \leq v,\) which is equivalent to

\[
E[1_{\{X > u, Y > v\}}] \geq E[1_{\{X > v, Y > u\}}] \text{ for all } u \leq v.
\]

Now consider \(g(u, v) = 1_{\{x > u, y > v\}}\) for fixed \(x\) and \(y,\) \(\Delta g(u, v) = 1_{\{x > u, y > v\}} - 1_{\{x > v, y > u\}}\) for \(u \leq v.\) Check that \(\Delta g(u, v)\) is a decreasing function of \(u.\) With this result, let us conjecture that

**Definition 1.2.5.** (Bi-variate Characterization of \(\leq_{hr}\))

\(X \leq_{hr} Y \iff E[g(X, Y)] \geq E[g(Y, X)]\) for all \(g \in G_{hr} = \{g(x, y): \Delta g(x, y) = g(y, x) - g(x, y)\) is decreasing in \(x\) for all \(x \geq y\}\). \(\diamondsuit\)

Another bi-variate characterization is:

**Definition 1.2.6.** (Bi-variate Characterization of \(\leq_{hr}\))

\(X \leq_{hr} Y \iff E[g(X, Y)] \geq E[g(Y, X)]\) for all \(g \in G_{hr}^2 = \{g(x, y): g(x, y)\) is decreasing in \(x\) for all \(x \geq y\) and increasing in \(y\) for all \(y \geq x\}\). \(\diamondsuit\)
1.3. Likelihood Ratio Order \( X \leq_{lr} Y \)

Likelihood ratio order is a strong form of stochastic order.

**Definition 1.3.1.** \( X \leq_{lr} Y \Leftrightarrow \)

\[
\frac{f(t)}{g(t)} \text{ is decreasing in } t \text{ over the union of the supports of } X \text{ and } Y.
\]

**Equivalent Definition 1.3.2.** \( X \leq_{lr} Y \Leftrightarrow \)

(a). \( f(u)g(v) \geq f(v)g(u) \), for all \( u \leq v \).

(b). \( GF^{-1} \) is a convex function.

(c). \( X \leq_{lr} Y \Leftrightarrow (X|a \leq X \leq b) \leq_{st} (Y|a \leq Y \leq b) \) whenever \( a \leq b \).

**Exercise 1.3.3.** Show that \( \leq_{lr} \Rightarrow \leq_{hr} \).

**Properties of Likelihood Ratio Order 1.3.4.**

(a). Closed under increasing functions: If \( X \leq_{lr} Y \) and \( f \) is an increasing function, then \( f(X) \leq_{lr} f(Y) \).

(b). Convolution: Let \((X_i, Y_i), i = 1, \ldots, m\), be independent pairs of random variables such that \( X_i \leq_{lr} Y_i, i = 1, \ldots, m \). If \( X_i, Y_i, i = 1, 2, \ldots, m \), all have logconcave densities, except possibly one \( X_i \) and one \( Y_k, i \neq k \), then \( \sum_{i=1}^{m} X_i \leq_{lr} \sum_{i=1}^{m} Y_i \).

(c). Convolution: Let \( \{X_i\} \) be a sequence of nonnegative independent random variables with logconcave densities. Let \( M \) and \( N \) be two discrete positive integer-valued random variable such that \( M \leq_{lr} N \). Then \( \sum_{i=1}^{M} X_i \leq_{lr} \sum_{i=1}^{N} Y_i \). The assertion holds even when \( M \leq_{hr} N \).

(d). Let \( \{X_i\} \) be a sequence of nonnegative IFR independent random variables. Let \( M \) and \( N \) be two discrete positive integer-valued random variables such that \( M \leq_{lr} N \). Then \( \sum_{i=1}^{M} X_i \leq_{hr} \sum_{i=1}^{N} Y_i \).

(e). Closed under ordered statistics: Let \( X_i, i = 1, \ldots, m \), be i.i.d. random variables, and \( Y_i, i = 1, \ldots, m \), be another set of i.i.d. random variables such that \( \{X_i\} \) is independent of \( \{Y_i\} \). Suppose that \( X_i \leq_{hr} Y_j \) for all choices of \( i \) and \( j \). Then \( X_{(k)} \leq_{hr} Y_{(k)}, k = 1, \ldots, m \).
(f). Let $X$, $Y$, and $\Theta$ be random variables such that $[X|\Theta = \theta_1] \leq_{lr} [Y|\Theta = \theta_2]$ for all $\theta_1$ and $\theta_2$ in the support of $\Theta$. Then $X \leq_{hr} Y$.

(g). Consider a family of density functions $\{g_\theta, \theta \in X\}$. Let $\Theta_1$ and $\Theta_2$ be two random variables with support in $X$ and distribution functions $F_1$ and $F_2$, respectively. Let $Y_1$ and $Y_2$ be two random variables such that $Y_i = X(\Theta_i)$, $i = 1, 2$, that is, suppose that the density function of $Y_i$ is given by $h_i = \int g_\theta(y)dF_i(\theta)$, $y \in R$, $i = 1, 2$. If $X(\Theta_1) \leq_{lr} X(\Theta_2)$ whenever $\theta_1 \leq \theta_2$, and if $\Theta_1 \leq_{hr} \Theta_2$, then $Y_1 \leq_{hr} Y_2$.

Bi-variate Characterizations 1.3.5.

(a). $X \leq_{hr} Y \iff E[g(X, Y)] \leq E[g(Y, X)]$ for all $g \in G_{lr} = \{g(x, y): g(x, y) \geq g(y, x) \text{ whenever } x \geq y\}$.

(b). $X \leq_{hr} Y \iff g(X, Y) \leq_{m} g(Y, X)$ for all $g \in G_{lr}$. 
1.4. Monotone Convex Order

This order is again developed from concept in reliability. Let $X$ and $Y$ be two random lifetimes. $X$ in some sense is smaller than $Y$ if at any $t$, the expected remaining lifetime of $X$ is less than that of $Y$. Such a notion of order is formalized and is extended to any pairs of random variables. We say that $X$ is smaller than $Y$ in monotone convex order.

**Definition 1.4.1.** $X \leq_c Y (F \leq_c G)$ if $E[(X-t)^+] \leq E[(Y-t)^+] \ \forall \ t$.

The name of this order is adopted from Shaked and Shanthikuma. This order is called *mean residual life order* in Wolff, *variability order* in Ross, and *convex order* in Stoyan. Note that Shaked and Shanthikuma also define a related stochastic order, the convex order.

The monotone convex order is a *partial ordering*, i.e., it is reflexive ($X \leq_c X$), transitive ($X \leq_c Y, Y \leq_c Z \Rightarrow X \leq_c Z$), and antisymmetric ($X \leq_c Y$ and $Y \leq_c X \Rightarrow X =_s Y$).

**Proof.** The reflexivity and transitivity are obvious. To show antisymmetry, we need to use the equivalent definition 1.4.2(b) below: $\int_0^\infty F^c(u)du = \int_0^\infty G^c(u)du \ (\leq \infty) \ \forall \ t \ \Rightarrow F^c(t) = G^c(t) \ \forall \ t$.

\[
\text{Proof:} \quad \int_0^\infty F^c(u)du = \int_0^\infty G^c(u)du \ (\leq \infty) \ \forall \ t \ \Rightarrow F^c(t) = G^c(t) \ \forall \ t.
\]

The monotone convex order is not a complete order since there exist random variables (distributions) that cannot be ordered through the ordering.

**Equivalent Definitions 1.4.2.**

(a). $X \leq_c Y (F \leq_c G) \iff E[g(X)] \leq E[g(Y)]$ for any increasing convex function $g$ with well-defined moments.

**Proof.** “$\Rightarrow$”: $g(x-t)^+$ is increasing, convex.

“$\Leftarrow$”: Consider piecewise linear increasing convex function $g_n(x)$ defined by $0 = a_0 < ... < a_n < a_{n+1} = \infty$ and $0 = s_0 < ... < s_{n-1} < s_n < \infty$ such that $g_n(x) = g_n(a_j) + s_j(x - a_j)$ for $x \in [a_j, a_{j+1})$. Then $g_n(x) = \sum_{j=0}^n (s_j - s_{j+1})(x - a_j)^+$. $X \leq Y \Rightarrow E[g_n(X)] \leq E[g_n(Y)]$. For any increasing convex function $g$, we can find a sequence of $\{g_n(x)\}$ of $g(x)$.

(b). $X \leq_c Y (F \leq_c G) \iff \int_0^\infty F^c(u)du \leq \int_0^\infty G^c(u)du \ (\leq \infty) \ \forall \ t$.

**Proof.** $E[(X-t)^+] = \int_0^\infty P[(X-t)^+ > u]du = \int_0^\infty P(X > u + t)du = \int_0^\infty F^c(u)du$.

(c). $X \leq_c Y (F \leq_c G) \iff E[\max(X, t)] \leq E[\max(Y, t)] \ (\leq \infty) \ \forall \ t$.

**Proof.** $E[\max(X, t)] = t + E[\max(X-t, 0)] = t + E[(X-t)^+]$. 

11
Properties of Monotone Convex Orders 1.4.3.

(a) From 1.4.2(a), \( X \leq_c Y, X \) and \( Y \geq 0 \Rightarrow E(X^r) \leq E(Y^r) \) for \( r \geq 1 \) (whenever these expectations exist).

(b) If \( a = E(X) \), then \( a \leq_c X \).

Proof. \( g(x) = \max(x, t) \) is an increasing convex function. By Jensen’s inequality, \( E[g(X)] \geq g[E(X)] \), i.e., \( E[\max(X, t)] \geq \max(E(X), t) \geq E(X) \).

(c) \( X \leq_c X + c \ \forall \ c > 0. \)

Combining 1.4.3(b) and (c), intuitively smaller in \( \leq_c \) generally means shorter, less variable, or both.

(d) \( \leq_c \) is weaker than \( \leq_{st} \), i.e., \( X \leq_{st} Y \) and \( E[Y^+] < \infty \Rightarrow X \leq_c Y. \)

Proof. \( E[Y^+] < \infty \) guarantee \( E[Y-t]^+ < E[Y^+ + |t|] < \infty \). The rest follows since \( g(x) = (x-t)^+ \) is an increasing convex function.

(e) Closed under Shift: \( X \leq_c Y \Rightarrow X+c \leq_c Y+c \ \forall \ c. \)

(f) Closed under Scaling: \( X \leq_c Y \Rightarrow cX \leq_c cY \ \forall \ c > 0; \)

(g) Closed under Positive Part: \( X \leq_c Y \Rightarrow X^+ \leq_c Y^+. \)

Proof. \( E[\max(X^+, t)] = E[\max(\max(0, X), t)] = E[\max(\max(0, t), X)] \leq E[\max(\max(0, t), Y)] = E[\max(Y^+, t)]; \) hence follows from (6).

(h) Closed under Negation: \( X \leq_c Y \) and \( E(X) = E(Y) \Rightarrow -X \leq_c -Y. \)

Proof. \( -X = -(X-t)^+ + (X-t)^+ - (X-t) = (X-t)^+ - (X-t)^+ + (X-t)^+ = (X-t)^+ - [-(X-t)^+]. \)

(i) Closed under Negative Part: \( X \leq_c Y \) and \( E(X) = E(Y) \Rightarrow -X \leq_c -Y. \)

Proof. \( E[\max(X, t)] = E[\max(\max(-X, 0), t)] = E[\max(-X, \max(0, t))]. \)

(j) Closed under Mixture: \( (X|\Theta = \theta) \leq_c (Y|\Theta = \theta) \ \forall \ \theta \in \text{support of } \Theta \Rightarrow X \leq_c Y. \)

Proof. \( E(X) = E[E(X|\Theta)] \leq E[E(X|\Theta)] = E(Y). \)

(k) Closed under Convolution: Let \( \{X_i\} \) and \( \{Y_i\} \), \( X_i \sim F_i \) and \( Y_i \sim G_i \), be two \( m \)-dimensional random vectors, with independent elements within each vector. \( X_i \leq_c Y_i \) \( F_i \leq_c G_i \) \( \Rightarrow \sum_{i=1}^m X_i \leq_c \sum_{i=1}^m Y_i. \)
Proof. It suffices to prove for two pairs of r.v.’s. Assume that all r.v.’s are independent. From closure under shift, \((X_1 + X_2) \leq_c (Y_1 + X_2)\), which, by closure under mixture, gives \((X_1 + X_2) \leq_c (Y_1 + X_2)\). By the same token, \((Y_1 + X_2) \leq_c (Y_1 + Y_2)\); and result follows transitivity of \(\leq_c\). Note that the marginal distributions of the convolution remain unchanged even if the two vectors are dependent.

(l). **Closed under Convergence in Distribution:** Suppose that \(X_i \xrightarrow{d} X\), \(Y_i \xrightarrow{d} Y\), \(X_i \leq_c Y_i \forall i\), \(E(X_i^+) \rightarrow E(X)\), and \(E(Y_i^+) \rightarrow E(Y)\). Then \(X \leq_c Y\).

**Proof.** \(E[\max(X_i, t^+)] = E(X_i^+) - \int_0^t F_i^c(u)du\). By bounded convergence, \(\int_0^t F_i^c(u)du \rightarrow \int_0^t F^c(u)du\).

Note. The conditions \(E(X_i^+) \rightarrow E(X)\) and \(E(Y_i^+) \rightarrow E(Y)\) are necessary. Consider \(X_i = 1\); \(Y_i = 0\) w.p. \((i-1)/i\), and \(= i\) w.p. \(1/i\). Then \(X_i \leq_c Y_i\), but \(X \not\leq_c Y\).

(m). **Closed under Transform by increasing convex function:** \(X \leq_c Y\) and \(g\) is an increasing convex function \(\Rightarrow g(X) \leq_c g(Y)\).

(n). Let \(\{X_i\}\) be \(m\) independent r.v.’s and \(\{Y_i\}\) be another \(m\) independent r.v.’s; \(X_i \leq_c Y_i\). \(\psi: \mathbb{R}^m \rightarrow \mathbb{R}\) being an increasing and componentwise convex function. Then \(\psi(X_1, ..., X_m) \leq_{st} \psi(Y_1, ..., Y_m)\).

**Proof.** It suffices to show the case for \(m = 2\). Just like proving (14), we can assume that all r.v.’s are independent. \(\psi(\cdot, x)\) is an increasing function of \(x\). By (16), \([\psi(x_1, X_2) | X_1 = x_1] \leq_c [\psi(x_1, Y_2) | X_1 = x_1]\); by (11), \(\psi(X_1, X_2) \leq_c \psi(X_1, Y_2)\). A similar argument gives \(\psi(X_1, Y_2) \leq_c \psi(Y_1, Y_2)\), and the result follows from the transitivity of \(\leq_c\).

**Example 1.4.4.**

(a). If \(X \sim \text{normal}(\mu_1, \sigma_1^2), Y \sim \text{normal}(\mu_2, \sigma_2^2)\), \(\mu_1 \leq \mu_2, \sigma_1 \leq \sigma_2\), then \(X \leq_c Y\).

(b). GI/G/1 Queue (Wolff, pp 489) Refer to two GI/G/1 queues \(A_1/G_1/1\) and \(A_2/G_2/1\) in the example for \(\leq_{st}\) (Example 1.1.7). Suppose that \(A_2 \leq_c A_1\) and \(G_1 \leq_c G_2\). Let \(D_n^{(i)}\) be \(D_n\) in the \(i\)th queue. As in Example 1.1.7, \(D_2^{(i)} \leq_c D_2^{(2)}\). By induction, \(D_n^{(i)} \leq_c D_n^{(2)}\). If the queues are stable and the limiting distributions of \(D_\infty^{(i)}\)’s exist, we can show that that \(D_\infty^{(1)} \leq_c D_\infty^{(2)}\).

(c). (Ross, pp 273) A random variable \(X\) is **new better than used in expected (NBUE)** if \(E[X-t|X > t] \leq E(X)\) for all \(t\). Suppose that \(X \sim \text{NBUE}\) and \(E(X) = 1/\mu\). Then \(X \leq_c Y \sim \text{exp}()\).
Solution. \( X \) is NBUE \( \implies \mathbb{E}[X-a|X>a] \leq \mathbb{E}(X) \ \forall \ a, \) i.e., \[ \int_{a}^{\infty} \frac{F^c(y)}{F^c(a)} dy \leq 1/\mu \iff \mu \leq \frac{F^c(a)}{\int_{a}^{\infty} F^c(y) dy} . \]

Integrating both sides w.r.t. \( a \) from 0 to \( s \), and changing variables by defining \( x = \int_{a}^{\infty} F^c(y) dy \), we have, \( s\mu \leq \int_{0}^{s} F^c(y) dy - \frac{1}{x} dx \iff s\mu \leq \ln \left[ \frac{1}{\mu} \int_{0}^{s} F^c(y) dy \right] \), which, after some algebra, gives

\[ \int_{s}^{\infty} F^c(y) dy \leq e^{-\mu s}/\mu = \int_{s}^{\infty} e^{-x\mu} dx. \]

(d). Combining the results from Example 2 and 3 (Ross, pp 275) Consider renewal arrivals to a stable single-server queue. If the inter-arrival distribution is NBUE with mean \( 1/\lambda \), then the mean delay \( d \leq \lambda \mathbb{E}(S^2)/[2(1-\lambda \mathbb{E}(S))] \), where \( S \) is the service distribution. The result follows since for Poisson arrivals with rate \( \lambda \), the mean delay is \( \lambda \mathbb{E}(S^2)/[2(1-\lambda \mathbb{E}(S))] \).

\[ \Box \Box \]

Exercise 1.4.5.

(a). *Closed under Random Sum:* Let \( \{X_i\} \) be \( m \) independent r.v.’s and \( \{Y_i\} \) be another \( m \) independent r.v.’s; \( M \) and \( N \) be two positive integer-valued r.v.’s such that \( M \leq_c N \). Then \( \sum_{i=1}^{M} X_i \leq_c \sum_{i=1}^{N} Y_i \).

(b). Suppose that \( X \leq_c Y \) and \( \mathbb{E}(X) = \mathbb{E}(Y) \). Show that \( \mathbb{E}[g(X)] \leq \mathbb{E}[g(Y)] \ \forall \) convex function \( g \).

Note. It follows that \( X \leq_c Y \) and \( \mathbb{E}(X) = \mathbb{E}(Y) \implies \mathbb{E}(X^r) \leq \mathbb{E}(Y^r) \) for \( r = 2, 4, \ldots \).

(c). A sufficient condition for \( \leq_c \): We say that \( X \sim F \) and \( Y \sim G \) satisfy the cut criterion if \( \mathbb{E}(X) \leq \mathbb{E}(Y) < \infty \) and there exists \( t_0 \) \( (< \infty) \) such that \( F(t) \leq G(t) \ \forall \ t < t_0 \) and \( F(t) \geq G(t) \ \forall \ t > t_0 \). It is possible that there exists an nondegenerate interval \( I \) such that \( F(t) = G(t) \ \forall \ t \in I \). Suppose that \( X \) and \( Y \) satisfy the cut criterion. Then \( X \leq_c Y \).

(d). Comparing Renewal Process (Ross, pp 257) Let \( \{M(t)\} \) and \( \{N(t)\} \) be renewal processes generated from inter-arrival distributions \( F \) and \( G \), respectively. If \( F \) is NBUE (NWUE) and \( G \) is exponential with the same mean as \( F \). Then \( M(t) \leq_c N(t) \).

\[ \Box \Box \]
1.5. The Notion of Regular Stochastic Convexity

Let \( \{P_\theta; \theta \in \Theta\} \) be a family of uni-variate distributions, and \( \{X(\theta); \theta \in \Theta\} \) be a set of random variables such that \( X(\theta) \sim P_\theta \). A lot of times we are interested in the monotone properties (increasing or decreasing), the second-order properties (convex or concave), or both the monotone and second-order properties of \( X(\theta) \) in some stochastic senses as \( \theta \) increases. As a result, we will mention the following concepts: stochastic increasing (SI), stochastic decreasing (SD), stochastic convex (SCV), stochastic concave (SCV), stochastic increasing convex (SICX), stochastic increasing concave (SICV), stochastic decreasing convex (SDCX), stochastic decreasing concave (SDCV), stochastic increasing linear (SIL), and stochastic decreasing linear (SDL).

Example 1.5.1. Exponential Tandem Queue Under Communication Blocking

(a) \[
\infty \text{ customers} \quad 1 \quad b_1 \quad 2 \quad b_2 \quad \cdots \quad b_{M-1} \quad M
\]

Given \( \mu_1, \ldots, \mu_M \), the throughput \( TH \) is a function of \( \theta = (b_1, \ldots, b_M) \). Knowing the condition under which \( TH(\theta) \) is concave with respect to \( \theta \) is useful in designing the line.

(b) \[
\infty \text{ customers} \quad 1 \quad 2 \quad b_2 \quad \cdots \quad M-1 \quad M
\]

Given \( b_1, \ldots, b_M \), and \( \Sigma \mu_i = c > 0 \). Will \( TH \) is a concave function of \( \theta = (\mu_1, \ldots, \mu_M) \)?

There are different notions of stochastic convexity that look very similar. One must be careful to differentiate one notion from another, particularly on how one notion implies (is stronger than) another one. Here we only show one notion, the regular stochastic convexity. Many performance measures of stochastic models, e.g., the throughput of a queueing network, can be expressed in terms of this notion. It is by no means the easiest to prove, and in fact even the regular stochastic convexity is sometime established by stronger notions. The notion is handy summarized by the following table.

<table>
<thead>
<tr>
<th>( {X(\theta)} )</th>
<th>( \rightleftharpoons )</th>
<th>( \cdot ) and</th>
<th>( \phi )</th>
<th>( \text{if} )</th>
<th>( E[\phi(X(\theta))] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>SI (SD)</td>
<td>-</td>
<td>inc.</td>
<td>inc. (dec.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCX (SCV)</td>
<td>-</td>
<td>convex (concave)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SICX (SICV)</td>
<td>SI</td>
<td>inc. convex (inc. concave)</td>
<td>inc. convex (inc. concave)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDCX (SDCV)</td>
<td>SD</td>
<td>inc. convex (inc. concave)</td>
<td>dec. convex (dec. concave)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIL</td>
<td>SI</td>
<td>increasing { convex concave }</td>
<td>increasing { convex in ( \theta ) concave in ( \theta ) }</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDL</td>
<td>SD</td>
<td>increasing { convex concave }</td>
<td>decreasing { convex in ( \theta ) concave in ( \theta ) }</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The table is read in the following way: \{X(\theta)\} is \textit{stochastically increasing} (SI) in \theta if for any increasing function \(\phi\), \(E[\phi(X(\theta))]\) is an increasing function in \(\theta\). \{X(\theta)\} is \textit{stochastically decreasing convex} (SDCX) in \theta if \{X(\theta)\} is SD and for any increasing convex function \(\phi\), \(E[\phi(X(\theta))]\) is decreasing convex in \(\theta\). Other items of the table are interpreted in a similar way.

**Equivalent Definitions 1.5.2.**

(a). \(\{X(\theta)\}\) is SICX (SICV) if, and only if \(\{X(\theta)\}\) is SI and \(\int_{x}^{\infty} F_{X}^{\ast}(s, \theta)ds\) is increasing (decreasing) convex in \(\theta\) for all \(x\).

(b). \(\{X(\theta)\}\) is SDCX (SDCV) if, and only if \(\{X(\theta)\}\) is SD and \(\int_{x}^{\infty} F_{X}^{\ast}(s, \theta)ds\) is decreasing (decreasing) convex in \(\theta\) for all \(x\).

**Properties of regular stochastic convexity 1.5.3.**

(a). \textit{Closed under addition}: Let \(\{X(\cdot)\}\) and \(\{Y(\cdot)\}\) be two independent processes. If \(\{X(\theta)\} \{Y(\theta)\}\) is SICX (SICV), then \(\{X(\theta)\}+\{Y(\theta)\}\) is SICX (SICV).

(b). \textit{Closed under composition}: If \(\{X(\theta)\}\) and \(\{Y(\lambda)\}\) be SICX (SICV, SIL), then \(\{Y(X(\theta))\}\) is SICX (SICV, SIL). If \(\{X(\theta)\}\) is SDCX (SDCV, SDL) and \(\{Y(\lambda)\}\) is SICX (SICV, SIL), then \(\{Y(X(\theta))\}\) is SDCX (SDCV, SDL).

Read Meester and Shanthikumar [1990] for a stronger notion of stochastic convexity.

Tail Properties of (Non-negative) Distributions

Let \( X \in [0, b] \sim F \) (\( b \) can be \( \infty \)). The remaining life of \( X \) at age \( t \) \( X_t \equiv (X-t|X > t) \). The mean residual life of \( X \) at age \( t \)

\[
E(X_t) = \int_0^t P(X > u|X > t)du = \int_0^t P(X-t > u|X > t)du = \int_u^\infty F^c(\gamma)d\gamma / F^c(t).
\]

(15)

\( X \) has bounded mean residual life (BMRL) if there exists a \( \gamma \) such that \( E(X_t) \leq \gamma \) for \( t < b \). If \( \gamma = E(X) \), then \( X \) has the property new better than used in expectation (NBUE). Similarly, \( X \) has the property new worse than used in expectation (NWUE) if \( E(X_t) \geq E(X) \forall t < b \). Similarly,

Example 6.
Erlang r.v.’s are NBUE; hyper-exponential r.v.’s are NWUE; exponential r.v.’s are both NBUE and NWUE. ♦

If \( X_t \leq \text{st} \ X \forall t < b \), then \( X \) has the property new better than used (NBU); if \( X_t \geq \text{st} \ X \forall t < b \), then \( X \) has the property new worse than used (NWU). An equivalent definition of NBU (NWU) is

\[ X \text{ is NBU (NWU)} \iff -\ln F^c \text{ is superadditive (subadditive) on } (0, \infty), \]

i.e., \( F^c(s+t) \leq (\geq) F^c(s) F^c(t) \forall s, t > 0 \)

(16)

Proof. Directly from definition of NBU (NWU). ♦

\( X \) has the property decreasing mean residual life (DMRL) if \( X \) is NBUE and \( E(X_t) \) is decreasing in \( t \forall t < b \). \( X \) has the property increasing mean residual life (IMRL) if \( X \) is NWUE and \( E(X_t) \) is increasing in \( t \forall t < b \).

Let \( F \) and \( f \) be the distribution and density functions of a continuous r.v. \( X \). \( \lambda(t) \equiv \lim_{\Delta t \to 0} \frac{P(X \in (t,t+\Delta t]}{f(t)} \) be the failure (hazard) rate of \( X \) at age \( t \) \( \geq 0 \). \( X \) has increasing failure rate (IFR) if \( \lambda(t) \) is increasing \( \forall t < b \); \( X \) has decreasing failure rate (IFR) if \( \lambda(t) \) is decreasing \( \forall t < b \).

The failure rate function completely determine the distribution of \( X \), since

\[
F^c(t) = e^{-\int_0^t \lambda(u)du}.
\]

(17)

Proof. \( \int_0^t \lambda(u)du = \int_0^t \frac{f(u)}{F^c(u)}du = -\ln F^c(t) \). ♦

An equivalent definition of IFR (DFR) is

\[ X \text{ has IFR (DFR)} \iff X_t \geq \text{st} X_t \text{ if } s \leq t < b \text{ } (X_t \leq \text{st} X_t \text{ if } s \leq t < b). \]

(18)

Proof. \( X_s \geq \text{st} X_t \iff P(X-s|X > s) \geq P(X-t|X > t) \iff F^c(s+v)/F^c(s) \geq F^c(t+v)/F^c(t) \forall v > 0 \iff e^{-\int_0^t \lambda(u)du} \geq e^{-\int_0^s \lambda(u)du} \forall v > 0 \). ♦

It is clear that from (17) and definitions above that
IFR ⇒ (NBU and DMRL) ⇒ NBUE (DFR ⇒ (NWU and IMRL) ⇒ NWUE). \hspace{10mm} (19)

We can show that NBU and DMRL (NWU and IMRL) do not imply each other.

Exercise 8. Give example to show that NBU and DMRL (NWU and IMRL) do not imply each other.

Another equivalent definition of IFR (DFR), which follows directly form differentiating $-\ln F^c$, is

\[ X \text{ has IFR (DFR)} \iff -\ln F^c \text{ is convex (concave) on } \{ t \mid F^c > 0 \}. \hspace{10mm} (20) \]

When $X$ and $Y$ are independent r.v.’s with IFR ⇒ min$(X, Y)$ has IFR. \hspace{10mm} (21)

\textit{Proof}. Straightforward calculation shows that failure rate of min$(X, Y) = \text{sum of failure rates of } X \text{ and } Y.$ ♦

IFR is closed under convolution, i.e.,

\[ X \text{ and } Y \text{ independent r.v.’s with IFR ⇒ } X+Y \text{ is IFR.} \hspace{10mm} (22) \]

The proof is a bit involved. See Barlow and Proshan, pp 100, or Shaked and Shanthikuma, pp 23.

DFR is closed under mixture. Let $X$ and $\Theta$ be two r.v.’s;

\[(X|\Theta=\theta) \text{ is DFR } \forall \theta \Rightarrow X \text{ is DFR.} \hspace{10mm} (23) \]

\textit{Proof}. Let $f$ and $f_\theta$ be the density function of $X$ and $(X|\Theta=\theta)$, respectively. $\lambda(t) = f(t)/F^c(t) = E(f_\theta)/F^c(t)$. By strict forward differentiation, $\frac{d\lambda(t)}{dt} = \frac{[\lambda(t)E(df_\theta/dt) + E^2(f_\theta)]}{[F^c(t)]^2}$. The failure rate is decreasing $\iff F^c(t)E(df_\theta/dt)+E^2(f_\theta) \leq 0 \iff E^2(f_\theta) \leq E(F^c_\theta)E(-df_\theta/dt)$. Cauchy-Schwartz inequality states that $E^2(XY) \leq E(X^2)E(Y^2)$, i.e., $E(F^c_\theta)E(-df_\theta/dt) \geq E^2(\sqrt{(-df_\theta/dt)(F^c_\theta)})$. Since $(X|\Theta=\theta)$ has decreasing failure rate, $-df_\theta/dt F^c_\theta \geq f_\theta^2$, which establish our result. ♦

Example 7.
From exponential r.v.’s, we see that IFR is not closed under mixture, while DFR is not closed under convolution.
Stochastic Ordering of Random Vectors and Stochastic Processes

Let \( X = (X_1, ..., X_m) \) and \( Y = (Y_1, ..., Y_m) \) be two random vectors.

**Definition.** \( X \leq_{st} Y \Leftrightarrow E[g(X)] \leq E[g(Y)] \) \( \forall \) increasing \( g: \mathbb{R}^m \to \mathbb{R} \). (24)

**Definition.** The stochastic processes \( \{X(t); t \geq 0\} \leq_{st} \{X(t); t \geq 0\} \)

\( \Leftrightarrow \{X(t_1), ..., X(t_m)\} \leq_{st} \{Y(t_1), ..., Y(t_m)\} \) \( \forall \) \( m, t_i \geq 0 \). (25)

**Example 8.** Comparing Renewal Process (Ross, pp 257)
Let \( \{M(t)\} \) and \( \{N(t)\} \) be renewal processes with interarrival distribution \( F \) and \( G \), respective. If \( F \leq_{st} G \), then \( \{M(t)\} \geq_{st} \{N(t)\} \).

**Solution.** Generate \( X_i \sim F \) and \( Y_i \sim G \) such that \( X_i \leq Y_i \). For all \( t \), \( M(t) \geq N(t) \).

**Example 9.** Stochastic Monotonicity for a Birth and Death Process (B&D) (Ross, pp 257)
Let \( \{X(t); t \geq 0\} \) be a B&D. Then \( \{X(t)\} \) is stochastically increasing in \( X(0) \).

**Solution.** Let \( \{X_1(t)\} \) and \( \{X_2(t)\} \) be two B&D’s with the same infinitesimal generators such that \( X_1(0) = i < j = X_2(0) \). We will construct a process \( \{X_3(t)\} =_{st} \{X_2(t)\} \) and \( X_3(t) \geq X_1(t) \) \( \forall \) \( t \). Define \( T \equiv \) the first time that \( X_1(t) = X_2(t) \); \( T \equiv \infty \) if \( X_1(t) \neq X_2(t) \) \( \forall \) \( t \). Construct \( X_3(t) \equiv X_2(t) \) for \( t < T \), and \( X_1(t) \) for \( t \geq T \). By the structure a B&D process, if \( T = \infty \), the states of \( X_1(t) \) and \( X_2(t) \) never cross; \( X_2(t) \geq X_1(t) \) and we are done. For finite \( T \), the Markovian property guarantees that \( \{X_3(t)\} \) has the same infinitesimal generator as \( X_1(t) \) and \( X_2(t) \).

**Exercise 9.** (Ross, pp 258)
For a B&D process, show that \( P(X(t) \geq j | X(0) = 0) \) is increasing in \( t \) \( \forall \) \( j \).

**Note.** This result gives an upper bound on \( P(X(t) \geq j | X(0) = 0) \).

**Example 10.** Exponential Convergence in Discrete-Time Markov Chain (DTMC) (Ross, pp 258)
For an irreducible finite-state DTMC, the rate of convergence of the \( n \)-step transition probability \( p^n_{ij} \) to the stationary distribution \( \pi_j \) is exponentially fast.

**Solution.** Suppose that \( P(X_0 = i) = 1 \), and the chain has \( N \) states. We will show that \( p^n_{ij} \to \pi_j \) in exponential rate.

First construct a process \( \{Y_n\} \) which has the same transition probability matrices as \( \{X_n\} \). Set \( \{Y_0\} \) (and hence \( \{Y_n\}\) ) \( \sim \) \( \pi \). Let \( T = \inf\{n; X_n = Y_n\} \). Since \( \{T > mK\} \subset \{X_K \neq Y_K, X_{2K} \neq Y_{2K}, ..., X_mK \neq Y_{mK}\} \); \( P(T > mK) \leq P(X_K \neq Y_K)P(X_{2K} \neq Y_{2K}|X_K \neq Y_K)...P(X_{mK} \neq Y_{mK}|X_K \neq Y_K, ..., X_{(m-1)K} \neq Y_{(m-1)K}) \). Observe that for an irreducible finite-state DTMC \( \{X_n\} \), there exists a \( K \) such that \( p^n_{ij} > 0 \) for all \( i \) and \( j \) (and hence \( p^n_{ij} > 0 \) \( \forall \) \( n \geq K \)). Choose an \( \epsilon > 0 \) such that \( p^n_{ij} \geq \epsilon \) \( \forall \) \( i, j \). So in \( K \) transitions, \( P(X_K = j) \) and \( P(Y_K = j) > \epsilon \); \( P(X_K = Y_K) > \epsilon \); \( P(X_K \neq Y_K) \leq (1- \epsilon)^2 \). By the same argument, \( P(X_{mK} \neq Y_{mK}|\cdot) \leq (1- \epsilon)^2 \); \( P(T > mK) \leq (1-\alpha)^m \), and hence \( T \) is a finite r.v.
Now define a second process \( \{Z_n\} \) such that \( Z_n \equiv Y_n \) for \( n \leq T \), and \( Z_n \equiv X_n \) for \( n > T \). It is clear that \( \{Z_n\} \) is a DRMC with the same transition probability matrices as \( \{X_n\} \) and \( \{Y_n\} \). By our initial setting, \( Z_0 \), and hence \( \{Z_n\} \), \( \sim \pi \). 

\[
p_{ij}^n = P(X_n = j) = P(X_n = j \mid T \leq n) P(T \leq n) + P(X_n = j \mid T > n) P(T > n);
\pi_j = P(Z_n = j) = P(Z_n = j \mid T \leq n) P(T \leq n) + P(Z_n = j \mid T > n) P(T > n)
\]

which means that \( |p_{ij}^n - \pi_j| \leq P(T > n) \leq (1 - \alpha)^{n/k-1} \). ♦

**Note.** The uniqueness of stationary distributions also follows from the argument.