Summer 7-25-2013

ENTROPY APPLICATIONS IN INDUSTRIAL ENGINEERING

Saeed Zamiri Marvizadeh
University of Nebraska-Lincoln, szamiri@unl.edu

ENTROPY APPLICATIONS
IN
INDUSTRIAL ENGINEERING

By

Saeed Zamiri Marvizadeh

A DISSERTATION

Presented to the Faculty of
The Graduate College at the University of Nebraska
In Partial Fulfillment of Requirements
For the Degree of Doctor of Philosophy

Major: Engineering

Under the Supervision of Professor Fred Choobineh

Lincoln, Nebraska

July, 2013

ENTROPY APPLICATIONS
IN
INDUSTRIAL ENGINEERING

Saeed Zamiri Marvizadeh, Ph.D.
University of Nebraska, 2013
Advisor: Fred Choobineh

Entropy is a fundamental measure of information content which has been applied in a wide variety of fields. We present three applications of entropy in the industrial engineering field: dispatching of Automatic Guided Vehicles (AGV), ranking and selection of simulated systems based on the mean performance measure, and comparison between random variables based on cumulative probability distributions.

The first application proposes three entropy-based AGV dispatching algorithms. We contribute to the body of knowledge by considering the consequence of potential AGV moves on the load balance of the factory before AGVs are dispatched. Kullback-Leibler directed divergence is applied to measure the divergence between load distribution after each potential move and load distribution of a balanced factory. Simulation experiments are conducted to study the effectiveness of suggested algorithms.

In the second application, we focus on ranking and selection of simulated systems based on the mean performance measure. We apply maximum entropy and directed divergence principles to present a two stage algorithm. The proposed method contributes to the ranking and selection body of knowledge because it relaxes the normality assumption for the underlying population which restricts the frequentist algorithms, it

does not assume any priori distribution which is assumed by bayesian approaches, and finally it provides ranking of systems based on their observed performance measures.

Finally, we present an entropy-based criterion for comparing two alternatives. Our comparison is based on directed divergence between alternatives' cumulative probability distributions. We compare the new criterion with stochastic dominance criteria such as first order stochastic dominance (FSD) and second order stochastic dominance (SSD). Since stochastic dominance rules may be unable to detect dominance even in situations when most decision makers would prefer one alternative over another, our criterion increases the probability of identifying the best system and reduces the probability of obtaining the nondominance set in such situations. Among two alternatives, we show that if one alternative dominates the other one by SSD, the dominating alternative will be dominated by our new criterion. In addition, we show that the probability associated with our new criterion is consistent with the probability corresponding to p almost stochastic dominance (p-AFSD).

List of Figures

List of Tables

Table of Contents

Chapter 1

Introduction

The concept of *entropy* was introduced by Claude E. Shannon in his 1948 paper "A Mathematical Theory of Communication". Wikipedia defines *entropy* as "*a measure of the uncertainty associated with a random variable. In this context, the term usually refers to the Shannon entropy, which quantifies the expected value of the information contained in a message, usually in units such as bits and a 'message' means a specific realization of the random variable. Equivalently, the Shannon entropy is a measure of the average information content one is missing when one does not know the value of the random variable*."

Entropy laid the foundation for a comprehensive understanding of communication theory and according to Kapur and Kesavan (1992), the introduction of Shannon entropy can be considered as one of the most important breakthroughs over the past fifty years in the literature on probabilistic uncertainty. The concept of entropy has been applied in a wide variety of fields such as statistical thermodynamics, urban and regional planning, business, economics, finance, operations research, queueing theory, spectral analysis, image reconstruction, biology and manufacturing which will be reviewed in the next chapter. In this chapter, entropy and two related concepts, maximum entropy and directed divergence, are reviewed.

1-1. Overview of the entropy concept

Entropy of a system can be described in different ways. The original idea was born from classical thermodynamics. Classical thermodynamics was developed during the 19th century and its primary architects were Sadi Carnot, Rudolph Clausius, Benoit Claperyon, James Clerk Maxwell and William Thomson. However, it was Clausius who first explicitly advanced the idea of entropy. The concept was then expanded by Maxwell. The specific definition which comes from Clausius, is shown in equation (1-1) and interprets the entropy, E, as the quantity of heat, Q, that is absorbed in a reversible system when temperature is T.

$$E = \frac{Q}{T} \tag{1-1}$$

As long as temperature is constant, it is simple enough to differentiate equation (1-1) and derive (1-2):

$$\Delta E = \frac{\Delta Q}{T} \tag{1-2}$$

Here Δ represents a finite increment, i.e. ΔE indicates a "change" or "increment" in E, as in $\Delta E = E_2 - E_1$, where E_1 and E_2 are entropies of two different states.

Clausius and the others, especially Carnot, were much interested in the ability to convert mechanical work into heat energy and vice versa. Hence they interpreted entropy as the amount of energy in a system that is unavailable to do work. They arrived at equation (1-3) where ΔU is the energy input to the system, and ΔW is the part of that energy which goes into doing work.

$$\Delta E = \frac{\Delta \mathrm{U} - \Delta \mathrm{W}}{T} \tag{1-3}$$

While physicists were laying the foundations for classical thermodynamics, chemists interpreted entropy in chemical reactions. Their real interest in entropy was to predict whether or not a given chemical reaction will take place. They defined entropy as equation (1-4) in which H is the enthalpy and F is the free energy (usually known as Gibb's free energy).

$$\Delta E = \frac{\Delta \mathrm{H} - \Delta \mathrm{F}}{T} \qquad 1\text{-}4$$

In the later 1800's, Maxwell, Ludwig Boltzmann and Josiah Willard Gibbs, through the new *molecular theory*, extended the ideas of classical thermodynamics to a new domain called *statistical mechanics* in which each system possesses macro-states and micro-states. For example, the temperature of a system defines a macro-state, while the kinetic energy of each molecule in the system defines its micro-state. Equation (1-5), first derived by Ludwig Boltzmann, is the general form of entropy in statistical mechanics where p_i is the probability that the i^{th} particle be in a given micro-state and all p_i's are evaluated for the same macro-state. k is an arbitrary constant, and in thermodynamics is the Boltzmann constant which is 1.380658×10^{-23}.

$$E = -k \sum p_i \ln p_i \qquad 1\text{-}5$$

Mathematical foundations of statistical mechanics are applicable to any statistical system, regardless of its status as a thermodynamic system. As an example (Abbas, 2006) consider tossing a die twice. The sum of two throws is considered as a macro-state of this system and the possible realizations can be considered as micro-states. In this case we have 11 macro-states (2,3,4,5,6,7,8,9,10,11,12) and 36 microstates. Figure 1-1 shows the possible micro-states of each macro-state.

2	3	4	5	6	7	8	9	10	11	12
(1,1)	(1,2)	(3,1)	(2,3)	(1,5)	(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)
	(2,1)	(1,3)	(3,2)	(5,1)	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	
		(2,2)	(1,4)	(2,4)	(2,5)	(3,5)	(4,5)	(5,5)		
			(4,1)	(4,2)	(5,2)	(5,3)	(5,4)			
				(3,3)	(3,4)	(4,4)				
					(4,3)					

Figure 1-1. Possible micro-states of each macro-state in die tossing experiment shown in Abbas (2006)

Entropy of each macro-state is defined to be proportional to the logarithm of the number of its micro-states. In addition, the number of micro-states of a macro-state is directly related to its probability of occurrence. Hence for a given macro-state, the entropy is a measure of the probability of its occurrence. Also the entropy of a system is the sum of its macro-state entropies. Shannon, used this conclusion as the basis for excursion of entropy into the new domain, information theory. He realized that entropy can be applied to quantify the uncertainty of a probability distribution, $P = (p_1, p_2, \dots, p_n)$. He first thought of the properties that a measure, $H(p_1, p_2, \dots, p_n)$, for quantifying the uncertainty of a probability distribution should have and prospected the following properties:

1) It should depend on all probabilities $p_1, p_2, \dots, p_n$. In the other words it should be a function of all $p_1, p_2, \dots, p_n$.
2) It should be a continuous function of $p_1, p_2, \dots, p_n$.
3) If $\mathrm{p_1, p_2, \dots, p_n}$ are reordered it should not change. This means that this measure should be permutationally symmetric. This property is desirable since the labeling of the outcomes should not affect the entropy.
4) $\mathrm{H}(\frac{1}{\mathrm{n}}, \frac{1}{\mathrm{n}}, \dots, \frac{1}{\mathrm{n}})$ should be a monotonic increasing function of n.
5) $\mathrm{H(p_1, p_2, \dots, p_n) = H(p_1 + p_2, p_3, \dots, p_n) + (p_1 + p_2)H}(\frac{\mathrm{p_1}}{\mathrm{p_1+p_2}}, \frac{\mathrm{p_2}}{\mathrm{p_1+p_2}})$.

Based on these desired properties, Shannon arrived at equation (1-6), which is exactly the Boltzmann entropy, and pointed out that this is the only measure which satisfies above properties:

$$H(p_1, p_2, \dots, p_n) = -k \sum_{i=1}^{n} p_i \ln p_i \qquad 1\text{-}6$$

where k is an arbitrary positive constant, which satisfies all the properties.

He not only proposed this measure, but also proved the theorem that this was the only function of $p_1, p_2, \dots, p_n$, which had all these properties. In other words, he showed that these properties characterized this measure.

In addition to the properties illustrated by Shannon, his measure also possesses some properties which were not initially intended. These additional properties are as follows (see Aczel and Daroczy (1975) and Mathai and Rathie (1975) for these properties and their corresponding proofs):

Property 1) *The entropy value does not change by adding an impossible event (an event with zero probability).*

Property 2) *When this measure is maximized subject to some linear constraints, the maximizing probabilities are all non-negative.*

Property 3) *Its value is always positive.*

Property 4) *Its value is minimum when* $p_1, p_2, \dots, p_n$ *is a degenerate distribution.*

Property 5) *It is a concave function of* $p_1, p_2, \dots, p_n$ *hence its local maximum will also be a global maximum.*

Property 6) *Its maximum value happens for the uniform distribution.*

Property 7) *Additivity: for two independent distributions, the entropy of the joint distribution, is the sum of the entropies of the two distribution.*

Property 8) *Strong additivity: for two not necessarily independent distributions, the entropy of the joint distribution, is the entropy of the first distribution plus the expected value of the conditional entropy of the second distribution.*

Property 9) *Subadditivity: for two not necessarily independent distributions, the joint entropy is less than or equal to the some of the uncertainties of the two distributions.*

Property 10) *The entropy value will be reduced if two outcomes are combined.*

1-2. Overview of the maximum entropy principle

The maximum-entropy principle (maxent) originated in statistical mechanics by Boltzmann (1871c,b,a) and Gibbs (1902). As an approach to density estimation, it was first proposed by Jaynes (1957b,a), and has since been used in many areas outside statistical mechanics (Kapur and Kesavan, 1992).

1-2-1. Maximum entropy in statistical mechanics

We begin with the work of Boltzmann (1871c,b,a), who studied properties of gas bodies, viewed as systems composed of a large number of molecules. One of his central concerns was how the macro-state of the system is influenced by its micro-states. The macro-state includes properties such as total volume, total number of molecules, and total energy. The micro-state is described by the properties of individual molecules such as their velocities and positions.

To simplify the discussion, assume that the molecules of the gas body occupy discrete states. These can be obtained, for example, by the discretization of positions and velocities of the molecules. A crucial quantity on both the macro-state and the micro-state is the energy. The energy of each molecule is the sum of the kinetic energy, which depends only on the velocity of the molecule, and the potential energy, which depends only on the position of the molecule within a force field. We assume that the division of the state space into discrete cells is fine enough so that the energy of molecules within the same cell is almost constant, but coarse enough to allow a large number of molecules per cell. The micro-state of the system can be viewed as a vector, listing for each molecule the cell it occupies. The macro-state is determined by the histogram of molecule counts across cells. Therefore, to describe the macro-state, it suffices to calculate the most likely histogram.

Boltzmann applied the "principle of indifference" and assumed that all the micro-states are equally likely. Thus, the most likely histogram is the one that can be realized by the largest number of micro-states.

Let's label the discrete cells as $1, 2, \dots, K$ where the number of molecules in the k^{th} cell is N_k and the total number of molecules is N. The total number of ways to realize a concrete allocation into cells is described by equation (1-7)

$$\frac{N!}{N_1!\,N_2!\,\dots\,N_K!} \qquad 1\text{-}7$$

Boltzmann looked for the set of occupancies N_k for which the number of possible realizations equation (1-7) is maximum, while respecting the law of conservation of energy

$$\sum_{k=1}^{K} N_k E_k = E \qquad 1\text{-}8$$

where E_k is the energy associated with the state k and E is the total energy.

Computationally, it is simpler to maximize the logarithm of equation (1-7). The logarithm of equation (1-7) plays a central role in thermodynamics and when multiplied by Boltzmann constant, it defines the thermodynamic entropy:

$$\text{Thermodynamic entropy} \propto \ln \frac{N!}{N_1!N_2!\dots N_K!} \approx \sum_{k=1}^{K} N_k \ln \frac{N}{N_k}$$

Replacing $\frac{N_k}{N}$ by p_k Boltzmann's problem can be rephrased as:

$$\text{Maximize} \sum_{k=1}^{K} N p_k \ln \frac{1}{p_k} \qquad 1\text{-}9$$

$$\text{Subject to the constraint} \sum_{k=1}^{K} p_k E_k = \frac{E}{N} \qquad 1\text{-}10$$

Using the method of Lagrange multipliers, we arrive at the solution to Boltzmann's problem: the Boltzmann distribution, $p_k \propto e^{\lambda E(k)}$, where λ is the corresponding Lagrangian multiplier for equation (1.8) ensuring that the average energy constraint is satisfied. Using the expression for the Boltzmann distribution, it is now possible to study various properties of gas bodies.

1-2-2. Jaynes-Kullback principle of maximum entropy

Jaynes (1957b,a) noticed that Boltzmann's reasoning can be re-interpreted using information theory and generalized to problems outside statistical mechanics. He

suggested that statistical mechanics "may become merely an example of statistical inference."

Jaynes, applied the information-theoretic work of Shannon and claimed that thermodynamic entropy in Boltzmann's problem can be replaced by information-theoretic entropy to quantify how uncertain we are about the system. Our only knowledge about the system is summarized by the average-energy constraint equation (1-10). Among all distributions satisfying this constraint, we should choose the one that is "maximally non-committal with regard to the missing information," i.e., the one with the largest information-theoretic entropy.

$$H(P) = -\sum_{i=1}^{n} p_i \ln p_i$$

Since the information-theoretic entropy is a multiple of the thermodynamic entropy, its maximization yields the result that is identical to Boltzmann's solution.

Moreover, the principle of maximum entropy can be viewed as a generalization of the principle of indifference applied by Boltzmann. In statistical inference, the principle of maximum entopy states that, subject to known descriptive statistics, the probability distribution which best represents the current state of knowledge, is the one with largest entropy. In the other words, it chooses the distribution which simultaneously maximizes a measure of entropy and is compatible with some constraints. If no information is available, the best probability distribution which is least committed to the information not given to us is the uniform distribution. Choosing a probability distribution with less value of entropy means that some data which are not given, are being used. On the other hand, having a descriptive statistic such as sample mean, maximum entropy principle will

construct a probability distribution with the same mean value and the maximum uncertainty.

Mathematically, this principle implies the maximization of Shannon entropy, subject to the following constraints.

$$\sum_{i=1}^{n} p_i = 1 \tag{1-11}$$

$$\sum_{i=1}^{n} p_i g_r(x_i) = a_r, \;\; r = 1,2, \dots, m \tag{1-12}$$

Where $X = (x_1, x_2, \dots, x_n)$, $g_1(X), g_2(X), \dots, g_m(X)$ are functions of X, and $a_1, a_2, \dots, a_m$ are related algebraic moment of each function.

The information-theoretic justification of Jaynes was generalized by Kullback (1959) who assumed that in addition to a set of constraints we are also given a distribution Q, serving as a default guess- the distribution we would choose if we had no data. He suggested choosing the distribution that is the closest to Q among all the distributions satisfying the constraints. The measure of closeness is the relative entropy,

$$KL(P, Q) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i},$$

also known as the Kullback-Leibler divergence, measuring how much information about the outcome could be gained by knowing P instead of approximating it by Q. If Q is uniform then the minimum relative entropy criterion is the same as the maximum entropy criterion.

1-3. Kullback-Leibler directed divergence measure

Alongside Shannon entropy, which quantifies the uncertainty of a probability distribution, Kullback-Leibler directed divergence, introduced by Kullback and Leibler (1951), is another concept which plays an important role in information theory. Kullback-Leibler divergence measures the difference between two probability distributions P and Q. If $P = (p_1, p_2, \dots, p_n)$ and $Q = (q_1, q_2, \dots, q_n)$ be probability distributions, then Kullback-Leibler divergence measure is defined as:

$$KL(P,Q) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} \qquad 1\text{-}13$$

Generally, metric divergence measures such as Euclidean distance, $\sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$, satisfy four conditions:

1) Non-negativity: $D(P,Q) \geq 0$.
2) Identity: $D(P,Q) = 0$ if and only if $P = Q$.
3) Symmetry: $D(P,Q) = D(Q,P)$.
4) Triangular inequality: $D(P,Q) + D(Q,R) \geq D(P,R)$.

Kullback-Leibler directed divergence satisfies the first two conditions of metric measures; but not the third and fourth conditions as they are not essential for a measure of discrepancy. Instead, it possesses two important conditions which are useful for optimization purposes:

5) $KL(P,Q)$ is a convex function of $p_1, p_2, \dots, p_n$.
6) When this measure is minimized subject to some linear constraints the minimizing probabilities are all non-negative.

Some properties of Kullback-Leibler's measure are as follows. The corresponding proofs can be found in Lexa (2004) and Kullback (1959).

Property 1) *$KL(P,Q)$ is a continuous function of $p_1, p_2, \dots, p_n$ and of $q_1, q_2, \dots, q_n$.*

Property 2) *$KL(P,Q)$ is permutationally symmetric, i.e. the value of this measure does not change if the outcomes are labeled differently if the pairs (p_1, q_1), (p_2, q_2), …, (p_n, q_n) are permuted among themselves.*

Property 3) *$KL(P,Q) \geq 0$, and is equal to zero if and only if $P = Q$.*

Property 4) *The minimum value of $KL(P,Q)$ is zero.*

Property 5) *$KL(P,Q)$ is a convex function of both P and Q.*

Property 6) *Since $KL(P,Q)$ is a convex function of P, its maximum for a given Q must occur at one of the degenerate distributions. The maximum value has to be*

$$max(-\ln q_1, -\ln q_2, \dots., -\ln q_n) = \ln \frac{1}{q_{min}}$$

where $q_{min} = min\ (q_1, q_2, \dots, q_n)$.

Similarly, $KL(P,Q)$ is a convex function of Q, its maximum for a given P can be made as large as we wish by making some values of q_i sufficiently small.

Property 7) *When this measure is minimized subject to some linear constraints the minimizing probabilities are all non-negative.*

Property 8) *$KL(P,Q) \neq D(Q,P)$, i.e., $KL(P,Q)$ is not symmetric.*

Property 9) If Q is a priori distribution, P is the probability distribution that minimizes the cross-entropy subject to the constraints (1-11) and

(1-12), and is any other distribution satisfying the same constraints, then

$$KL(R,P) + KL(P,Q) = KL(R,Q)$$

1-3-1. Relationship with Shannon entropy

In addition to the above properties, If Q is a uniform distribution $U = (\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$, then

$$KL(P,U) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{1/n} = \sum_{i=1}^{n} p_i \ln p_i + \ln n = -H(P) + \ln(n)$$

Hence,

$$KL(P,U) = -H(P) + C \qquad 1\text{-}14$$

Where $C = \ln n$ is constant. Thus, from this perspective, the Shannon entropy measure can be considered a special case of Kullback-Leibler directed divergence measure, however they are different conceptually as Shannon entropy is an uncertainty concept; but Kullback-Leibler divergence measures the directed divergence between to probability distributions.

1-3-2. Symmetric divergence

As stated in property 8, Kullback-Leibler's measure, $KL(P,Q)$, is not symmetric. In order to define a symmetric divergence measure $J(P,Q)$ can be defined as follows:

$$J(P,Q) = KL(P,Q) + KL(Q,P) \qquad 1\text{-}15$$

This measure is symmetric since $J(P,Q) = J(Q,P)$, obviously. $J(P,Q)$ is called measure of symmetric cross-entropy or measure of symmetric divergence.

1-4. Generalized measures of entropy and directed divergence

However, we noted that Kullback-Leibler measure satisfies conditions (1), (2), (5) and, (6), there are also other measures that satisfy those four conditions and thus qualify as legitimate measures of directed divergence. Even if a measure satisfies only conditions (1), (2), and (5), but not (6), it can still be considered as a measure of directed divergence. These measures are called generalized measures of directed divergence.

Csiszer (1972), introduced $D(P,Q)=\sum_{i=1}^{n} q_i \phi(\frac{p_i}{q_i})$ as a family of measures for directed divergence. In this family ϕ must be a twice differentiable convex function with $\phi(1)=0$. Measure $D(P,Q)$ defined by Csiszer satisfies conditions (1), (2), (5), but not (6) (See Csiszer,1972).

If $\phi(x)=x\ln x$ then $D(P,Q)=\sum_{i=1}^{n} q_i\left(\frac{p_i}{q_i}\ln\frac{p_i}{q_i}\right)=\sum_{i=1}^{n} p_i \ln\frac{p_i}{q_i}$ which is Kullback-Leibler directed divergence. Thus, Kullback-Leibler directed divergence is a special case of Csiszer's family of measures when $\phi(x)=x\ln x$. Table 1-1shows some special cases of Csiszer's directed divergence measure.

Table 1-1. Some special cases of Csiszer's directed divergence measure

$\phi(x)$	$D(P,Q)$	Introduced by
$x \ln x$	$\sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i}$	Kullback and Leibler (1951)
$\frac{x^{\alpha}-x}{\alpha-1}, \alpha>0, \alpha \neq 1$	$\frac{1}{\alpha-1}\sum_{i=1}^{n} q_i \left[\left(\frac{p_i}{q_i}\right)^{\alpha} - \frac{p_i}{q_i}\right]$	Havrada and Charvat (1967)
$x \ln x - \frac{1+ax}{a}\ln(1+ax) + \frac{1+a}{a}\ln(1+a)$	$\sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} - \sum_{i=1}^{n} \frac{q_i+ap_i}{a} \ln \frac{q_i+ap_i}{q_i} + \frac{1+a}{a}\ln(1+a)$	Kapur and kesavan (1992)
$\frac{x^{\alpha}-x^{\beta}}{\alpha-\beta}, \alpha>1, \beta \leq 1 \; or \; \alpha<1, \beta \geq 1$	$\frac{1}{\alpha-\beta}\left(\sum_{i=1}^{n} {p_i}^{\alpha} {q_i}^{1-\alpha} - \sum_{i=1}^{n} {p_i}^{\beta} {q_i}^{1-\beta}\right)$	Sharma and Mittal (1975)

Csiszer family of divergence measures are not defined when $q_i = 0$ and the corresponding $p_i \neq 0$. To overcome this problem P' and Q' can be defined such that $P' = \frac{aP+b}{a+nb}$ and $Q' = \frac{aQ+b}{a+nb}$ where $a > 0$ and $b > 0$. P'andQ' are also probability distributions and $D(P', Q') = \sum_{i=1}^{n} \frac{aq_i+b}{a+b}\ \phi\left(\frac{ap_i+b}{aq_i+b}\right)$ can be used as a measure of directed divergence of P and Q. Thus, the Csiszer family of measures can be generalized by equations (1-16) and (1-17).

$$D(P,Q) = \sum_{i=1}^{n} (q_i + c)\, \phi\left(\frac{p_i + c}{q_i + c}\right), \qquad c > 0 \tag{1-16}$$

$$D(P,Q) = \sum_{i=1}^{n} (aq_i + 1)\, \phi\left(\frac{ap_i + 1}{aq_i + 1}\right), \qquad a > 0 \tag{1-17}$$

The measures presented in Table 1-1 can be generalized by replacing p_i by $p_i + c$ and q_i by $q_i + c$ or p_i by $1 + ap_i$ and q_i by $1 + aq_i$. For example, generalized forms of Kullback-Leibler divergence measure would be:

$$D(P,Q) = \sum_{i=1}^{n} (p_i + c) \ln\left(\frac{p_i + c}{q_i + c}\right), \qquad c > 0 \tag{1-18}$$

$$D(P,Q) = \sum_{i=1}^{n} (ap_i + 1) \ln\left(\frac{ap_i + 1}{aq_i + 1}\right), \qquad a > 0 \tag{1-19}$$

In addition of being defined even when $q_i = 0$ and the corresponding $p_i \neq 0$, these generalized forms give greater flexibility in applications because of the parameters that they have.

A generalized measure of entropy of a distribution P can be defined as a monotonic decreasing function of the generalized directed divergence of P from the uniform distribution. Hence, corresponding to every generalized measure of directed divergence,

there is a unique measure of generalized entropy and according to Kapur and Kesavan (1992), if $U = (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$ is a uniform distribution, then generalized measures of directed divergence and entropy are related by equation (1-20).

$$H(P) = \max D(P, U) - D(P, U) \qquad 1\text{-}20$$

Table 1-2 shows some generalized measures of directed divergence and their corresponding measure of entropy.

Table 1-2. Some generalized measures of directed divergence and their corresponding measure of entropy

$D(P,Q)$	$H(P)$
$\sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i}$	$-\sum_{i=1}^{n} p_i \ln p_i$
$\frac{1}{\alpha-1}\sum_{i=1}^{n} q_i \left[\left(\frac{p_i}{q_i}\right)^{\alpha} - \frac{p_i}{q_i}\right]$	$\frac{1}{\alpha-1}\left(\sum_{i=1}^{n} {p_i}^{\alpha} - 1\right)$
$\sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} - \sum_{i=1}^{n} \frac{q_i + ap_i}{a} \ln \frac{q_i + ap_i}{q_i} + \frac{1+a}{a} \ln(1+a)$	$-\sum_{i=1}^{n} p_i \ln p_i - \frac{1+an}{an} \ln(1+an) + \sum_{i=1}^{n} \frac{1+anp_i}{an} \ln(1+np_i)$
$\frac{1}{\alpha-\beta}\left(\sum_{i=1}^{n} {p_i}^{\alpha} {q_i}^{1-\alpha} - \sum_{i=1}^{n} {p_i}^{\beta} {q_i}^{1-\beta}\right)$	$\frac{n^{\alpha-1}}{\beta-\alpha}\left(\sum_{i=1}^{n} {p_i}^{\alpha} - 1\right) - \frac{n^{\beta-1}}{\beta-\alpha}\left(\sum_{i=1}^{n} {p_i}^{\beta} - 1\right)$

1-5. Outline and contribution

Although entropy concept has been applied in wide variety of fields (see Kapur (1993)), there exist even more areas where it can be employed. In this work we study applications of entropy in industrial engineering. Three problems are addressed and for each problem an entropy-based approach is suggested.

In the first problem we address the dispatching issue of a material handling system within the context of Automatic Guided Vehicles (AGV) in a discrete part manufacturing system. The dispatching issue is about allocating available AGVs to move requests to ensure efficient part flow in the factory. We believe that the objective of this resource allocation problem should be load balancing among the factory work centers. Using Kullback-Leibler directed divergence, we present entropy-based resource allocation algorithms that consider the consequence of potential moves on the load balance of the factory before resources are allocated. The proposed algorithms are suitable for real-time implementation and strive to even the load in the factory while satisfying the move requests generated by the factory work centers.

In the second problem we focus on ranking and selection based on the mean performance measure. We use maximum entropy and Kullback-Leibler directed divergence principles to present a two stage algorithm for this problem. The proposed method contributes to the ranking and selection body of knowledge because it relaxes the normality assumption for underlying population which restricts the frequentist algorithms, it does not assume any priori distribution assumed by bayesian methods, and finally it provides ranking of systems based on their observed performance measures.

Finally, we present an entropy-based criterion for comparing two alternatives. Our comparison is based on directed divergence between alternatives' cumulative probability distributions. We compare the new criterion with stochastic dominance criteria such as first order stochastic dominance (FSD) and second order stochastic dominance (SSD). Since stochastic dominance rules may be unable to detect dominance even in situations when most decision makers would prefer one alternative over another, our criterion reduces the probability of obtaining the nondominance set in such situations. Among two alternatives, we show that if one alternative dominates the other one by SSD, the dominating alternative will be dominated by our new criterion. In addition, we show that the probability associated with our new criterion is consistent with the probability corresponding to p almost stochastic dominance (p-AFSD).

The remainder of this dissertation is organized as follows. Chapter 2 reviews the applications of entropy in industrial engineering and specially manufacturing context. Entropy-based dispatching algorithms for automatic guided vehicles are presented in chapter 3. Ranking and selection via simulation is studied in chapter 4 and the new entropy-based ranking and selection procedure is offered. Chapter 5, focuses on the stochastic dominance issue. An entropy-based measure for stochastic dominance is introduced in this chapter.

Chapter 2

Background and literature review

Soon after entropy and maximum entropy concepts were introduced by Shannon and Jaynes respectively, they began to be used in wide variety of fields such as statistical thermodynamics, urban and regional planning, business, economics, finance, operations research, queueing theory, spectral analysis, image reconstruction, biology and manufacturing (see Kapur (1993) for some applications).

Within Manufacturing, the entropy concept has been applied to measure diversity of production systems from two different points of views: Flexibility and Complexity.

a) Flexibility

Flexibility is defined as the ability of a system to cope with changes (Mandelbaum, 1990). Sethi and Sethi (1990) defined different types of flexibility as shown in Table 2-1.

Table 2-1. definitions of different types of flexibility

Flexibility	Definition
Machine	Various types of jobs can be performed by machine
Operation	Different ways can be used to perform a job
Process	Different parts can be made without a major setup
Routing	A part can be produced through different routes
Volume	Different overall output levels can be produced
Market	The system can easily adapt itself with market changes
Production	Different parts can be produced without adding major capital equipment
Program	System can be run unattended for a period of time
Material handling	Different parts can be moved efficiently through a manufacturing facility
Product	New products can be added or substituted easily
Expansion	Capability and capacity can be increased easily

Entropy has been widely used to measure and quantify flexibility in manufacturing systems. A system facing uncertainty uses flexibility as an adaptive response to cope with changes. The flexibility in the action of the system depends on the decision options or the choices available and on the freedom with which various choices can be made (Kumar, 1987). A greater number of choices leads to more uncertainty of outcomes, and hence, increased flexibility. According to Pereira and Paulre (2001), the level of flexibility of a system depends on the set of possible outcomes and therefore possible ways that these outcomes can be obtained. This inference has been the main driver to apply entropy as a measure of flexibility by different researchers. Thus for all the proposed measures, the possible states of the system and their related probabilities of occurrence is defined. Entropy concepts are then applied to the obtained probability distribution to measure flexibility.

Yao (1985), and Yao and Pei (1990) applied entropy to the measurement of the routing flexibility of a flexible manufacturing system. Kumar (1987, 1988) adopted it for the measurement of operation flexibility. Chang et al. (2001) suggested that two attributes of flexibility, namely routing efficiency and routing versatility, should be considered in the measurement models of routing and single-machine flexibility. Chang (2007) proposed a multi attribute approach for routing flexibility by considering three attributes: routing efficiency, routing versatility and routing variety. Chang (2009) offered a multi attribute approach for machine-group flexibility. Rao and Gu (1994) used entropy as a means of measuring production volume and production flexibility. Shuiabi et al. (2005) applied entropy as a measure of the flexibility of production operations. Extended from Yao and Pei's approach (Yao and Pei, 1990), Piplani and Wetjens (2007) proposed two

dispatching rules, namely 'least reduction in entropy' and 'least relative reduction in entropy' for operations dispatching based on entropic measures of part routing flexibility. Table 2-2 shows the proposed entropy measures of flexibility and their related flexibility type.

Table 2-2. Applications of entropy in different types of flexibility

Flexibility	Entropy-based approach
Machine	Chang et al. (2001), Chang (2009)
Operation	Shuiabi et al. (2005)
Process	----
Routing	Chang (2007), Piplani and Wetjens (2007), Yao (1985), Yao and Pei (1990)
Volume	Chang (2004), Rao and Gu (1994), Olivella et al. (2010)
Market	----
Production	Rao and Gu (1994)
Program	----
Material handling	----
Product	----
Expansion	----

a) Complexity

Complexity can be associated with systems that are difficult to understand, describe, predict or control. As noted by Scuricini (1987) who states that: "*Complexity is a subjective quality, its meaning and its value change following the scope of the system being taken under consideration*", it is difficult to define complexity in a precise formal sense.

Generally, the complexity of a system can be described in terms of several interconnected aspects of the system such as: number of elements or sub-systems, degree

of order within the structure of elements or sub-systems, degree of interaction or connectivity between the elements, sub-systems and the environment, level of variety, in terms of the different types of elements, sub-systems and interactions, and degree of predictability and uncertainty within the system.

Authors have analyzed complexity in two different aspects: structural complexity which is associated with the system configuration (Deshmukh, 1993; Deshmukh et al., 1998) and operational complexity which is defined as the uncertainty associated with the dynamical aspects while system is running (Frizelle and woodcock, 1995; Scuricini, 1987; Sivadasan el al., 2002).

Entropy provides a means of quantifying complexity. The complexity of a system increases with increasing levels of disorder and uncertainty. Therefore, a higher complexity system requires a larger amount of information to describe its state. According to the entropy concept, the structural complexity is thus defined as the expected amount of information (entropy) necessary to describe the state of a planned system; while operational complexity is defined as the expected amount of information necessary to describe the state of the system's deviation from the schedule.

The idea of using entropy as a measure for complexity in manufacturing was first introduced by Frizelle and Woodcock (1995), and Frizelle (1996), for operational complexity. Calinescu et al. (1998) applied and assessed the operational complexity measures offered by Frizelle and Woodcock (1995), and Frizelle (1996). Deshmukh (1993), and Deshmukh et al. (1998) used entropy to suggest an entropy-based measure for structural complexity. Frizelle and Suhov (2001) offered some measures for both structural and operational complexities. Sivadasan et al. (2002) offered an entropy-based

methodology to measure the operational complexity of supplier-customer systems associated with the uncertainty of material and information. Fujimoto et al. (2003) applied entropy to study structural complexity because of product variety. They used this measure to manage assembly process design strategies. Yu and Efstathiou (2006) introduced entropy-based operational complexity as a new way to assess the performance for rework cells. Martinez-Olvera (2008) proposed an entropy formulation to assess information sharing approaches in supply chain environments. Sivadasan et al. (2006) modeled the operational complexity of supplier-customer systems from an information-theoretic perspective. Wu et al. (2007) studied the relationship between cost and the operational complexity measures offered by Frizelle and Woodcock (1995). Sivasadan et al. (2010) studied the effect of closer supply chain integration on operational complexity of production scheduling.

In addition to manufacturing, the entropy concept has been applied in different fields within decision making under uncertainty context including portfolio selection and measures of risk.

- Portfolio selection

Portfolio selection is about assigning a certain amount of wealth to different assets so that the investment can bring a most profitable return. Markowitz (1952) proposed the mean-variance (M-V) analysis model and created a fundamental basis for modern portfolio analysis. M-V model tries to maximize the expected return for a given level of risk or to minimize the level of risk for a given level of expected return. The portfolio variance decreases as portfolio diversification increases. Entropy has become a well-established measure of diversification. Higher portfolio diversification yields a greater

entropy value. Bera and Park (2005, 2008) applied entropy and cross entropy to provide a well-diversified portfolio. Huang (2008) introduced two types of fuzzy mean-entropy models. Zhang et al. (2009) considered a multi-period portfolio selection problem by taking into account four criteria: return, risk, transaction cost, and diversification degree of portfolio. They offered a possiblistic mean-semivariance entropy model in which entropy is applied to measure diversification degree of portfolio. Jana et al. (2009) considered the portfolio selection problem by taking into account four criteria and added entropy as the objective function to generate a well diversified asset. Qin et al. proposed three portfolio selection methods based on fuzzy cross-entropy. Wu et al. (2009) applied the maximum entropy principle to obtain a numerical solution for their min-max model to investigate the optimal portfolio with riskless assets. Rodder et al. (2010) used an information theoretical inference mechanism under maximum entropy and minimum cross entropy principles, respectively, in order to propose an entropy-driven expert system for portfolio selection. Bhattacharyya (2013) offered a fuzzy portfolio selection model by minimizing mean-skewness as well as minimizing variance- cross entropy.

- Measures of risk

According to Jones and Zitikis (2007), a risk measure is a mapping from the set of all random variable to the extended real numbers in order to quantify the degree of risk involved in each random variable. Although many authors have proposed suitable methods to measure risk in the past two decades, entropy applications to measure risk have been constrained to a limited number of research studies.

Yang and Qiu (2005) introduce the expected utility entropy (EU-E) measure of risk and suggest a decision-making model based on expected utility and entropy. The EU-E

reflects an individual's intuitive attitude toward risk while the decision model incorporates the expected utility decision criterion as a special case. Föllmer (2011) proposes a new coherent risk measure called the iso-entropic risk measure, which is based on cross entropy under the theory framework of Artzner et al.(1999). Ahmadi-Javid (2012) introduces the concept of entropic value-at-risk (EVaR), a new coherent risk measure that corresponds to the tightest possible upper bound obtained from the Chernoff inequality for the value-at-risk (VaR) as well as the conditional value-at-risk (CVaR). Chengli and Yan (2012) study a coherent version of the entropic risk measure, both in the law invariant case and in a situation of model ambiguity.

Chapter 3

Entropy-based dispatching for automatic guided vehicles

Material handling is a nonvalue adding, necessary function for the production of discrete parts. It entails moving work in progress among work centers and raw material into and finished products out of the factory. Inefficient implementation of material handling could substantially add to production costs. For example, delays in work-in-process movement would increase parts' factory flow time, resulting in higher inventory costs. Therefore, efforts have been made to improve both the technology and efficient implementation of material handling systems.

An Automatic Guided Vehicle System (AGVS) is an example of a material handling system that has benefited from technological innovation and has created opportunities to improve the efficiency of material movement. An AGVS is considered by many as the most flexible automated material handling system (Hwang and Kim, 1998). This flexibility stems from the intelligence imbedded in the AGVS and on board of each automated guided vehicle (AGV). This intelligence allows the system to be responsive, in real time, to the material move requests generated by the work centers.

Operation control of an AGVS consists of resolving vehicle routing, vehicle dispatching, and vehicle scheduling issues. A routing issue (Kim and Tanchoco, 1991; Nishi et al., 2005; Nishi et al., 2009) is identifying the best path for the assigned AGV. A dispatching issue (Egbelu and Tanchoco, 1984; Kim and Klein, 1996; Hwang and Kim, 1998; Ho and Chein, 2006) is assigning AGVs to pickup or delivery requests. A scheduling issue (Zaremba et al., 1997; Sabuncuoglu, 1998; Veeravalli et al., 2002; Gaur

et al., 2003) is determining the arrival or departure times of AGVs at pickup or delivery points. Each of these control issues has its own relevant objective(s) that myopically could be optimized. However, an AGVS controller must contribute toward optimizing the objectives of the factory, since material handling is a supporting function within a factory. An important objective for factory management is minimization of the time that parts spend in the factory, i.e., parts flow time. The material handling system can contribute to achieving the factory objective by addressing work center move requests in a timely manner while avoiding creation of temporary bottlenecks or aggravating structural bottlenecks. Temporary bottlenecks are created when workload is not properly distributed among the work centers, and structural bottlenecks are aggravated when a scarce capacity of work centers is ignored when dispatching decisions are made.

In this chapter, we focus on the dispatching issue of a material handling system within the context of an AGVS in a discrete part manufacturing system. Since the dispatching issue is about allocating available AGVs to move requests to ensure efficient part flow in the factory, the objective of this resource allocation solution is load balancing among the factory work centers. Specifically, we use an entropy-based resource allocation rule that considers the consequence of potential moves on the load balance of the factory before resources are allocated. The proposed approach is suitable for real-time implementation and strives to even the load in the factory while satisfying the move requests generated by the factory work centers.

The remainder of this chapter is organized as follows. In Section 3.1, we present a brief review of prior work. The application Kullback–Leibler divergence measure is explained in Section 3.2. Three dispatching algorithms are proposed in Sections 3.3. In

Section 3.4, performance of proposed algorithms is studied and compared with other algorithms by conducting simulation experiments. Finally, conclusions and future research directions are presented in Section 3.5.

3-1. Prior work

Dispatching is a resource allocation problem in that idle AGVs (resource) are assigned to move requests (demand). At a given time, the resource vector consists of a possibly ordered list of idle AGVs, and the demand vector consists of a possibly ordered list of move requests. The problem is to determine the one-to-one pairing of the elements of these two vectors.

AGV dispatching rules have been investigated by many researchers. Simple heuristic dispatching rules are discussed by Egbelu and Tanchoco (1984). They divided the dispatching rules into two categories: work-center-initiated (mapping from the demand vector to the resource vector) and vehicle-initiated (mapping from the resource vector to the demand vector) rules and showed that in busy production settings vehicle-initiated dispatching rules are preferable. Their vehicle-initiated category consists of the following rules: random work center (RW); shortest travel time/distance (STT/D); longest travel time/distance (LTT/D); maximum outgoing queue size (MOQS); minimum remaining outgoing queue size (MROQS); modified first-come, first-served (MFCFS); and unit load shop arrival time (ULSAT). Each of these heuristic rules optimizes an objective. These objectives are summarized in Table 3-1.

Table 3-1. Objectives of heuristic dispatching rules

Heuristic Rule	Objective
RW	Maximizing long term dispatching entropy
STT/D	Minimizing the percentage of vehicles' time/distance empty travel
LTT/D	Maximizing the percentage of vehicles' time/distance empty travel
MOQS	Minimizing the percentage of time parts spend in output queue
MFCFS	Minimizing the elapsed time between placing a move request by work center and its satisfaction
MROQS	Minimizing the possibility of work center blockage
ULSAT	Minimizing the percentage of time parts spend in input queues

Bartholdi and Platzman (1989) propose a decentralized dispatching rule for a simple closed-loop system. Their suggested rule, First Encountered, First Served (FEFS), assigns the AGV traveling along the loop to the move request it encounters first. The basis of this rule is to minimize the percentage of AGV empty travel time. Yamashita (2001) suggests two dispatching policies: "the nearest vehicle in time" and "the nearest vehicle in distance." The objective of these rules is to minimize the AGV empty travel time. Kim et al. (2004) propose two dispatching rules for a single-loop, single vehicle AGV system: MAED (Minimum Average of Empty Distance) and MSED (Minimum Sum of Empty Distance). The MAED rule minimizes the average vehicle travel time, and the MSED rule minimizes the sum of AGV vehicle travel time.

Some authors have extended single-objective dispatching rules to multiobjective decision rules by considering several decision criteria recognizing, due to interdependencies within a manufacturing process, some objectives may have conflicts. In general, multiobjective dispatching rules are supposed to perform better than single-objective rules. Kim and Hwang (1999) propose an AGV dispatching rule based on three bidding functions defined for travel time, input buffer size, and output buffer size. The objective is to minimize a combination of these three bidding functions. Naso and

Turciano (2005) suggest a hierarchical fuzzy dispatching rule in which, if a critical move request (e.g., requests from work centers with a saturated output buffer or directed to starved destinations) is detected, it is selected as the final decision. Otherwise, a multicriteria rule is called with the objective of optimizing the AGV utilization by minimizing the empty vehicle travel time. Tan and Tang (2001) suggest an AGV dispatching rule to strike a compromise between satisfaction of several simple objectives using a hybrid Fuzzy-Taguchi approach. Bozer and Yen (1996) develop two algorithms, Modified Shortest Travel Time First (MOD STTF) and Bidding-Based Dynamic Dispatching (B2D2). The basis of these policies is to minimize the empty vehicle travel time. Kim and Klein (1996) propose several multiattribute decision rules and compare them with the single-attribute dispatching rules for different performance measures. Jeong and Randhawa (2001) propose a multiattribute dispatching rule by considering the unloaded travel distance to the pickup point, the remaining space in the input buffer of the delivery point and the remaining space in the outgoing buffer of the pickup point. Guan and Dai (2009) offer a deadlock-free multia-attribute dispatching method by considering three criteria: traveling distance, input queue size and output queue size. These criteria are weighted and combined into a single criterion. Criteria are weights influenced by transportation loads and processing loads.

The objectives of most of the proposed dispatching rules in the literature are optimization of material handling rather than factory performance measures. Since a material handling system has a supporting role in the factory, its control should be aligned with optimizing the factory's performance. A highly desirable performance objective of a factory is to achieve a laminar flow of parts within the factory. The notion

of laminar flow has been of much interest in operation of assembly lines and has led to the creation of a large body of knowledge generally known as assembly line balancing (e.g., Stecke, 1983; Mukhopadhyay et al., 1992; Becker and Scholl, 2006). The promise is that when factory work centers (or assembly line work centers) are balanced in work content, no work center is overburdened or starved; and thus the flow of parts through the factory is smooth.

We propose a look-ahead AGV dispatching approach that considers the contribution of a potential material move to the laminar flow within the factory before an AGV is dispatched. The only prior work which considers the factory load balancing concept when dispatching decisions are made is Kim et al. (1999). They define a balancing index based on the difference between the number of parts in the pickup and the destination work centers. Their rule selects the job with the highest balancing index. Their decision is based on the current state of the system, while they do not consider the resulting state after the dispatched AGV delivers the parts to the destination. Since movement of parts into or out of the system or among work centers impacts the state of the system, we take into consideration the impact of making such moves as we plan an AGV's dispatch. Hence, in this research, the resulting state of the system after each possible dispatching decision is predicted, and the dispatch decision that contributes the most to the laminar flow of the factory is selected. The Kullback–Leibler divergence principle is used to measure the contribution of a dispatch toward the system laminar flow, and simulation experiments are conducted to compare the performance of the proposed approach with three simple dispatching rules: Shortest Travel Distance (STT), Maximum Outgoing Queue Size (MOQS), and Modified First-Come, First-Served (MFCFS).

3-2. Applying the Kullback–Leibler divergence measure

We represent the work content of work centers at a given time by a vector whose elements are the existing work content in each work center. Moving a work load between two work centers changes the work content vector. Under the ideal balanced factory operating conditions, elements of the work content vector should be equal at all times. Manufacturing system literature suggests that a balanced production system has the highest throughput, that is, bottlenecks are avoided (laminar flow). This notion of balanced factory gives us an ideal benchmark, although a practical utopia, for operating a factory.

We utilize this notion of balanced factory as our reference vector and order the move requests based on the distance of their resultant workload vector from the balanced factory work content vector. The move request that results in a workload vector closest to the balanced factory vector is deemed to be the best move at the time of the dispatch decision.

If elements of the work content vector are represented as a proportion of the total work content of the factory, then the work content vector can be viewed as a probability distribution. Under this scenario, the work content vector of an ideal factory is a uniform distribution. Therefore, determining the distance between the resultant work content vector of a move request and the ideal work content vector becomes determining the degree of divergence between two probability distributions.

We will apply the Kullback-Leibler directed divergence measure explained in chapter 1 to tackle this problem. As mentioned in chapter 1, their measure, measures the directed

divergence between a probability distribution,P, and a reference distribution, Q. If $P = (p_1, \dots, p_n)$ and the reference distribution is $Q = (q_1, \dots, q_n)$, then $KL(P, Q) = \sum_{i=1}^{n} p_i \ln\left(\frac{p_i}{q_i}\right)$. Moreover, $KL(P, Q) = 0$ if, and only if, $P = Q$. When the reference probability distribution is uniform, i.e., $q_i = q \ \forall \ i$, then $KL(P, Q) = -H(P) + \ln(n)$. The first term of $KL(P, Q)$ is the Shannon entropy, and the second term is a constant. Thus, when operation utopia is a balanced factory, it suffices to calculate the Shannon entropy of the potential resultant work content distributions and select the dispatch that has the largest Shannon entropy.

3-3. Proposed dispatch algorithms

A dispatch action reduces the number of parts in the output queue of the originating work center and increases the number of parts in the input queue of the receiving work center by the same amount. Increasing the size of a work center's input queue increases the work content of the work center, while reducing the size of the output queue does not impact the work content of the work center. Additionally, process completion of a part at a work center which results in moving the part to the output queue changes the work content of the work center. In general, the addition of parts to an input queue of a work center has a greater impact on the work content of that work center than moving one or more parts to the output queue of the work center. The primary reason is that batches of parts are moved into the input queue, and single parts are moved into the output queue.

In this section, we present three algorithms for a dispatching decision. The first algorithm attempts to balance the factory work content among all work centers by assessing the impact of increasing the work content of a work center after a dispatch.

The second algorithm attempts to balance the number of loads in the output queue of work centers by assessing the impact of removing a load from the output queue of a work center. The third algorithm attempts to balance the combined number of loads in the input and output queues of the work centers by assessing the impact of a potential dispatch on the factory balance. Algorithms two and three may be considered work content balancing algorithms, rather than number of load balancing algorithms, if the work content of all loads in the factory is the same. For all three algorithms, the base reference is an ideal factory where work content or number of loads are uniformly distributed among the work centers. Before we present the algorithms, we define the notations used in the algorithms and state the assumptions under which the algorithms were developed.

3-3-1. Notation

N_p: Number of part types

N_s: Number of work centers

N_a: Number of vehicles

V: Vehicle speed

C: Vehicle capacity

LT: Vehicle loading time

UT: Vehicle unloading time

$S(l,i)$: Mean service time of part type l at work center i

PD: An $N_s \times N_s$ matrix where $PD_{d,i}$ is the distance between the drop-off point of work center d to pickup point of work center i

DP: An $N_s \times N_s$ matrix where $DP_{d,i}$ is the distance between the pickup point of work center d to drop-off point of work center i

λ_l: Arrival rate of part l into the system

$\widehat{M}_{liqt}$: Number of undispatched and unassigned parts of type l in the output queue of work center i that is requested to be moved to the input queue of work center q at time t

$\widehat{M}_{it}$: Number of undispatched and unassigned parts at the output queue of work center i at time t, $\widehat{M}_{it} = \sum_{l=1}^{N_p} \widehat{M}_{liqt}$

M_t: Set of move requests at time t, $M_t = \{\widehat{M}_{liqt} | \widehat{M}_{liqt} > 0\}$

$\tilde{I}_{lit}$: Estimated number of parts of type l at the input queue of work center i at time t

$\tilde{O}_{lit}$: Estimated number of parts of type l at the output queue of work center i at time t

$\tilde{O}_{it}$: Estimated number of parts at the output queue of work center i at time t, $\tilde{O}_{it} = \sum_{l=1}^{N_p} \tilde{O}_{lit}$

$W_{it}^h(j, t_a)$: The h^{th} waiting position of part type j that arrived at time t_a at the input queue of work center i at observed time t

$\tau_{lit}^h(t_p, n)$: The arrival time, t_p, at the output queue of work center i of the h^{th} vehicle dispatched at time t to pick up n units of part type l

$\gamma_{lit}^{h}(t_d, n)$: The arrival time, t_d, at the input queue of work center i of the h^{th} vehicle dispatched at time t to deliver n units of part type l

E_{liq}: Estimated system entropy if part type l is moved from output queue of work center i to the input queue of work center q

3-3-2. Assumptions

The following assumptions are used:

1) All AGVs are single load vehicles.
2) The factory layout and AGVS guide paths are known.
3) The number of AGVs available in the system is known.
4) Number and location of pickup and delivery points are known and fixed.
5) An idle AGV stays at the work center of last delivery before being dispatched.
6) AGV failure time is negligible.
7) The longest idle AGV will be assigned to a dispatch first.

Assumption 7 expresses the work-center-initiated rule which will be applied if more than one AGV is idle at a given time.

3-3-3. Algorithm one: work content balancing algorithm (WCBA)

Suppose a vehicle is idle at location d at time t_0, and $\left|M_{t_0}\right| \neq 0$ move requests have been made. Let T_{liq} be the total time required to complete move request M_{liqt_0} and $t_0 + T_{liq}$ be the clock time when it is completed. For each move request $M_{liqt_0} \in M_{t_0}$, the expected work content of work centers at time $t_0 + T_{liq}$ is computed. The algorithm

selects the move request that results in the work content of work centers to be as equal as possible (closest to an ideal factory).

The pseudo code of *WCBA* consists of the *Main* procedure and the *Work Content Estimation* procedure. The *Work Content Estimation* procedure calculates the work content of work centers at the completion time of each requested move.

WCBA*: *Main Procedure

Step 1. Identify the set of move requests at time t_0, M_{t_0}.

Step 2. For each move request $M_{liqt_0} \in M_{t_0}$:

- Compute $T_{liq} = \frac{DP_{d,i}}{V} + LT + \frac{PD_{i,q}}{V} + UT$,
- Call *Work-Content-Estimation procedure* to compute $\tilde{I}_{jk,t_0+T_{liq}}$ for all $j \in \{1, \dots, N_p\}$ and $k \in \{1, \dots, N_s\}$,
- Compute $E_{liq} = -\sum_{k=1}^{N_s} \left(\frac{\sum_{j=1}^{N_p} S(j,k).\tilde{I}_{jk,t_0+T_{liq}}}{\sum_{w=1}^{N_s}\sum_{j=1}^{N_p} S(j,w).\tilde{I}_{jw,t_0+T_{liq}}} \ln \frac{\sum_{j=1}^{N_p} S(j,k).\tilde{I}_{jk,t_0+T_{liq}}}{\sum_{w=1}^{N_s}\sum_{j=1}^{N_p} S(j,w).\tilde{I}_{jw,t_0+T_{liq}}} \right)$.

Step 3. Set $(l^*, i^*, q^*) = Max\{E_{liq}\}$.

Step 4. Dispatch the vehicle to the work center i^* to pickup part type l^*.

The *Main* procedure consists of four steps. Step 1 identifies M_{t_0}, the set of move requests at time t_0. Step 2 first calculates the total time required to complete each move request. Then the *Work Content Estimation procedure* is called to compute the expected number of part types at the input queues of work centers at the move request completion time. Finally, the work content of the work center is calculated and system entropy is obtained. In Steps 3 and 4, respectively, the move request with the largest entropy measure is chosen as the best dispatch decision and a vehicle is dispatched.

WCBA*: *Work-Content-Estimation Procedure

Step 1. Set $U = Min\left(C, \hat{O}_{lit_0}\right)$.

Step 2. $\Upsilon_j = \left\{\gamma^h_{jkt_0}(t_d, n) \middle| t_d < t_0 + T_{liq}\right\}, j \in \{1, \ldots, N_p\}$,

$\tilde{I}_{j,k,t_0+T_{liq}} = I_{jkt_0} + \sum_{\gamma^h_{jkt_0}(t_d,n)\in\Upsilon_j} n$,

$W_{kt_0} = W_{kt_0} \bigcup_{j\in\{1,\ldots,N_p\}} \bigcup_{\gamma^h_{jkt_0}(t_p,n)\in\Upsilon_j} \left\{\overbrace{(j,t_d),\ldots,(j,t_d)}^{n}\right\}$

Step 3. If $k = 1$, then for each part type j with mean interarrival time $\frac{1}{\lambda_j}$ generate n_j interarrival times $\tau_{1_j}, \ldots, \tau_{n_j}$ such that $\sum_{m_j=1}^{n_j} \tau_{m_j} < T_{liq}$.

Set $W_{kt_0} = W_{kt_0} \bigcup_{\substack{\forall m_j\in\{1,\ldots,n_j\} \\ \forall j\in\{1,\ldots,N_p\}}} \left(j, \sum_{r=1}^{m_j} \tau_{j_r}\right)$ and sort it by the ascending arrival time.

Set $\tilde{I}_{jk,t_0+T_{liq}} = \tilde{I}_{jk,t_0+T_{liq}} + n_j$.

Step 4. Set $t = t_0$.

Step 5. If $\sum_{j=1}^{N_p} \tilde{I}_{jk,t_0+T_{liq}} > 0$ set $t = t_a | {W_{kt_0}}^1(j, t_a)$ and go to Step 6, or else go to Step 7.

Step 6. If $t < t_0 + T_{liq}$, then

$\tilde{I}_{j|{W_{kt_0}}^1(j,t_a),k,t_0+T_{liq}} = \tilde{I}_{j|{W_{kt_0}}^1(j,t_a),k,t_0+T_{liq}} - 1$,

$t = t + S(k, t_a | {W_{kt_0}}^1(j, t_a))$,

$W_{kt_0} = W_{kt_0} / \left\{{W_{kt_0}}^1(j, t_a)\right\}$, and go to Step 5, or else go to Step 7.

Step 7. If $k = q$ then, $\tilde{I}_{lk,t_0+T_{liq}} = \tilde{I}_{lk,t_0+T_{liq}} + U$.

Work Content Estimation procedure is a subprocedure which is called by the *Main* procedure to compute $\tilde{I}_{jk,t_0+T_{liq}}$. Step 1 compares the number of parts of type l at the output queue of work center i with C, the vehicle capacity, to find, the number of parts U that can be moved. Step 2 checks whether there are any dispatched vehicles to deliver parts to the input queue of work center k before move completion time $t_0 + T_{liq}$. If a

move could be accomplished, then the number of parts and ordered set of parts waiting to be processed at the input queue of this work center are updated. If work center k is the work center at which parts enter the system, Step 3 computes the expected number and the order of parts which enter the input queue of this work center during the move completion time. The number of parts and ordered set of parts waiting to be processed at the input queue of work center k are also updated in this Step. Steps 4, 5, and 6 calculate the number of parts that will be processed from time t_0 to time $t_0 + T_{liq}$. Finally, if work center k is the work center to which the parts will be transported, Step 7 updates the number of parts at the input queue of the work center by adding U units of parts to it.

3-3-4. Algorithm two: output queue balancing algorithm (OQBA)

OQBA looks at the number of parts at the output queues of work centers and selects a move request in which the output queue sizes become closest to each other. The *Main* procedure of this algorithm is similar to that of WCBA; but instead of balancing the work content, it tries to balance the output queue sizes. Hence, the only difference in the *Main* procedure takes place at the second step, as shown below:

Step 2 of *OQBA: Main Procedure*

Step 2. For each move request $M_{liqt_0} \in M_{t_0}$:

- Compute $T_{(l,i,q)} = \frac{DP_{d,i}}{V} + LT + \frac{PD_{i,q}}{V} + UT$,
- Call *Number In Queue Estimation procedure* to compute $\tilde{O}_{k,t_0+T_{liq}}$ for all $k \in \{1, \dots, N_s\}$,
- Compute $E_{liq} = -\sum_{k=1}^{N_s} \left(\frac{\tilde{O}_{k,t_0+T_{liq}}}{\sum_{j=1}^{N_s} \tilde{O}_{j,t_0+T_{liq}}} \ln \frac{\tilde{O}_{k,t_0+T_{liq}}}{\sum_{j=1}^{N_s} \tilde{O}_{j,t_0+T_{liq}}} \right)$.

Number in Queue Estimation procedure is a subprocedure inside the Main procedure. Step 1 verifies that the number of parts of type l that are to be moved from the output queue of work center i does not exceed the vehicle capacity. Step 2 checks if there are any dispatched vehicles to pick up parts from the output queue of work center k before move completion time $t_0 + T_{liq}$. If such vehicles exist, the scheduled number of parts to pick up is deducted from the output queue of work center k at the scheduled pickup time. Step 3 checks if there are any dispatched vehicles to deliver parts to the input queue of work center k before move completion time, $t_0 + T_{liq}$. If there are any, then the number of parts and ordered set of parts waiting to be processed at the input queue of this work center are updated. Step 4 computes the expected number and the order of parts which enter the input queue of work center k during the move completion, if k is a work center at which parts enter the system. The number of parts and the ordered set of parts waiting to be processed at the input queue of work center k are then updated in this step. Steps 5, 6, and 7 calculate the number of parts that will be processed from time t_0 to time $t_0 + T_{liq}$. If work center k is the work center to which parts are moved, Step 8 updates the number of parts at its input queue by adding U units of parts to it. Finally, the number of parts processed by work center k during move completion time, is added to the number of parts at its output queue at Step 9. The pseudo code of the *Number in Output Estimation procedure* is as follows*:*

OQBA*: *Number in Queue Estimation Procedure

Step 1. Set $U = Min\left(C, \hat{O}_{lit}\right)$,

Step 2. Set $\Gamma = \left\{ {\tau^h}_{jkt_0}(t_p, n) \middle| t_p < t_0 + T_{liq}, j \in \{1, \ldots, N_p\} \right\}$,

$$\tilde{O}_{k,t_0+T_{liq}} = O_{k,t_0} - \sum_{{\tau^h}_{jkt_0}(t_p,n)\in\Gamma} n.$$

Step 3. Set $\Upsilon = \left\{ \gamma^h_{jkt_0}(t_d, n) \middle| t_d < t_0 + T_{liq}, j \in \{1, \ldots, N_p\} \right\}$,

$$\tilde{I}_{k,t_0+T_2} = I_{kt_0} + \sum_{\gamma^h_{jkt_0}(t_d,n)\in\Upsilon} n \ ,$$

$$W_{kt_0} = W_{kt_0} \bigcup\nolimits_{\gamma^h_{jkt_0}(t_d,n)\in M} \left\{ \overbrace{(j,t_d),\ldots,(j,t_d)}^{n} \right\}$$

Step 4. If $k = 1$, then for each part type j with mean interarrival time $\frac{1}{\lambda_j}$ Simulate n_j interarrival times $\tau_{1_j}, \ldots, \tau_{n_j}$ such that $\sum_{m_j=1}^{n_j} \tau_{m_j} < T_{liq}$.

Set $W_{kt_0} = W_{kt_0} \bigcup_{\substack{\forall m_j\in\{1,\ldots,n_j\} \\ \forall j\in\{1,\ldots,N_p\}}} \left(j, \sum_{r=1}^{m_j} \tau_{j_r}\right)$ and sort it.

Set $\tilde{I}_{k,t_0+T_{liq}} = \tilde{I}_{k,t_0+T_{liq}} + \sum_{j=1}^{N_p} n_j$.

Step 5. Set $t = t_0$, $a = 0$.

Step 6. If $\tilde{I}_{k,t_0+T_{liq}} > 0$ set $t = t_a | {W_{kt_0}}^1(j, t_a)$ and go to step 7.else go to Step 8.

Step 7. If $t < t_0 + T_{liq}$, then $\tilde{I}_{k,t_0+T_{liq}} = \tilde{I}_{k,t_0+T_{liq}} - 1$, $t = t + S(k, t_a | {W_{kt_0}}^1(j, t_a))$,

3-3-5. Algorithm three: input and output queues balancing algorithm (IOQBA)

The IOQBA looks at the combined number of parts in both the input and output queues of all work centers and selects a move which makes the combined number of parts in both queues as close as possible among all work centers. Thus the only change in procedure *Main* would be in the second step as shown below:

Step 2 of *IOQBA: Main Procedure*

Step 2. For each move request $M_{liqt_0} \in M_{t_0}$:

- Compute $T_{(l,i,q)} = \frac{DP_{d,i}}{V} + LT + \frac{PD_{i,q}}{V} + UT$,
- Call *Number in Queue Estimation procedure* to compute $\tilde{O}_{k,t_0+T_{liq}}$ and $\tilde{I}_{k,t_0+T_{liq}}$ for all $k \in \{1, \dots, N_s\}$,
- Compute:

$$E_{(l,i,q)} = -\left[\sum_{k=1}^{N_s}\left(\frac{\tilde{O}_{k,t_0+T_{liq}}}{\sum_{j=1}^{N_s}\tilde{O}_{j,t_0+T_{liq}} + \sum_{j=1}^{N_s}\tilde{I}_{j,t_0+T_{liq}}}\ln\frac{\tilde{O}_{k,t_0+T_{liq}}}{\sum_{j=1}^{N_s}\tilde{O}_{j,t_0+T_{liq}} + \sum_{j=1}^{N_s}\tilde{I}_{j,t_0+T_{liq}}}\right) + \sum_{k=1}^{N_s}\left(\frac{\tilde{I}_{k,t_0+T_{liq}}}{\sum_{j=1}^{N_s}\tilde{O}_{j,t_0+T_{liq}}+\sum_{j=1}^{N_s}\tilde{I}_{j,t_0+T_{liq}}}\ln\frac{\tilde{I}_{k,t_0+T_{liq}}}{\sum_{j=1}^{N_s}\tilde{O}_{j,t_0+T_{liq}}+\sum_{j=1}^{N_s}\tilde{I}_{j,t_0+T_{liq}}}\right)\right].$$

IOQBA *Number in Queue Estimation* procedure is the same as OQBA *Number in Queue Estimation* procedure.

In the Appendix IV, a simple numerical example demonstrates the steps of the proposed algorithms.

3-4. Comparison with other algorithms

In this section, the performance of the three proposed algorithms is compared with the following dispatching rules.

Shortest Travel Time (STT) rule: If a vehicle is idle at location d at time t_0 and $|M_{t_0}| \neq 0$ move requests have been made, the decision is to dispatch the vehicle to the

move request $\widehat{M}_{l^*i^*q^*t_0}$ such that $T_{l^*i^*q^*} = \text{Min}\{T_{liq}\}$, where T_{liq} is the total time required to accomplish the move request $\widehat{M}_{liqt_0}$.

Maximum Outgoing Queue Size (MOQS) rule: If a vehicle is idle at location d at time t_0 and $|M_{t_0}| \neq 0$ move requests have been made, the decision is to dispatch the vehicle to the move request $\widehat{M}_{l^*i^*q^*t_0}$ such that $\widehat{M}_{i^*t_0} = \text{Max}\{\widehat{M}_{it_0}\}$ and $\widehat{M}_{l^*i^*q^*t_0} = \text{Max}\{\widehat{M}_{li^*qt_0}\}$.

Modified First Come First Served (MFCFS) rule: When a part becomes available at the output queue of a work center and no vehicle is free at that time, the time that the request was generated is saved. When a vehicle becomes available, it is assigned to the work center with the earliest saved request time. Also, each work center can have at most one saved request at a time.

3-5. Simulation experiments

To evaluate the performance of the proposed algorithms, two simulation models are developed. Experiments are conducted using different system parameters for each model including: number of vehicles, vehicle capacity, and parts arrival rates to the system.

Model 1: Figure 3-1 shows the factory layout for the first model. This factory consists of 11 work centers. Parts enter the system from Work Center 1 and exit from Work Center 11. The pickup and drop-off points (circles) are arranged in a way so that when a vehicle is approaching a work center, it will reach the work center's drop-off point first. Five part types are produced in the factory. Table 3-2 shows the operation sequence and production volume percentage of each part type. It is assumed that each part type has a fixed process time at each work center. The processing time of each part

type at each work center is illustrated in Table 3-3. The average vehicle speed is 150 feet per minute and it takes 0.5 minute to perform loading or unloading operations. Simulation experiments are conducted using different values for number of vehicles (4 or 5), vehicle capacity (5, 20 or unlimited), and mean interarrival time of parts to the factory (4.6, 4.8, 5, 5.2, 5.4, 5.6, 5.8 or 6 minutes). Parts arrival to the factory is exponentially distributed. Initial simulation test runs show that the factory will be oversaturated when the mean interarrival time is smaller than 4.6 minutes.

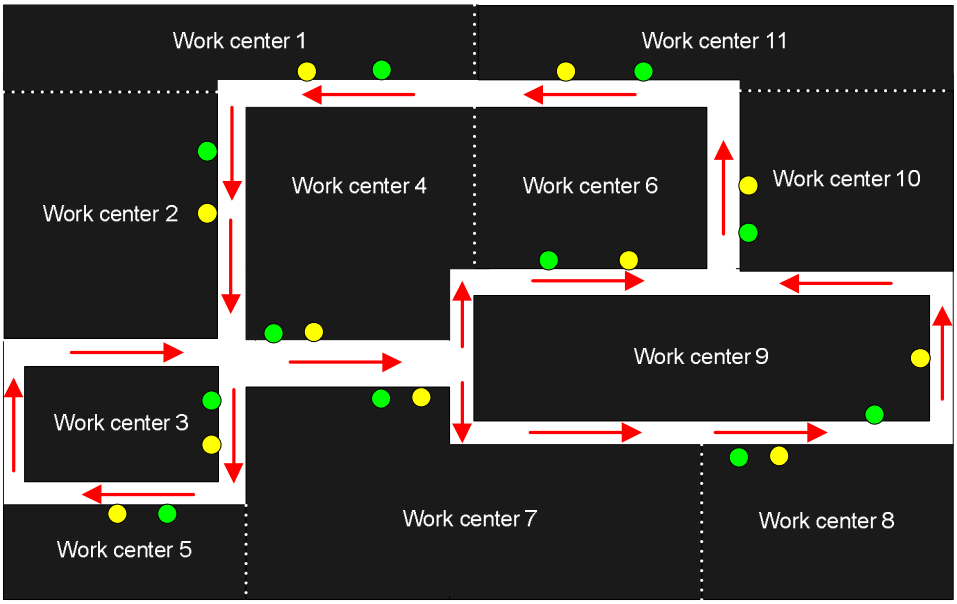


Figure 3-1. Factory layout in Model 1

Table 3-2. Operation sequences and production volume percentage of part types.

Part Type	Operation Sequence	Production Volume Percentage
1	1-2-4-9-8-10-11	%25
2	1-2-4-7-9-6-10-11	%30
3	1-2-7-9-6-10-11	%10
4	1-2-3-5-9-6-11	%10
5	1-2-4-8-10-11	%25

Table 3-3. Part types operation times at work centers (minutes)

		Work Center										
		1	2	3	4	5	6	7	8	9	10	11
Part Type	1	1.6	2	-	3.2	-	-	-	4.6	2	2	0
	2	1.6	2	-	1.3	-	4.7	2.5	-	2	2	0
	3	1.2	1.5	-	-	-	3.2	4.7	-	1.8	1.5	0
	4	1.6	2	17.2	-	20.4	8.7	-	-	1.7	-	0
	5	1.6	2	-	3.8	-	-	-	4.6	-	2	0

Model 2: This model consists of a factory producing 6 part types using 12 work centers with Work Centers 1 and 12 as entrance and exit points, respectively. The factory layout, which is adapted from Ho and Chein (2006), is shown in Figure 3-2. The pickup and drop-off points are assumed to be at the same location for each work center. Table 3-4 shows the operation sequence and production volume percentage of each part type. The processing times are normal random variables, and their distributions are shown in Table 3-5. The average vehicle speed is 120 feet per minute, and it takes 0.5 minute to perform loading or unloading operations. Like Model 1, simulation experiments are conducted for each combination of different values for different factors: number of AGVs (3 or 5), AGV capacity (5, 20 or unlimited), and mean interarrival time of parts to the system (4, 4.4, 4.8, 5.2, 5.6 or 6 minutes). Initial simulation test runs show that 4 minutes is almost the least possible value for the mean interarrival time before the system becomes over saturated.

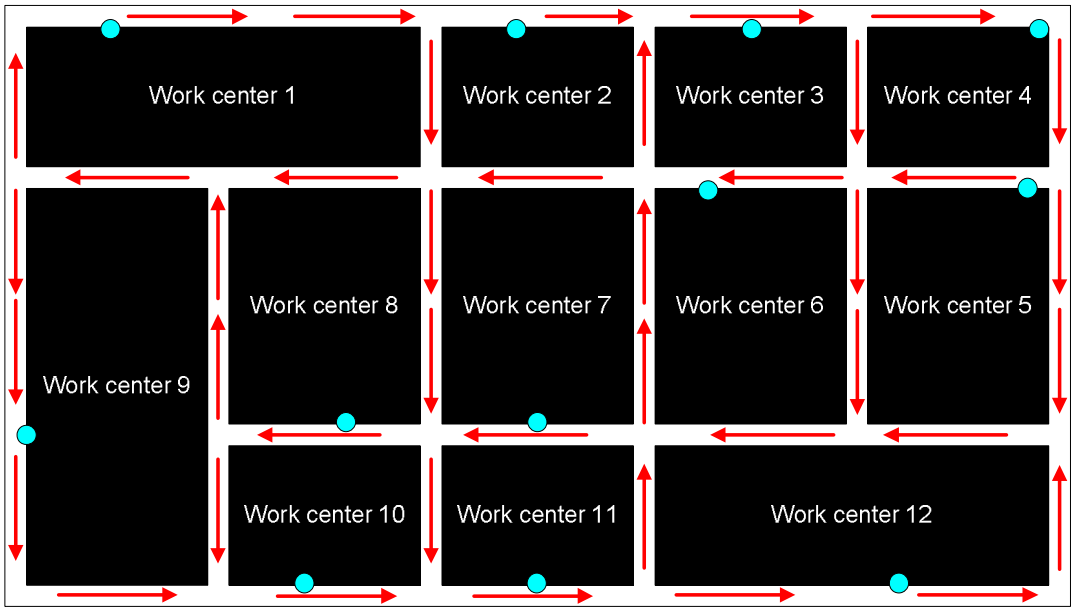


Figure 3-2. The factory layout in Model 2

Table 3-4. Operation sequences and production volume percentage of part types

Part Type	Operation Sequence	Production Volume
1	1-3-5-7-9-11-12	%16
2	1-2-4-6-8-10-12	%17
3	1-4-5-7-9-10-12	%18
4	1-3-4-5-9-11-12	%15
5	1-2-3-6-8-9-12	%14
6	1-5-6-7-10-11-12	%20

Table 3-5. Part types operation times at work centers.

N(μ, σ) represents the normal distribution with mean μ and standard deviation σ

		Work Center											
		1	2	3	4	5	6	7	8	9	10	11	12
Part Type	1	1	-	N(2.5,0.15)	-	N(3,0.1)	-	N(4,0.2)	-	N(1,0.15)	-	N(1,0.1)	0
	2	1	N(2.5,0.1)	-	N(3,0.2)	-	N(1.75,0.2)	-	N(3.5,0.15)		N(2,0.2)	-	0
	3	1	-	-	N(4,0.2)	N(1.5,0.1)	-	N(3.75,0.2)	-	N(3,0.15)	N(2,0.2)	-	0
	4	1	-	N(3.5,0.15)	N(2,0.2)	N(2.5,0.1)	-	-	-	N(3,0.15)	-	N(1,0.1)	0
	5	1	N(2,0.1)	N(4,0.15)	-	-	N(2.5,0.2)	-	N(3,0.15)	N(1,0.15)	-	-	0
	6	1	-	-	-	N(3,0.1)	N(2.5,0.2)	N(3.5,0.2)	-	-	N(2,0.2)	N(1.5,0.1)	0

Simulation experiments are conducted for each model with all combinations of different factors (number of vehicles, vehicle capacity, and mean interarrival time). The number of replications is set to be 10 for each configuration; and simulation length and warm-up time are 14,400 minutes and 960 minutes, respectively.

Table 3-6. Test statistics for the pair-wise T-test

	WCBA	OQBA	IOQBA	STT	MOQS	MFCFS
WCBA	-	-2.329	-2.535	-3.065	-9.887	-35.309
OQBA	-	-	-1.357	-0.554	1.389	-18.293
IOQBA	-	-	-	0.792	1.978	-11.204
STT	-	-	-	-	2.031	-19.063
MOQS	-	-	-	-	-	-34.004
MFCFS	-	-	-	-	-	-

The average time parts spend in the factory minus total processing times and transfer times, i.e. a part's average waiting time, is chosen as the performance measure. The waiting time average of 10 replications for various configurations of models one and two are shown in the Appendix V. To statistically compare dispatching rules, two-sided, pair-wise T-tests are applied using waiting time data for 84 configurations (48 configurations for Model 1 and 36 configurations for Model 2, shown in the Appendix V). For these tests, the hypotheses are H_0: $\mu_x = \mu_y$ and H_1: $\mu_x \neq \mu_y$ at level of significance, α, equal to 0.05. If $t_0 > |t_{0.025,83}| = 1.989$, the null hypothesis is rejected; and we conclude that there is significant difference between the two dispatching rules. Table 3-6 shows the calculated t-values for the pair-wise T-tests where *x* represents the row and *y* represents the column.

Based on the results of Table 3-6, we conclude that there is no significant difference between pairs: (OQBA and IOQBA), (OQBA and STT), (OQBA and MOQS), (IOQBA and STT), and (IOQBA and MOQS), while other pairs are significantly different. For those pairs that are significantly different, a one-sided, pair-wise T-test is applied. The one-sided test has the form of $H_0: \mu_x < \mu_y$ against $H_0: \mu_x > \mu_y$ at level of significance, α, equal to 0.05. If $t_0 > t_{0.05,83} = -1.663$, the null hypothesis is rejected; and we conclude that μ_x is significantly smaller than μ_y. Table 3-7 shows the results of performing a one-sided T-test on the pairs with mean different performance measures as identified in Table 3-6.

Table 3-7. Results of one sided, pair-wise t-tests p-values at 95% level of significance
WCBA: WCBA is better; OQBA: OQBA is better; IOQBA: IOQBA is better; STT: STT is better; MOQS: MOQS is better; MFCFS: MFCFS is better; N: No significant difference.

	WCBA	OQBA	IOQBA	STT	MOQS	MFCFS
WCBA	-	WCBA	WCBA	WCBA	WCBA	WCBA
OQBA	-	-	N	N	N	OQBA
IOQBA	-	-	-	N	N	IOQBA
STT	-	-	-	-	MOQS	STT
MOQS	-	-	-	-	-	MOQS
MFCFS	-	-	-	-	-	-

Table 3-7 shows that at 5% level of significance, WCBA performs better than all of the other rules, while OQBA's performance is not significantly different from IOQBA.

3-6. Conclusion

In this chapter we consider a discrete part manufacturing system with automated guided vehicles material handling system. Three look-ahead dispatching algorithms are proposed which predict the state of the system after each possible dispatching decision

and select the dispatch decision that contributes the most to the laminar flow of the factory. The Kullback–Leibler divergence principle is used to measure the contribution of a dispatch toward the system laminar flow. The first algorithm, *Work Content Balancing Algorithm*, attempts to balance the factory work content among all work centers by assessing the impact of increasing the work content of a work center after a dispatch. The second algorithm, *Output Queue Balancing Algorithm,* attempts to balance the number of loads in the output queue of work centers by assessing the impact of removing a load from the output queue of a work center. The third algorithm, *Input and Output Queues Balancing Algorithm,* attempts to balance the combined number of loads in the input and output queues of the work centers by assessing the impact of a potential dispatch on the factory balance. Simulation experiments have shown that WCBA performs better than the other two proposed algorithms (OQBA and IOQBA) and three simple dispatching rules (STT, MOQS and MFCFS), while OQBA's performance is not significantly different from IOQBA.

Chapter 4

An entropy-based ranking and selection method

Ranking and Selection (R&S) via simulation is a useful method to identify the best choice among several alternative systems. The selection issue tries to answer “which one of k competing systems is the “best”? “, while the ranking issue is concerned with the question: “what is the t^{th} $(1 \leq t \leq k)$ best system among k competing systems?”. In a majority of problems, "best" can be defined as the system which maximizes (or minimizes) the expected value of a performance measure. Chen et al. (2000), Chick and Inoue (2001), Kim and Nelson (2001) and Pichitlamken et al. (2006) have proposed algorithms to select the best system in which "best" is defined as the system with maximum (or minimum) mean value of the simulation output. Two major statistical approaches, frequentist (classic) and bayesian, have been used in previous papers.

The frequentist approaches (Bechhofer et al., 1995; Rinott, 1978) build a confidence interval containing the sample mean with the certain probability of correct selection. Algorithms using this approach assume that samples are independently identically (IID) normally distributed. For steady-state simulations (where data are not independent) or when the simulation data are not approximately normally distributed the method of batch means is employed (Kim and Nelson, 2001). On the other hand among those systems lying in the indifference zone, one is selected randomly and presented as the best system with a certain probability of correct selection while there might be other systems with almost the same properties.

Bayesian approaches (Chen et al., 2000; Gupta and Miescke 1994, 1996) assume a prior distribution which is supposed to describe the knowledge about the sample mean

before any sampling. The posterior, the conditional distribution of the uncertain quantity when data is given, is updated each time and the purpose is to maximize the posterior probability of correct selection with a constraint on budget. Algorithms using this method also consider the samples to be independently identically (IID) normally distributed (Chen et al., 2000; Chick and Inoue, 2001).

Both frequentist and bayesian methods consider some assumptions before performing any sampling. In practical problems no information about the systems' distributions is available and hence it would not be correct to use information such as normality of the samples or a priori distribution which is not given. The only information that can be used are descriptive statistics like sample moments obtained from observations. Moreover, none of the previous algorithms provide the mean-based ranking of systems; however not only selecting the best system, but also providing a ranking based on mean performance is useful and vital in most cases.

According to the number of sampling stages, existing Ranking and Selection procedures can be classified by two types: two-stage procedures (Nelson et al., 2001; Nelson and Staum, 2006; Tsai et al., 2009) and fully sequential procedures (Hong and Nelson, 2005, 2007; Pichitlamken et al., 2006; Tsai and Nelson, 2010). Two-stage procedures are developed based on the least-favorable-configuration assumption. Although they are easy to implement, they usually prescribe more samples than needed. In contrast, fully sequential procedures reach a selection decision more quickly, but may incur more switching cost among simulated systems.

In this research the normality assumption of the parent population which restricts the previous algorithms is relaxed. Also, the priori distribution is not assumed. The key idea

is to find, for each system, a distribution which is most ignorant using only the available information summarized in descriptive statistics. We present a two stage algorithm. In stage 1 we generate a set of observations and find the probability distribution that maximizes entropy or in other words a distribution that represents the most conservative current state of knowledge for each system. This distribution is used to determine the final number of observations required to rank systems within a specific confidence level. In the second stage, the distribution that maximizes the uncertainty of each system is obtained based on this final number of observations and directed divergence with respect to a reference distribution is used to compare the systems. Also the algorithm provides a ranking of systems based on the mean value of simulation outputs. Moreover, two systems will have the same ranking if the user is indifferent among their mean values.

The remainder of this chapter is organized as follows. Preliminaries and the two stage mean based algorithm are presented and explained in sections 4.1 and 4.2, respectively. Section 4.3 shows the performance of the algorithm using a simple example including two systems. The simulation results are presented in section 4.4. Finally, the conclusion is presented in section 4-5.

4-1. Preliminaries

Suppose that the performance measure of a system follows a probability distribution with unknown mean, μ. Assuming that n observations are generated from this system and $\bar{a}$ shows the sample mean of those n observations, as n approaches infinity, by the central limit theorem, $\sqrt{n}(\bar{a} - \mu)$ converges in distribution to a normal distribution with mean zero.

Now suppose that $P^* = (p_1^*, p_2^*, \dots, p_B^*)$ is the optimal solution of the mathematical programming model, Model (I), and $P^{**} = (p_1^{**}, p_2^{**}, \dots, p_B^{**})$ is the optimal solution of Model (I) as n approaches infinity and $\bar{a}$ converges in distribution to a normal distribution with mean μ.

$$H(P^*) = \max \sum_{i=1}^{B} -(1 + p_i)\ln(1 + p_i)$$

subject to:

$$(1) \quad \sum_{i=1}^{B} p_i = 1 \qquad \text{Model (I)}$$

$$(2) \quad \sum_{i=1}^{B} p_i x_i = \bar{a}$$

$$(3) \quad p_i \geq 0, \quad i \in \{1, \dots, B\}$$

Model (I) is a maximum entropy model (see section 1-2 for more details). The maximum entropy principle states that, subject to some known descriptive statistics such as sample mean, the probability distribution which best represents the current state of knowledge is the one with the largest entropy. The objective function of Model (I) is maximizing $\sum_{i=1}^{B} -(1 + p_i)\ln(1 + p_i)$ which is a generalized form of entropy. This form of generalized entropy is preferred because its corresponding directed divergence measure is defined even when the denominator is zero. Model (I) consists of two constraints in addition to the nonnegativity constraints. The first constraint guarantees that the sum of probabilities of all outcomes is one, $\sum_{i=1}^{B} p_i = 1$. The second constraint ensures that the mean of the optimal probability distribution which is calculated by multiplication of each possible outcome, x_i where $i = 1,2, \dots, B$ is equal to the sample mean of the generated n observations, $\bar{a}$. In addition, since the generalized form of

entropy does not guarantee that the optimal probabilities are greater than or equal to zero, nonnegativity constraints should be written to warrant that the optimal probabilities are nonnegative.

Theorem 4.1 shows that $\sqrt{n}(P^{**} - P^{*})$ converges to a multivariate normal distribution with mean zero as n goes to infinity.

Theorem 4.1.

Suppose that $a_1, a_2, \ldots, a_n$ are independent and identically distributed observations of a system which follows a distribution with mean μ. Let $\bar{a}$ be the sample mean of these observations. Moreover let $P^{*} = (p_1^{*}, p_2^{*}, \ldots, p_B^{*})$ be the optimal solution for Model (I). Moreover assume that $P^{**} = (p_1^{**}, p_2^{**}, \ldots, p_B^{**})$ is the optimal solution for model (I) if n approaches infinity. Also let λ_0 and λ_1 respectively be the lagrangian multipliers for the first and second constraints in Model (I) when n approaches infinity and $\bar{a}$ converges in distribution to a normal distribution with mean μ. $\sqrt{n}(P^{**} - P^{*}) \sim MVN(0, \Sigma)$ where $A^T = (A_1, \ldots, A_B)$ and

$$A_i = \left(\frac{(B+1)^{-1}\left(\sum_{i=1}^{B} e^{-\lambda_1 x_i}\right)^2}{(-\sum_{i=1}^{B} {x_i}^2 e^{-\lambda_1 x_i})(\sum_{i=1}^{B} e^{-\lambda_1 x_i}) + (\sum_{i=1}^{B} x_i e^{-\lambda_1 x_i})^2} \right) e^{-\lambda_0 - \lambda_1 x_i - 1} \left(\frac{\bar{a} + \sum_{i=1}^{B} x_i}{B+1} - x_i \right) \qquad 4\text{-}1$$

$$\Sigma = A^T \frac{\sum_{i=1}^{n} (a_i - \bar{a})^2}{n-1} A \qquad 4\text{-}2$$

The proof of theorem 4.1 is presented in Appendix VI.

Now let's define $KL(P, U)$ as the Kullback-Leibler directed divergence between a discrete probability distribution P with B possible outcomes and discrete uniform

distribution, U, with probability $\frac{1}{\left\lfloor \frac{B}{K} \right\rfloor}$ for the largest $\left\lfloor \frac{B}{K} \right\rfloor$ possible outcomes and zero elsewhere, where K is a positive number less than B.

$$P_U(x = x_i) = \begin{cases} \frac{1}{\left\lfloor \frac{B}{K} \right\rfloor} & if \;\; i \in \left\{ B - \left\lfloor \frac{B}{K} \right\rfloor + 1, B - \left\lfloor \frac{B}{K} \right\rfloor + 2, \dots, B \right\} \\ 0 & otherwise \end{cases}$$

In the set of possible outcomes, If large outcomes are more likely to happen than small outcomes according to P, then the directed divergence between P and U is smaller than if small outcomes are more likely to happen than large outcomes. In the other words a larger mean will result in smaller directed divergence with respect to U. $KL(P, U)$ can be written as below.

$$KL(P, U) = \sum_{i=1}^{B-\left\lfloor \frac{B}{K} \right\rfloor} (1 + p_i)\ln(1 + p_i) + \sum_{i=B-\left\lfloor \frac{B}{K} \right\rfloor + 1}^{B} (1 + p_i)\ln\left(\frac{1 + p_i}{1 + \frac{1}{\left\lfloor \frac{B}{K} \right\rfloor}} \right).$$

Theorem 4.2 shows that $\sqrt{n}\big(KL(P^*, U) - KL(P^{**}, U)\big)$ converges to a normal distribution with mean zero as n goes to infinity.

Theorem 4.2.

Define $KL(P, U) = \sum_{i=1}^{B-\left\lfloor \frac{B}{K} \right\rfloor}(1 + p_i)\ln(1 + p_i) + \sum_{i=B-\left\lfloor \frac{B}{K} \right\rfloor + 1}^{B}(1 + p_i)\ln\left(\frac{1+p_i}{1+\frac{1}{\left\lfloor \frac{B}{K} \right\rfloor}} \right)$.

The asymptotic distribution of $KL(P^*, U)$ is given by $\sqrt{n}\big(KL(P^*, U) - KL(P^{**}, U)\big)$ $\rightarrow N(0, \sigma^2)$ where:

$$\sigma^2 = h\Sigma h^T \qquad (4\text{-}3)$$

$$h_i = \begin{cases} \ln(1+p_i^*)+1 & if\ 1 \le i \le B - \left\lfloor \frac{B}{K} \right\rfloor \\ \ln(1+p_i^*)+1-\ln\left(1+\frac{1}{\left\lfloor \frac{B}{K} \right\rfloor}\right) & if\ B - \left\lfloor \frac{B}{K} \right\rfloor + 1 \le i \le B. \end{cases} \quad 4\text{-}4$$

Now Assume that we have K systems to rank based on their mean performance measure and select the system with the largest performance measure as the "best".

n observations $(a_{1k}, \dots, a_{nk})$ are generated from each system. $\bar{a}_k = \frac{a_{1k}+\cdots,+a_{nk}}{n}$ shows the sample mean of the k^{th} system. Then for each competing system Model (I) is constructed and solved. Since the resulting discrete probability distributions will be compared later with each other, we will use the same set of possible outcomes $(x_1, x_2, \dots, x_B)$ to construct Model (I) for all probability distributions. In order to find this common set of possible outcomes, we put all observations from all systems together, rank the Kn_0 observations from smallest to largest and divide them into B buckets. Bucket means are used as the possible outcomes for each probability distribution. Hence, K discrete probability distributions which maximize the generalized form of entropy are available where the mean of the k^{th} probability distribution is equal to the k^{th} system sample mean.

Now we have to figure out how many additional observations are needed in order to detect any difference larger than or equal to a user defined indifference parameter δ between mean performance measures of any pair of systems, with specific probability of correct selection, $1-\alpha$. We generate the additional number of observations, update $\bar{a}_k$, and construct and solve Model (I) for each system. Our goal is to calculate $KL(P,U)$ for each system and sort them from the smallest to the largest. We should ensure that we are able to detect any difference between $KL(P,U)$ of two systems larger than or equal to an

entropy-based indifference parameter (δ_I). Hence the mean-based user-determined indifference parameter (δ) has to be converted to an entropy-based indifference parameter. Theorem 4.3 is used to convert δ to δ_I**.** The proof is presented in Appendix VI.

Theorem 4.3. (Converting δ to δ_I)

Assume that $P^{**} = (p_1^{**}, p_2^{**}, \dots, p_B^{**})$ is the optimal solution of Model (I) for a system with unknown mean μ. Then:

$$F_Y(y) = 1 - \left(\sum_{i=1}^{\frac{\lfloor \frac{B}{K} \rfloor}{y} \ln[1+\frac{1}{\lfloor \frac{B}{K} \rfloor}]} \left(e^{-\lambda_0 - \lambda_1 x_i - 1} - 1 \right) \right)^B \quad 4\text{-}5$$

$$\frac{\delta_I}{\delta} - \lambda_1 \geq {F_Y}^{-1}(1-\alpha) \quad 4\text{-}6$$

Where λ_0, λ_1 are the Lagrangian multipliers for the first and second constraint in Model (I) respectively, and K is the number of competing systems.

Since each system has its own optimal probability distribution obtained from maximum entropy model, according to Theorem 4.3, each system k produces its own entropy-based indifference parameter (δ_I^k). Using Theorems 4.1 and 4.2 we find the additional number of buckets and additional number of observations required to be generated from each system in order to detect any difference larger than or equal to a user defined indifference parameter δ between mean performance measures of any pair of systems, with specific probability of correct selection, $1 - \alpha$.

4-2. The proposed R&S algorithm

In this section we present the new two stage algorithm for ranking and selection of K systems.

Setup

Step 1. Specify the initial number of observations (n_0), the mean based indifference parameter (δ), the number of observations in each bucket (N) and confidence interval level $1-\alpha$.

First Stage

Step 2. Generate $r = n_0$ observations from each system $k \in K$.

Step 3. Calculate the sample mean $\bar{a}_k$ of each system and set the number of buckets, $B = \left\lceil \frac{Kr}{N} \right\rceil$.

Step 4. Put all observations from all systems together, rank the Kr observations from smallest to largest and divide them into B buckets where each bucket consists of N observations except the last bucket which has $KR - (B-1)N$ observations.

Step 5. For each bucket $i, i = 1, \dots, B$ calculate the bucket mean, x_i.

Step 6. For each system $k \in K$ solve the following optimization problem to find $P_k^* = (p_{k_1}^*, \dots, p_{k_B}^*)$.

$$H(P_k^*) = \max \sum_{i=1}^{B} -\left(1 + p_{k_i}\right) \ln(1 + p_{k_i})$$

subject to:

$$(1)\quad \sum_{i=1}^{B} p_{k_i} = 1$$

$$(2)\quad \sum_{i=1}^{B} p_{k_i} x_i = \bar{a}_k$$

$$(3)\quad p_{k_i} \geq 0,\ \ i \in \{1, \dots, B\}$$

Step 7. Calculate δ_I^k, σ_k^2 and $B(k) = \left(\frac{z\alpha/_2 \sigma_k}{\delta_I^k}\right)^2$ for each system $k \in K$ and set $B^* = Max\ (\lceil B(k) \rceil)$.

Second Stage

Step 8. Generate $\left\lceil \frac{B^* N}{k} \right\rceil - r$ more observations from each system and set $r = \left\lceil \frac{B^* N}{k} \right\rceil$.

Step 9. Update the sample mean, $\bar{a}_k$, of each system and redo steps 4-6.

Ranking and Selection

Step 10. For each system $k \in K$ calculate:

$$KL(P_k^*) = \sum_{i=1}^{B-\left\lfloor \frac{B}{K} \right\rfloor} \left(1 + p_{k_i}^*\right) \ln(1 + p_{k_i}^*) + \sum_{i=B-\left\lfloor \frac{B}{K} \right\rfloor + 1}^{B} \left(1 + p_{k_i}^*\right) \ln\left(\frac{1 + p_{k_i}^*}{1 + \frac{1}{\left\lfloor \frac{B}{K} \right\rfloor}}\right)$$

Step 11. For each system $k \in K$ calculate δ_I^{*k}. Rank systems based on their $KL(P_k^*)$ value from largest to smallest. If for systems $i, j \in K$, $\left|KL(P_i^*) - KL\left(P_j^*\right)\right| \leq \min_{k \in K}(\delta_I^{*k})$, then systems i and j are non-dominated.

At the first step the initial number of observations(n_0) to be taken from each system $k \in K$, the mean based indifference parameter (δ), the number of observations in each bucket (N) and confidence interval level $1-\alpha$, are specified. n_0 observations are generated from each system in step 2 and based on these observations, the sample mean $\bar{a}_k$ of each system is calculated in step 3. Also since the total number of observations from all systems (Kn_0) and the number of observations in each bucket are known, the initial number of buckets is set to $B = \left\lceil \frac{Kn_0}{N} \right\rceil$ in step 3. Since we are going to compare the probability distributions found for each system $k \in K$, they should be found on the same set for possible outcomes. In order to find this set, we put all observations from all systems together, rank the Kn_0 observations from smallest to largest and divide them into B buckets in step 4. Moreover, in step 5, the bucket mean is calculated for each bucket. In step 6, for each system $k \in K$, we find the probability distribution which maximizes the generalized entropy measure $\sum_{i=1}^{B} -(1+p_{k_i})\ln(1+p_{k_i})$ and possesses the mean equal to system sample mean, $\bar{a}_k$. Theorem 4.3 is applied to calculate the entropy-based indifference parameter, δ_I^k, for each system $k \in K$ in step 7. Also theorem 4.2 is used to calculate the final number of buckets needed, B^*. In steps 8 and 9, the additional number of observations that has to be generated from each system $k \in K$ is found. These observations are added to the initial observations from each system $k \in K$ to update each system sample mean and the probability distribution that maximizes the entropy measure $\sum_{i=1}^{B^*} -(1+p_{k_i})\ln(1+p_{k_i})$ while possessing the mean equal to the updated system sample mean. In step 10, for each system, $KL(P_k^*)$ is calculated, which is the Kullback-

Leibler directed divergence between the maximum entropy probability distribution for each system $k \in K$ and the following discrete uniform distribution:

$$p(x = x_i) = \begin{cases} \frac{1}{\left\lfloor \frac{B^*}{K} \right\rfloor} & if\ \ i \in \left\{ B^* - \left\lfloor \frac{B^*}{K} \right\rfloor + 1, B^* - \left\lfloor \frac{B^*}{K} \right\rfloor + 2, \dots, B^* \right\} \\ 0 & otherwise \end{cases}$$

In step 11 the systems are ranked based on their $KL(P_k^*)$ and the systems i and j are considered to be non-dominated if the difference between $KL(P_i^*)$ and $KL\left(P_j^*\right)$ is less than the minimum of their entropy-based indifference parameter.

4-3. Numerical example

In this section the steps of the proposed procedure are shown by a numerical example. Assume that we have three systems with the following distributions for their performance:

$System\ \#1 \sim N(0.204, 1.0)$,

$System\ \#2 \sim N(0, 1.097)$,

$System\ \#3 \sim N(-0.204, 1.186)$,

Where $N(\mu,\ \delta^2)$ represents a normal distribution with mean μ and variance δ^2.

Suppose that the actual distributions for the performance of the above systems are unknown. The goal is to choose the systems with the smallest mean performance measure through observations being taken.

In order to perform the proposed distribution let's assume that the initial number of observations for each system under consideration is 25, the mean based indifference

parameter is 0.2041, the number of observations in each bucket is three, and the confidence interval level is 0.95.

Initial 25 observations generated from each system are shown in Table 4-1.

Table 4-1. Initial observations generated from each system under study

System #1	System #2	System #3
-1.345	-1.029	0.076
-0.183	-1.006	-1.507
0.479	0.413	2.662
1.030	0.998	-0.681
-0.775	0.169	0.881
0.100	-0.222	-0.366
0.332	1.634	1.355
0.267	-0.682	0.178
0.576	0.888	-0.770
0.100	2.117	-0.114
1.573	-0.922	0.233
0.621	0.308	-0.920
0.273	0.900	-1.215
0.498	1.347	0.573
0.677	-0.070	1.541
1.987	0.708	-1.166
0.465	-1.944	-1.719
1.715	0.065	1.840
0.521	-0.834	-0.083
1.001	-1.856	-1.156
-0.493	0.435	-0.329
-0.183	-0.945	1.477
0.220	2.669	0.636
0.785	1.211	-1.689
1.983	1.605	1.265

According to the above generated observations, $\bar{a}_1 = 0.4893, \bar{a}_2 = 0.2383, \bar{a}_3 = 0.0402$. Moreover, $B = \frac{(25)(3)}{(3)} = 25$. Hence, we put all observations from all three systems together, rank the 25 observations from the smallest to the largest and divide them into 25 buckets where each bucket consists of three observations (

Table 4-2).

Table 4-2. Three observations in each bucket and bucket sample means

	Observation #1	Observation #2	Observation #3	Bucket sample mean
Bucket #1	-1.944	-1.856	-1.719	-1.839
Bucket #2	-1.689	-1.507	-1.345	-1.513
Bucket #3	-1.215	-1.166	-1.156	-1.179
Bucket #4	-1.029	-1.006	-0.945	-0.994
Bucket #5	-0.922	-0.920	-0.834	-0.892
Bucket #6	-0.775	-0.770	-0.682	-0.742
Bucket #7	-0.681	-0.493	-0.366	-0.513
Bucket #8	-0.329	-0.222	-0.183	-0.245
Bucket #9	-0.183	-0.114	-0.083	-0.127
Bucket #10	-0.070	0.065	0.076	0.024
Bucket #11	0.100	0.100	0.169	0.123
Bucket #12	0.178	0.220	0.233	0.210
Bucket #13	0.267	0.273	0.308	0.282
Bucket #14	0.332	0.413	0.435	0.393
Bucket #15	0.465	0.479	0.498	0.481
Bucket #16	0.521	0.573	0.576	0.557
Bucket #17	0.621	0.636	0.677	0.645
Bucket #18	0.708	0.785	0.881	0.791
Bucket #19	0.888	0.900	0.998	0.929
Bucket #20	1.001	1.030	1.211	1.081
Bucket #21	1.265	1.347	1.355	1.322
Bucket #22	1.477	1.541	1.573	1.530
Bucket #23	1.605	1.634	1.715	1.651
Bucket #24	1.840	1.983	1.987	1.936
Bucket #25	2.117	2.662	2.669	2.482

Three maximum entropy models are constructed for the systems under study:

$$H(P_1) = \max \sum_{i=1}^{25} -(1+p_i)\log(1+p_i)$$

$$(1)\ \sum_{i=1}^{25} p_{1i} = 1$$

$$(2)\ \sum_{i=1}^{25} p_{1i}x_i = 0.4893$$

$$(3)\ p_{1i} \geq 0 \qquad i = 1, 2, \ldots, 25$$

Corresponding model for system #1

$$H(P_2) = \max \sum_{i=1}^{25} -(1+p_i)\log(1+p_i)$$

$$(1)\ \sum_{i=1}^{25} p_{2i} = 1$$

$$(2)\ \sum_{i=1}^{25} p_{2i}x_i = 0.2383$$

$$(3)\ p_{2i} \geq 0 \qquad i = 1, 2, \ldots, 25$$

Corresponding model for system #2

$$H(P_3) = \max \sum_{i=1}^{25} -(1+p_i)\log(1+p_i)$$

$$(1)\ \sum_{i=1}^{25} p_{3i} = 1$$

$$(2)\ \sum_{i=1}^{25} p_{3i}x_i = 0.0402$$

$$(3)\ p_{3i} \geq 0 \qquad i = 1, 2, \ldots, 25$$

Corresponding model for system #3

Figure 4-1 illustrates the optimal solutions, P_1^*, P_2^*, and P_3^*.

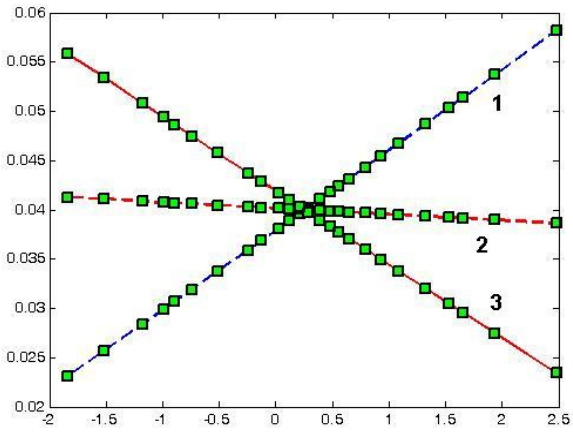


Figure 4-1. Optimal solutions for P_1^*, P_2^*, and P_3^* at the first stage

Estimated variances, σ_1^2, σ_2^2, and σ_3^2 are 8.41×10^{-6}, 4.09×10^{-5}, and 3.84×10^{-5} respectively. Moreover, δ_I^1, δ_I^2, and δ_I^3 are 4.62×10^{-5}, 8.47×10^{-5}, and 8.10×10^{-5}. Hence, $B_{(1)} = 123$, $B_{(2)} = 148$ and $B_{(3)} = 150$. Therefore, $B^* = Max\{123,148,150\} = 150$ i.e., $150 - 25 = 125$ additional observations should be generated from each system in order to be able to make decision within the 95% confidence level. Figure 4-2 shows P_1^*, P_2^*, and P_3^*, after generating the additional observations and redoing steps 4-6.

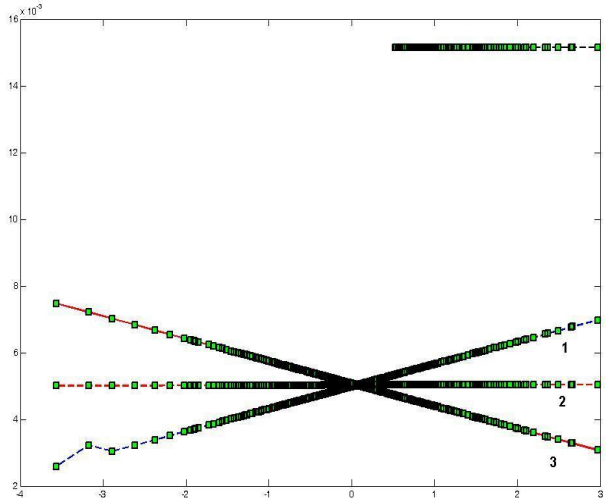


Figure 4-2. Updated P_1^*, P_2^*, and P_3^* at the second stage

In addition to P_1^*, P_2^*, and P_3^*, Figure 4-2 shows the reference discrete uniform distribution, $p(x = x_i) = \begin{cases} \frac{1}{50} & if \ \ i \in \{101,102, \dots, 150\} \\ 0 & otherwise \end{cases}$. Finally $U(P_k^*)$, which is the Kullback-Leibler directed divergence between P_k^* and the discrete uniform distribution is computed and the three systems are ranked based on their $KL(P_k^*)$. $U(P_1^*) = 0.0296$, $KL(P_2^*) = 0.0419$ and $KL(P_3^*) = 0.0503$. Since the difference between each pair is significantly larger than the minimum of their entropy-based indifference parameter and $KL(P_1^*) < KL(P_2^*) <$

$KL(P_3^*)$, we conclude that within 95% confidence level, system #1 has the largest mean, system #2 has the second largest mean, and system 3 possesses the smallest mean among these three systems.

4-4. Simulation experiments

In order to investigate the performance of the proposed algorithm simulation experiments are conducted. The number of systems in each experiment $K = 3$ or $K = 5$. We choose, the initial number of observations from each system to be $n_0 = 24$, the mean based indifference parameter to be $\delta = 0.204$, the number of observations in each bucket to be $N = 1, 2, 3$ and the confidence interval level to be $1 - \alpha = 0.95$.

Two configurations of the true means are used suggested by Kim and Nelson (2001):

- Slippage configuration (SC), in which μ_1 is set to δ and $\mu_2 = \mu_3 = \cdots = \mu_K = 0$. According to Kim and Nelson (2001), "*this is a difficult configuration for procedures that try to eliminate systems because all of the inferior systems are close to the best*".
- Monotone Decreasing Means configuration (MDM), in which the means of all systems were spaced evenly apart according to the following formula $\mu_i = \mu_1 - \mathrm{a}(i-1)$, where $a = \delta$. Kim and Nelson (2001) state that *this* configuration is good *"to investigate the effectiveness of the procedures in eliminating noncompetitive systems"*.

For each configuration of the means, the variance of the best system is set both higher and lower than the variances of the other systems. Hence experiments are run with

variances of all systems either monotonically decreasing or monotonically increasing. The results are shown in Table 4-3:Table 4-10.

Table 4-3. Simulation results for MDM configuration with increasing variance when K=3
NS: Neighborhood Switch

MDM Configuration, Increasing Variance, with K=3					
True Mean	True Variance	True Ranking	Ranking given by algorithm		
			N=1	N=2	N=3
0.204	1.000	1	1.06	1.03	1.02
0.000	1.097	2	1.94	1.96	1.96
-0.204	1.187	3	2.902	2.92	2.94
Average number of observations			128	134	145
Selecting the best system			92%	94%	96%
Correct Ranking			91%	94%	95%
Correct Ranking after NS			100%	100%	100%

Table 4-4. Simulation results for MDM configuration with decreasing variance when K=3
NS: Neighborhood Switch

MDM Configuration, Decreasing Variance, with K=3					
True Mean	True Variance	True Ranking	Ranking given by algorithm		
			N=1	N=2	N=3
0.204	1.000	1	1.02	1	1
0.000	0.911	2	1.92	1.94	1.96
-0.204	0.843	3	2.88	2.96	2.98
Average number of observations			164	178	196
Selecting the best system			93%	100%	100%
Correct Ranking			92%	94%	98%
Correct Ranking after NS			99%	100%	100%

Table 4-5. Simulation results for SC configuration with increasing variance when K=3
NS: Neighborhood Switch

SC Configuration, Increasing Variance, with K=3					
True Mean	True Variance	True Ranking	Ranking given by algorithm		
			N=1	N=2	N=3
0.204	1.000	1	1	1	1
0.000	1.097	2	2.04	2.03	2.02
0.000	1.187	2	2.05	2.04	2.02
Average number of observations			129	142	157
Selecting the best system			100%	100%	100%
Correct Ranking			95%	97%	97%
Correct Ranking after NS			100%	100%	100%

Table 4-6. Simulation results for SC configuration with decreasing variance when K=3
NS: Neighborhood Switch

SC Configuration, Decreasing Variance, with K=3					
True Mean	True Variance	True Ranking	Ranking given by algorithm		
			N=1	N=2	N=3
0.204	1.000	1	1	1	1
0.000	0.911	2	2.04	2.04	2.01
0.000	0.843	2	2.03	2.02	2.02
Average number of observations			179	184	193
Selecting the best system			100%	100%	100%
Correct Ranking			95%	96%	98%
Correct Ranking after NS			100%	100%	100%

Table 4-7. Simulation results for MDM configuration with increasing variance when K=5
NS: Neighborhood Switch

MDM Configuration, Increasing Variance, with K=5					
True Mean	True Variance	True Ranking	Ranking given by algorithm		
			N=1	N=2	N=3
0.204	1.000	1	1.02	1.01	1.01
0.000	1.097	2	1.87	1.92	1.96
-0.204	1.187	3	2.84	2.92	2.94
-0.408	1.270	4	3.87	3.91	3.94
-0.612	1.348	5	4.85	4.87	4.91
Average number of observations			398	415	431
Selecting the best system			91%	93%	96%
Correct Ranking			76%	78%	82%
Correct Ranking after NS			98%	100%	100%

Table 4-8. Simulation results for MDM configuration with decreasing variance when K=5
NS: Neighborhood Switch

MDM Configuration, Decreasing Variance, with K=5					
True Mean	True Variance	True Ranking	Ranking given by algorithm		
			N=1	N=2	N=3
0.204	1.000	1	1.02	1.02	1.01
0.000	0.911	2	1.84	2.02	2.02
-0.204	0.843	3	2.86	2.96	2.98
-0.408	0.788	4	4.1	4.05	4.01
-0.612	0.742	5	4.8	4.85	4.99
Average number of observations			795	857	985
Selecting the best system			93%	96%	98%
Correct Ranking			84%	89%	96%
Correct Ranking after NS			99%	100%	100%

Table 4-9. Simulation results for SC configuration with increasing variance when K=5
NS: Neighborhood Switch

SC Configuration, Increasing Variance, with K=5					
True Mean	True Variance	True Ranking	Ranking given by algorithm		
			N=1	N=2	N=3
0.204	1.000	1	1.08	1.05	1.04
0.000	1.097	2	1.85	1.91	1.94
0.000	1.187	2	1.96	1.94	1.98
0.000	1.270	2	2.08	2.1	2.02
0.000	1.348	2	2.16	2.1	2.04
Average number of observations			505	581	614
Selecting the best system			89%	94%	96%
Correct Ranking			62%	71%	75%
Correct Ranking after NS			90%	95%	99%

Table 4-10. Simulation results for SC configuration with decreasing variance when K=5
NS: Neighborhood Switch

SC Configuration, Decreasing Variance, with K=5					
True Mean	True Variance	True Ranking	Ranking given by algorithm		
			N=1	N=2	N=3
0.204	1.000	1	1	1	1
0.000	0.911	2	1.8	1.82	1.89
0.000	0.843	2	1.84	1.91	1.92
0.000	0.788	2	1.99	2.12	1.95
0.000	0.742	2	2.14	2.08	2.04
Average number of observations			512	621	686
Selecting the best system			88%	90%	94%
Correct Ranking			65%	69%	77%
Correct Ranking after NS			92%	99%	100%

According to Table 4-3: Table 4-6, with three competing systems, when number of observations in each bucket is one, the probability of correct selection is less than 95%

for MDM configuration with increasing and decreasing variances and more than 95% for SC configuration with both increasing and decreasing variances. Increasing the number of observations in each bucket, increases the probability of correct selection to more than 95% in all cases, however it generates more observations from each system. Moreover, when there is one observation in each bucket, the probability of obtaining the correct ranking is less than or equal to 95% in all configurations. Increasing the number of observations in each bucket will increase the probability of obtaining the correct ranking to more than 99%.

According to Table 4-7: Table 4-10, with five competing systems, when number of observations in each bucket is one, the probability of correct selection is less than 95% in all configurations. Increasing the number of observations in each bucket to N=3, increases the probability of correct selection to more than 95% all configurations except one configuration. It can be shown than N should be at least equal to 5 in order to obtain 95% correct selection in all configurations. Moreover, when there is one observation in each bucket, the probability of obtaining the correct ranking is less than 85% in all configurations. Increasing the number of observations in each bucket will increase the probability of obtaining the correct ranking. However, with N=3, the probability of correct ranking is still less than 85% for three configurations.

If $(r_1, r_2, \dots r_i, r_{i+1}, \dots, r_K)$ is the ranking obtained from the algorithm, then $(r_1, r_2, \dots r_{i+1}, r_i, \dots, r_K)$ is defined as the ranking after neighborhood switch. If we perform neighborhood switch after obtaining the ranking, when K=3, then the probability of obtaining the correct ranking after one neighborhood switch is more than 99% in all configurations even when N=1. When K=5, after performing the neighborhood switch,

the probability of obtaining the correct ranking is more than 90% when N=1 and more than or equal to 99% when N=3.

4-5. Conclusion

In this chapter we used the maximum entropy principle to present a two-stage ranking and selection algorithm with the following contributions:

a) Relaxing the normality assumption which restricts the previous algorithms.

b) No a priori distribution is assumed.

c) Providing a ranking of systems based on the mean value of simulation outputs, and Moreover two systems will have the same ranking if the user is indifferent among their mean values.

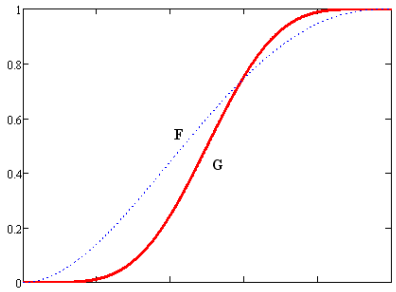
At the first stage of this two-stage algorithm, based on the initial observations obtained at the first stage, the probability distribution which maximizes uncertainty or in other words a distribution which represents the current state of knowledge for each system is found via using maximum entropy principle. This distribution is used to determine the final number of observations required to rank systems in an specific confidence level and the distribution which maximizes the uncertainty of each system is obtained based on this final number of observations. The second stage applies directed divergence on distributions found in the first stage to compare the systems. The efficiency of the offered algorithm was shown by simulation experiments.

Chapter 5

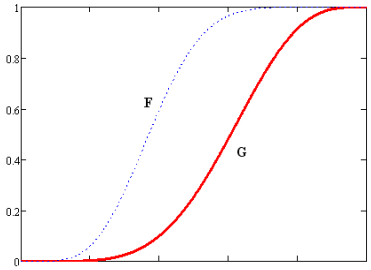
Entropy-Based measure for stochastic dominance

Let F and G be two cumulative probability distributions and let s be a positive integer. For each distribution function H let $D_H^1(x) = H(x)$ and for $i \in \{2,3, \dots\}$ let $D_H^i(x) = \int_{-\infty}^{x} D_H^{i-1}(z)dz$. F is said to be dominated by G by stochastic dominance of order s if $D_G^s(x) \leq D_F^s(x)$ for all values of x with strict inequality for at least for one point x_0(see for example Levy (1992)). In the past two decades stochastic dominance has been widely applied in different areas such as finance, risk management, and economics (see Sriboonchita et al. (2009)).

Two possible cases for first order stochastic dominance (FSD) are illustrated in Figure 5-1. Two possible cases for FSDIn Figure 5-1.a, F and G are nondominant based on FSD because for some values of x, $F(x) < G(x)$ and for some other values $F(x) > G(x)$. In Figure 5-1.b, G dominates F based on FSD since for all values of x, $G(x) \leq F(x)$ with strict inequality at least for one point.



a. F and G are nondominant based on FSD

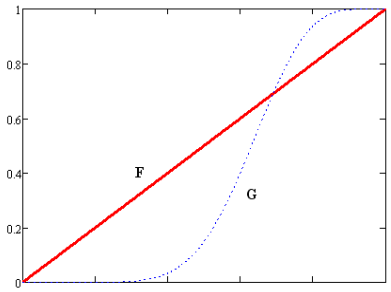


b. G dominates F based on FSD

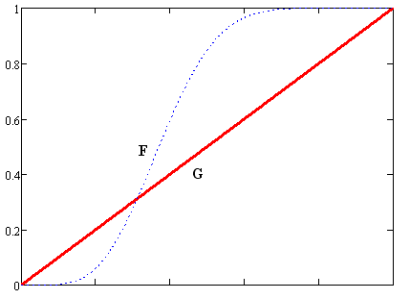
Figure 5-1. Two possible cases for FSD

Two possible cases for second order stochastic dominance (SSD) are shown Figure 5-2. In Figure 5-2.a, G dominates F based on SSD since for all values of x,

$\int_{-\infty}^{x} G(x)dx \leq \int_{-\infty}^{x} F(x)dx$ with strict inequality at least for one point. In Figure 5-2.b, F and G are nondominant based on SSD because for some values of x, $\int_{-\infty}^{x} G(x)dx < \int_{-\infty}^{x} F(x)dx$ and for some other values $\int_{-\infty}^{x} G(x)dx > \int_{-\infty}^{x} F(x)dx$.



a. G dominates F based on SSD



b. F and G are nondominant based on SSD

Figure 5-2. Two possible cases for SSD

Although stochastic dominance have been widely used to rank probability distributions (see Levy (1992, 1998) for surveys on stochastic dominance), they may be unable to determine dominance even in situations when most decision makers would prefer one alternative over another. For example according to Leshno and Levy (2002) suppose that two lotteries are available. Lottery X returns zero dollars with probability 0.01 and returns one million dollars with probability 0.99, while lottery Y yields one dollar with probability 1.00. Most individuals prefer X to Y, however X does not stochastically dominates Y according to any stochastic dominance rule of order s. This drawback has motivated Leshno and Levy (2002) to offer stochastic dominance rules weaker than FSD or SSD. They offered a stochastic dominance criterion that applies the proportion of the two bounded areas, created by the crossing of the two probability distributions, to the sum of the two areas. They named their criterion p almost first order stochastic dominance (p-AFSD) which has been applied by several researchers such as

Bali et al. (2009, 2011), Tzheng et al. (2012), and Levy et al. (2010). The main advantage of p-AFSD is the possibility for reduction of a set of nondominant alternatives.

We offer a stochastic dominance criterion weaker than FSD and SSD based on the concept of entropy. Like p-AFSD, its advantage over the stochastic dominance criteria is to reduce the possibility of obtaining a nondominated set of alternatives. Although, unlike p-AFSD which compares alternatives based on surrendered areas and requires calculation of the corresponding areas, based on the new offered criterion alternatives are compared to a utopia and the alternative which closer to the utopia based on our definition of closeness is chosen. The new criterion is named entropy-based stochastic dominance (ESD). We investigate its relation with FSD, SSD, and p-AFSD. The remainder of this chapter is organized as follows. In section 5-1, Kullback and Leibler information is explained. Entropy-based stochastic dominance measure is presented in section 5-2. In section 5-3 the relationship between ESD and FSD, SSD, and p-AFSD is investigated. Finally conclusion is explained in section 5-4.

5-1. Kullback-Leibler information

Kullback-Leibler (KL) information is a measure (a 'distance' in an heuristic sense) between an observed distribution, f, and a reference distribution, g (see section 1-3 for more details). Being denoted by $KL(f, g)$, it measures the amount of 'information' lost when distribution f is used to approximate a reference distribution, g. The analyst seeks a distribution that its information content is as close as possible to the reference distribution. Moreover, their measure possesses several useful properties (see Appendix

II) including the fact that it is always nonnegative and it is equal to zero if and only if $f(x) = g(x)$.

Several authors have studied extensions of KL information. Asadi et al. (2004) considered KL information for a residual distribution function with density function $\frac{f(x)}{\bar{F}(t)}$ where $x \geq t$ where F is the cumulative probability distribution and $\bar{F}(t) = 1 - F(x)$. Barapour and Rad (2012) have proposed an extension of KL information for a nonnegative random variable, called cumulative residual KL information, in terms of survival function as follows.

$$CRKL(\bar{F}, \bar{R}) = \int_0^\infty \bar{F}(x) \ln \frac{\bar{F}(x)}{\bar{G}(x)} dx - (E(X) - E(Y))$$

Where $E(X) = \int_0^\infty \bar{F}(x)\, dx$ and $E(Y) = \int_0^\infty \bar{G}(x)\, dx$.

Park et al. (2012) have considered another extension, called cumulative KL information, in terms of cumulative distribution function as

$$CKL(F, G) = \int_0^\infty F(x) \ln \frac{F(x)}{G(x)} dx - (E(X) - E(Y))$$

They showed that CKL is nonnegative and is equal to zero if and only if $F(x) = G(x)$. Keeping these two properties (nonnegativity and being equal to zero if and only if $F(x) = G(x)$) are the main reason that the above authors did not use $\int_0^\infty F(x) \ln \frac{F(x)}{G(x)} dx$, as it does not keep these well-known KL information properties. If for all x, $G(x) \leq F(x)$ then since $\int_0^\infty F(x) \ln \frac{F(x)}{G(x)} dx$ is nonnegative and right continuous, when $\int_0^\infty F(x) \ln \frac{F(x)}{G(x)} = 0$ we have $F(x) \ln \frac{F(x)}{G(x)}$ for all x. Thus we have either $F(x) = 0$ which gives $G(x) = 0$ (because $0 \leq G(x) \leq F(x)$) or we have $\ln \frac{F(x)}{G(x)} = 0$ which means that

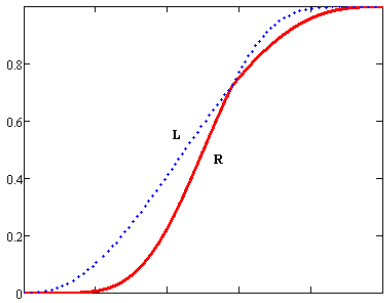
$F(x) = G(x)$. Throughout this chapter we use $\int_0^\infty F(x) \ln \frac{F(x)}{G(x)} dx$ as our measure for directed divergence between F and G.

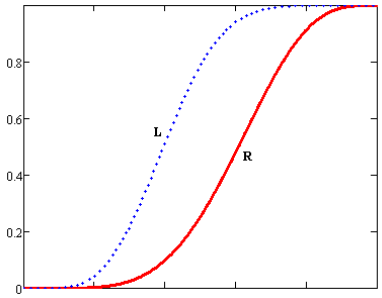
5-2. Entropy-based criterion for stochastic dominance (ESD)

We use $a \vee b$ to denote $max(a, b)$ and use $a \wedge b$ to denote $min(a, b)$. For two functions, f and g, the function $f \vee g$ is called the upper envelope, or the left envelope of f and g, and the function $f \wedge g$ is called the lower envelope, or the right envelope of f and g.

For a real valued function f we use $\int f$ to denote $\int_{-\infty}^{\infty} f(x)dx$. We use $\int_A f$ to denote the Lebesgue integral $\int_A f(x)d\lambda(x)$, where λ denotes the Lebesgue measure on the line. For two functions, f and g, the set $\{x: f(x) < g(x)\}$ is denoted by $[f < g]$. We will abbreviate $[f < g]$ to denote $f < g$ when we write an integral on $[f < g]$. The notation $[f > g]$ is used in a similar way.

Assume that X and Y are two nonnegative continuous random variables with cumulative probability distributions F and G respectively and let $L = F \vee G$ denote their left envelope and let $R = F \wedge G$ denote their right envelope. Figure 5-3 shows right and left envelopes of distributions shown in Figure 5-1.





a. Right and left envelopes of distributions in Figure 5-1.a b. Right and left envelopes of distributions in Figure 5-1.b

Figure 5-3. Right and left envelopes of distributions in Figure 5-1

Definition.

Let H be a reference cumulative probability distribution.

We define F is closer to H than G by ESD, if

$$\left|\int F \ln\left(\frac{F}{H}\right)\right| \leq \left|\int G \ln\left(\frac{G}{H}\right)\right| \qquad 5\text{-}1$$

When (5-1) is not true, we say F is not closer to H than G.

We say F strictly dominates G by ESD if $\left|\int F \ln\left(\frac{F}{R}\right)\right| = 0$ and $\left|\int G \ln\left(\frac{G}{R}\right)\right| \neq 0$.

We say F weakly dominates G by ESD when one of the following is true.

(i) F is closer to R and not closer to L than G.

(ii) G is closer to L and not closer to R than F.

We say F and G are nondominated by ESD when one of the following is true.

(i) F is closer to R and L than G.

(ii) G is closer to R and L than F.

The following two propositions are useful when two distributions are compared by ESD.

Proposition 5-1. G is closer to R than F if and only if $\int (F \vee G) \ln\left(\frac{G}{F}\right) \leq 0$.

Proof. G is closer to R than F when

$\left|\int G \ln\left(\frac{G}{R}\right)\right| \leq \left|\int F \ln\left(\frac{F}{R}\right)\right|$ which means that $\int G \ln\left(\frac{G}{R}\right) \leq \int F \ln\left(\frac{F}{R}\right)$, because the integrands and hence the integrals on both sides are nonnegative. This means

$$\int_{F<G} G \ln\left(\frac{G}{R}\right) + \int_{F>G} G \ln\left(\frac{G}{R}\right) \leq \int_{F<G} F \ln\left(\frac{F}{R}\right) + \int_{F>G} F \ln\left(\frac{F}{R}\right),$$

$$\int_{F<G} G \ln\left(\frac{G}{F}\right) + \int_{F>G} G \ln\left(\frac{G}{G}\right) \leq \int_{F<G} F \ln\left(\frac{F}{F}\right) + \int_{F>G} F \ln\left(\frac{F}{G}\right),$$

$$\int_{F<G} G \ln\left(\frac{G}{F}\right) \leq \int_{F>G} F \ln\left(\frac{F}{G}\right),$$

$$\int_{F<G} G \ln\left(\frac{G}{F}\right) + \int_{F>G} F \ln\left(\frac{G}{F}\right) \leq 0,$$

Note that on the set $[F > G]$ we have $F = F \vee G$ and on the set $[F < G]$ we have $G = F \vee G$. Thus,

$$\int_{F<G} (F \vee G) \ln\left(\frac{G}{F}\right) + \int_{F>G} (F \vee G) \ln\left(\frac{G}{F}\right) \leq 0,$$

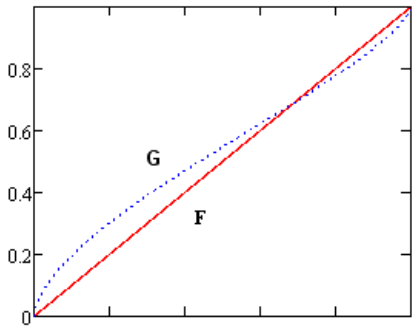
$$\int (F \vee G) \ln\left(\frac{G}{F}\right) \leq 0. \blacksquare$$

Proposition 5-2. G is closer to R than F if and only if $\int (F \wedge G) \ln\left(\frac{G}{F}\right) \geq 0$.

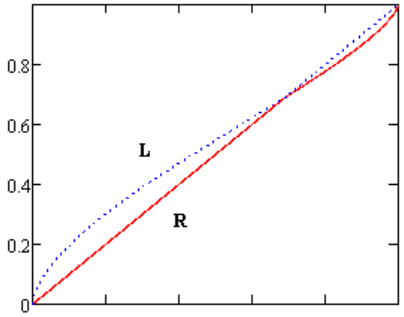
The proof of Proposition 2 is similar to the proof of proposition 1.

Consider the case when $sup\{x: H(x) = 0\} > sup\{x: F(x) = 0\}$ then there exists some points x such that $F(x) > 0$ and $H(x) = 0$. Hence, $\int F \ln\left(\frac{F}{H}\right)$ will be infinite. In this case inverse cumulative distribution functions are compared by ESD instead of cumulative distribution functions. We interpret the results as follows: If G^{-1} (strictly or

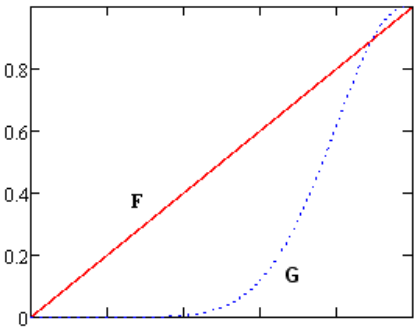
weakly) dominates F^{-1} by ESD then F dominates G and if G^{-1} and F^{-1} are nondominated then G and F are nondominated as well. Three different scenarios are illustrated in Figure 5-4.



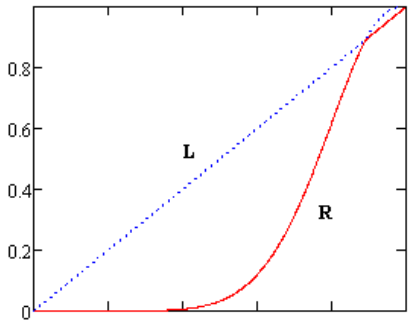
a. F and G in scenario I



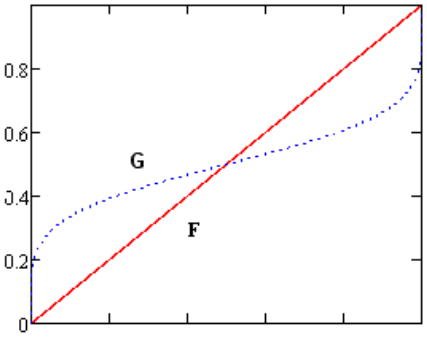
b. R and L in scenario I



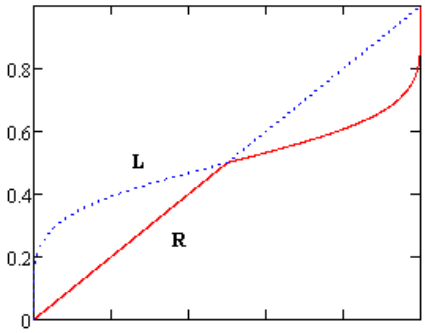
c. F and G in scenario II



d. R and L in scenario II



e. F and G in scenario III



f. R and L in scenario III

Figure 5-4. Three different scenarios when ESD is measured

In the first and third scenarios (parts a and b, and e and f of Figure 5-4), $sup\{x: R(x) = 0\}, sup\{x: F(x) = 0\}, sup\{x: G(x) = 0\}$, and $sup\{x: L(x) = 0\}$ are

equal, therefore $\int F \ln\left(\frac{F}{R}\right)$, $\int G \ln\left(\frac{G}{R}\right)$, $\int F \ln\left(\frac{F}{L}\right)$, and $\int G \ln\left(\frac{G}{L}\right)$ have finite values. In the second scenario $sup\{x: R(x) = 0\} > sup\{x: F(x) = 0\}$. Thus, $\int F \ln\left(\frac{F}{R}\right)$ is not defined in real numbers.

5-3. Relationship between ESD and FSD, SSD, and p-AFSD

In this section the relationship between ESD, FSD, SSD, and p-AFSD is studied. Throughout this section assume that F and G are two continuous cumulative distributions.

Theorem 5-1. F strictly dominates G according to ESD if and only if F dominates G according to FSD.

Proof. If F strictly dominates G by ESD, then $\left|\int F \ln\left(\frac{F}{F \wedge G}\right)\right| = 0$ and $\left|\int G \ln\left(\frac{G}{F \wedge G}\right)\right| \neq 0$. This implies that $F = F \wedge G$ and $\mathrm{G} > F \wedge G$. Hence, F dominated G by FSD ($F < G$). Inversely, if F dominated G then $F \leq G$ with strict inequality at least for one point implying that $F = F \wedge G$ and $\mathrm{G} > F \wedge G$. Therefore, $\left|\int F \ln\left(\frac{F}{F \wedge G}\right)\right| = 0$ and $\left|\int G \ln\left(\frac{G}{F \wedge G}\right)\right| \neq 0$ which means that F strictly dominates G by ESD.■

Proposition 5-3.

$$\int G \, ln \, \frac{G}{F} \geq \int (G - F) \qquad 5\text{-}2$$

Proof. Applying Taylor series expansion to $\ln x$ at $x = 1$, we obtain: $\ln x = x - 1 - \frac{(x-1)^2}{2\xi^2}$ where ξ is a point between 1 and x. This implies that $\ln x \leq x - 1$ for all $x > 0$ where equality holds when $x = 1$. Thus,

$$\int G \ln \frac{F}{G} \leq \int G\left(\frac{F}{G} - 1\right) = \int (F - G)$$

Therefore,

$$\int G \ln \frac{G}{F} = -\int G \ln \frac{F}{G} \geq -\int (F - G) = \int (G - F)\,. \blacksquare$$

Conjecture 5-1. If F dominates G by ESD, then G does not dominate F by SSD.

Conversely, if F dominates G by SSD, then G does not dominate F by ESD.

Assume that F dominates G by ESD and

$$\left|\int F \ln\left(\frac{F}{F \wedge G}\right)\right| \leq \left|\int G \ln\left(\frac{G}{F \wedge G}\right)\right| \text{ and } \left|\int F \ln\left(\frac{F}{F \vee G}\right)\right| > \left|\int G \ln\left(\frac{G}{F \vee G}\right)\right|.$$

$$\int F \ln\left(\frac{F}{F \wedge G}\right) \leq \int G \ln\left(\frac{G}{F \wedge G}\right) \text{ and } \int F \ln\left(\frac{F \vee G}{F}\right) > \int G \ln\left(\frac{F \vee G}{G}\right),$$

$$\int_{F>G} F \ln\left(\frac{F}{G}\right) \leq \int_{F<G} G \ln\left(\frac{G}{F}\right) \tag{5-3}$$

and,

$$\int_{F<G} F \ln\left(\frac{G}{F}\right) > \int_{F>G} G \ln\left(\frac{F}{G}\right) \tag{5-4}$$

According to proposition 5-3,

$$\int_{F>G} F \ln\left(\frac{F}{G}\right) \geq \int_{F>G} (F - G) \tag{5-5}$$

$$\int_{F<G} F \ln\left(\frac{G}{F}\right) \leq \int_{F<G} (G - F) \tag{5-6}$$

If G dominates F based on SSD then

$$\int_{F>G} (F - G) > \int_{F<G} (G - F) \tag{5-7}$$

From (5-5), (5-6) and (5-7) we have

$$\int_{F<G} F \ln\left(\frac{G}{F}\right) \leq \int_{F<G} (G-F) < \int_{F>G} (F-G) \leq \int_{F>G} F \ln\left(\frac{F}{G}\right) \quad 5\text{-}8$$

Also from (5-3), (5-4), and (5-8) we obtain

$$\int_{F>G} G \ln\left(\frac{F}{G}\right) < \int_{F<G} F \ln\left(\frac{G}{F}\right) \leq \int_{F<G} (G-F) < \int_{F>G} (F-G) \leq \int_{F>G} F \ln\left(\frac{F}{G}\right) \leq \int_{F<G} G \ln\left(\frac{G}{F}\right) \quad 5\text{-}9$$

or by removing the third inequality,

$$\int_{F>G} G \ln\left(\frac{F}{G}\right) < \int_{F<G} F \ln\left(\frac{G}{F}\right) < \int_{F>G} F \ln\left(\frac{F}{G}\right) \leq \int_{F<G} G \ln\left(\frac{G}{F}\right) \quad 5\text{-}10$$

We tried a wide range of distributions and could not find any pair of distributions that satisfy inequalities in (5-9) or (5-10). But we could not bring a proof that shows inequalities in (5-9) or (5-10) never occur. We show by three sets of distributions that G does not dominate F by SSD if F dominates G by ESD. Similarly we show that if F dominates G by SSD, then G does not dominate F by ESD.

Distributions set #1 consists of negatively skewed unimodal beta cumulative distributions and a uniform [0,1] cumulative distribution which are shown in Figure 5-5. These negatively skewed unimodal beta cumulative distributions are compared to the uniform cumulative distributions based on SSD and ESD. The comparison results are summarized in Table 1. As shown in Table 5-1, all the negatively skewed unimodal beta cumulative distributions dominate the uniform [0,1] cumulative distribution by SSD. According to the ESD rule, the unimodal beta distributions dominate the uniform cumulative distribution except when no significance difference is distinguished. In that case two distributions are nondominated. Hence when unimodal beta distributions dominate the uniform distributions based on SSD then, then the uniform distribution does not dominate the unimodal beta distributions based on ESD. Assuming that G is a

negatively skewed unimodal beta cumulative distribution and F is uniform [0,1] cumulative distribution, computational results for $\left|\int F \ln\left(\frac{F}{R}\right)\right|$, $\left|\int G \ln\left(\frac{G}{R}\right)\right|$, $\left|\int F \ln\left(\frac{F}{L}\right)\right|$, and $\left|\int G \ln\left(\frac{G}{L}\right)\right|$ which led to the decisions by ESD in Table 5-1 are shown in Table 5-2.

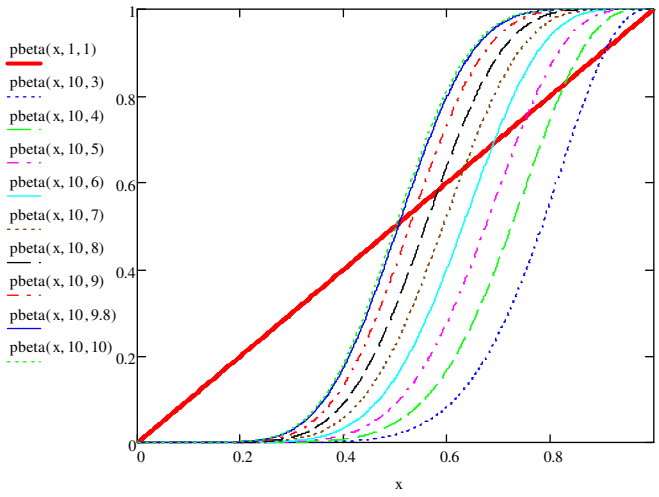


Figure 5-5. distributions set #1: negatively skewed unimodal beta and uniform [0,1] cumulative distributions

Table 5-1. Comparison between SSD and ESD rules on distributions set #1

	SSD	*ESD*
$Beta(1,1)\ vs\ Beta\ (10,3)$	$Beta(10,3)$ is dominant	$Beta(10,3)$ is dominant
$Beta(1,1)\ vs\ Beta\ (10,4)$	$Beta(10,4)$ is dominant	$Beta(10,4)$ is dominant
$Beta(1,1)\ vs\ Beta\ (10,5)$	$Beta(10,5)$ is dominant	$Beta(10,5)$ is dominant
$Beta(1,1)\ vs\ Beta\ (10,6)$	$Beta(10,6)$ is dominant	$Beta(10,6)$ is dominant
$Beta(1,1)\ vs\ Beta\ (10,7)$	$Beta(10,7)$ is dominant	$Beta(10,7)$ is dominant
$Beta(1,1)\ vs\ Beta\ (10,8)$	$Beta(10,8)$ is dominant	$Beta(10,8)$ is dominant
$Beta(1,1)\ vs\ Beta\ (10,9)$	$Beta(10,9)$ is dominant	$Beta(10,9)$ is dominant
$Beta(1,1)\ vs\ Beta\ (10,9.8)$	$Beta(10,9.8)$ is dominant	Nondominated set
$Beta(1,1)\ vs\ Beta\ (10,10)$	$Beta(10,10)$ is dominant	Nondominated set

Table 5-2. computational results for $\left|\int F \ln\left(\frac{F}{R}\right)\right|$, $\left|\int G \ln\left(\frac{G}{R}\right)\right|$, $\left|\int F \ln\left(\frac{F}{L}\right)\right|$, and $\left|\int G \ln\left(\frac{G}{L}\right)\right|$ of decisions made in Table 5-1 based on ESD, Assuming that G is a negatively skewed unimodal beta cumulative distribution and F is uniform [0,1] cumulative distribution.

	$\left\|\int G \ln\left(\frac{G}{L}\right)\right\|$	$\left\|\int F \ln\left(\frac{F}{L}\right)\right\|$	$\left\|\int G \ln\left(\frac{G}{R}\right)\right\|$	$\left\|\int F \ln\left(\frac{F}{R}\right)\right\|$
$Beta(1,1)\ vs\ Beta\ (10,3)$	0.158	0.002	0.002	0.428
$Beta(1,1)\ vs\ Beta\ (10,4)$	0.130	0.007	0.007	0.381
$Beta(1,1)\ vs\ Beta\ (10,5)$	0.108	0.015	0.017	0.313
$Beta(1,1)\ vs\ Beta\ (10,6)$	0.090	0.026	0.029	0.260
$Beta(1,1)\ vs\ Beta\ (10,7)$	0.076	0.037	0.044	0.218
$Beta(1,1)\ vs\ Beta\ (10,8)$	0.065	0.048	0.059	0.185
$Beta(1,1)\ vs\ Beta\ (10,9)$	0.058	0.051	0.073	0.164
$Beta(1,1)\ vs\ Beta\ (10,9.8)$	0.052	0.065	0.084	0.148
$Beta(1,1)\ vs\ Beta\ (10,10)$	0.048	0.070	0.093	0.138

Distributions set #2 consists of positively skewed bimodal beta cumulative distributions and a uniform [0,1] cumulative distribution (Figure 5-6). The positively skewed bimodal beta cumulative distributions are compared to the uniform cumulative distributions based on SSD and ESD. As shown in Table 5-3, all the positively skewed bimodal beta cumulative distributions are dominated by uniform [0,1] cumulative distribution by SSD. According to the ESD rule, the bimodal beta distributions are dominated by the uniform cumulative distribution except when no significance difference is distinguished. In that case two distributions are nondominated according to ESD. Thus, when the uniform distribution dominates the bimodal beta distributions based on SSD, then, the bimodal beta distributions do not dominate the uniform distribution based on ESD. Assuming that G is a positively skewed bimodal beta cumulative distribution and F is uniform [0,1] cumulative distribution, computational results for $\left|\int F \ln\left(\frac{F}{R}\right)\right|$, $\left|\int G \ln\left(\frac{G}{R}\right)\right|$, $\left|\int F \ln\left(\frac{F}{L}\right)\right|$, and $\left|\int G \ln\left(\frac{G}{L}\right)\right|$ which led to the decisions by ESD in Table 5-3 are shown in Table 5-4.

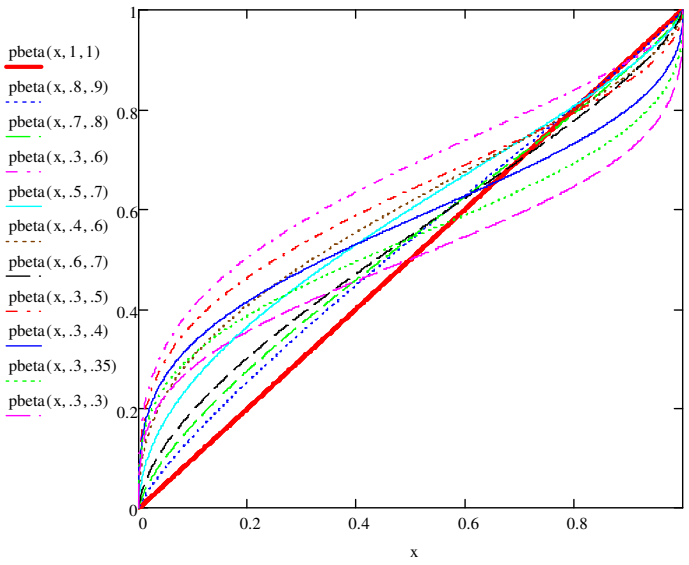


Figure 5-6. Distributions set #2: positively skewed bimodal beta and uniform [0,1] cumulative distributions

Table 5-3. Comparison between SSD and ESD rules on distributions set #2

	SSD	ESD
$Beta(1,1)\ vs\ Beta\ (0.8,0.9)$	$Beta(1,1)$ is dominant	$Beta(1,1)$ is dominant
$Beta(1,1)\ vs\ Beta\ (0.7,0.8)$	$Beta(1,1)$ is dominant	$Beta(1,1)$ is dominant
$Beta(1,1)\ vs\ Beta\ (0.3,0.6)$	$Beta(1,1)$ is dominant	$Beta(1,1)$ is dominant
$Beta(1,1)\ vs\ Beta\ (0.5,0.7)$	$Beta(1,1)$ is dominant	$Beta(1,1)$ is dominant
$Beta(1,1)\ vs\ Beta\ (0.4,0.6)$	$Beta(1,1)$ is dominant	$Beta(1,1)$ is dominant
$Beta(1,1)\ vs\ Beta\ (0.6,0.7)$	$Beta(1,1)$ is dominant	$Beta(1,1)$ is dominant
$Beta(1,1)\ vs\ Beta\ (0.3,0.5)$	$Beta(1,1)$ is dominant	$Beta(1,1)$ is dominant
$Beta(1,1)\ vs\ Beta\ (0.3,0.4)$	$Beta(1,1)$ is dominant	$Beta(1,1)$ is dominant
$Beta(1,1)\ vs\ Beta\ (0.3,0.35)$	$Beta(1,1)$ is dominant	$Beta(1,1)$ is dominant
$Beta(1,1)\ vs\ Beta\ (0.3,0.3)$	$Beta(1,1)$ is dominant	Nondominated set

Table 5-4. computational results for $\left|\int F \ln\left(\frac{F}{R}\right)\right|$, $\left|\int G \ln\left(\frac{G}{R}\right)\right|$, $\left|\int F \ln\left(\frac{F}{L}\right)\right|$, and $\left|\int G \ln\left(\frac{G}{L}\right)\right|$ of decisions made in distributions set #2 based on ESD, Assuming that G is a positively skewed bimodal beta cumulative distribution and F is uniform [0,1] cumulative distribution.

	$\left\|\int G \ln\left(\frac{G}{L}\right)\right\|$	$\left\|\int F \ln\left(\frac{F}{L}\right)\right\|$	$\left\|\int G \ln\left(\frac{G}{R}\right)\right\|$	$\left\|\int F \ln\left(\frac{F}{R}\right)\right\|$
$Beta(1,1)\ vs\ Beta\ (0.8,0.9)$	0.000	0.027	0.033	0.000
$Beta(1,1)\ vs\ Beta\ (0.7,0.8)$	0.003	0.031	0.043	0.003
$Beta(1,1)\ vs\ Beta\ (0.3,0.6)$	0.001	0.085	0.343	0.001
$Beta(1,1)\ vs\ Beta\ (0.5,0.7)$	0.002	0.060	0.123	0.002
$Beta(1,1)\ vs\ Beta\ (0.4,0.6)$	0.005	0.064	0.017	0.005
$Beta(1,1)\ vs\ Beta\ (0.6,0.7)$	0.007	0.036	0.059	0.007
$Beta(1,1)\ vs\ Beta\ (0.3,0.5)$	0.007	0.067	0.268	0.007
$Beta(1,1)\ vs\ Beta\ (0.3,0.4)$	0.022	0.048	0.192	0.024
$Beta(1,1)\ vs\ Beta\ (0.3,0.35)$	0.035	0.038	0.155	0.040
$Beta(1,1)\ vs\ Beta\ (0.3,0.3)$	0.053	0.029	0.119	0.063

Distributions set #3 consists of positively skewed unimodal beta cumulative distributions and a uniform [0,1] cumulative distribution (Figure 5-7). The positively skewed unimodal beta cumulative distributions are compared to the uniform cumulative distributions based on SSD and ESD. As shown in Table 5-5, when the positively skewed unimodal beta cumulative distributions are dominated by uniform [0,1] cumulative distribution by ESD, then the unimodal beta distributions does not dominate the uniform distribution based on SSD. Assuming that G is a positively skewed unimodal beta cumulative distribution and F is uniform [0,1] cumulative distribution, computational results for $\left|\int F \ln\left(\frac{F}{R}\right)\right|$, $\left|\int G \ln\left(\frac{G}{R}\right)\right|$, $\left|\int F \ln\left(\frac{F}{L}\right)\right|$, and $\left|\int G \ln\left(\frac{G}{L}\right)\right|$ which led to the decisions by ESD in Table 5-5 are shown in Table 5-6.

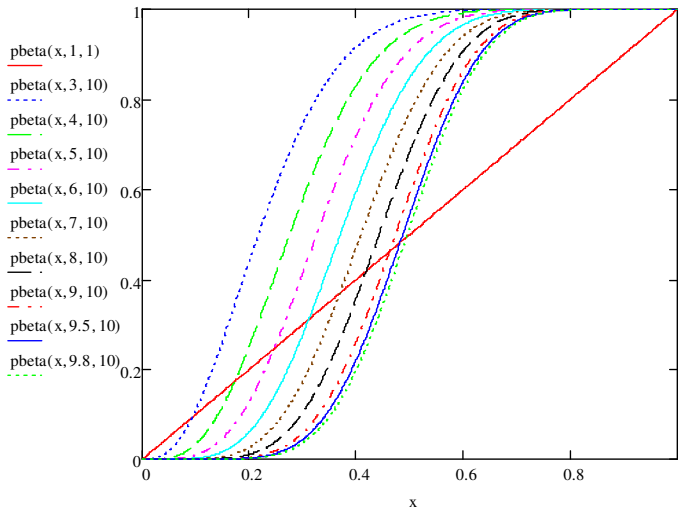


Figure 5-7. Distributions set #3: positively skewed unimodal beta and uniform [0,1] cumulative distributions

Table 5-5. Comparison between SSD and ESD rules on distributions set #3

	SSD	*ESD*
$Beta(1,1)\ vs\ Beta\ (3,10)$	Nondominated set	$Beta(1,1)$ is dominant
$Beta(1,1)\ vs\ Beta\ (4,10)$	Nondominated set	$Beta(1,1)$ is dominant
$Beta(1,1)\ vs\ Beta\ (5,10)$	Nondominated set	$Beta(1,1)$ is dominant
$Beta(1,1)\ vs\ Beta\ (6,10)$	Nondominated set	$Betal(1,1)$ is dominant
$Beta(1,1)\ vs\ Beta\ (7,10)$	Nondominated set	$Betal(1,1)$ is dominant
$Beta(1,1)\ vs\ Beta\ (8,10)$	Nondominated set	$Beta(1,1)$ is dominant
$Beta(1,1)\ vs\ Beta\ (9,10)$	Nondominated set	$Beta(1,1)$ is dominant
$Bleta(1,1)\ vs\ Beta\ (9.5,10)$	Nondominated set	$Beta(1,1)$ is dominant
$Belta(1,1)\ vs\ Beta\ (9.8,10)$	Nondominated set	Nondominated set

Table 5-6. computational results for $\left|\int F \ln\left(\frac{F}{R}\right)\right|$, $\left|\int G \ln\left(\frac{G}{R}\right)\right|$, $\left|\int F \ln\left(\frac{F}{L}\right)\right|$, and $\left|\int G \ln\left(\frac{G}{L}\right)\right|$ of decisions made in distributions set #3 based on ESD, Assuming that G is a positively skewed unimodal beta cumulative distribution and F is uniform [0,1] cumulative distribution.

	$\left\|\int G \ln\left(\frac{G}{L}\right)\right\|$	$\left\|\int F \ln\left(\frac{F}{L}\right)\right\|$	$\left\|\int G \ln\left(\frac{G}{R}\right)\right\|$	$\left\|\int F \ln\left(\frac{F}{R}\right)\right\|$
$Beta(1,1)\ vs\ Beta\ (3,10)$	0.001	0.175	0.398	0.003
$Beta(1,1)\ vs\ Beta\ (4,10)$	0.005	0.157	0.302	0.011
$Beta(1,1)\ vs\ Beta\ (5,10)$	0.010	0.138	0.236	0.025
$Beta(1,1)\ vs\ Beta\ (6,10)$	0.017	0.120	0.190	0.044
$Beta(1,1)\ vs\ Beta\ (7,10)$	0.025	0.105	0.155	0.066
$Beta(1,1)\ vs\ Beta\ (8,10)$	0.033	0.091	0.129	0.089
$Beta(1,1)\ vs\ Beta\ (9,10)$	0.040	0.098	0.124	0.113
$Beta(1,1)\ vs\ Beta\ (9.5,10)$	0.044	0.102	0.120	0.117
$Beta(1,1)\ vs\ Beta\ (9.8,10)$	0.047	0.072	0.096	0133

Relationship with p-AFSD

p-AFSD applies the proportion of the two bounded areas, created by the crossing of the two probability distributions, to the sum of the two areas. In the other words p, associated with p-AFSD is obtained by $\frac{\int_{F<G}(G-F)}{\int_{F<G}(G-F)+\int_{F>G}(F-G)}$. If p is larger than a predetermined probability, then F dominates G, otherwise they are nondominated. In case of ESD, a probability which is shown to be too close to the probability associated with p-ASFD can be found.

$$\left|\int F \ln\left(\frac{F}{R}\right)\right| = \int F \ln\left(\frac{F}{F \wedge G}\right) = \int_{F<G} F \ln\left(\frac{F}{F \wedge G}\right) + \int_{F>G} F \ln\left(\frac{F}{F \wedge G}\right) =$$

$$= \int_{F<G} F \ln\left(\frac{F}{F}\right) + \int_{F>G} F \ln\left(\frac{F}{G}\right) = \int_{F>G} F \ln\left(\frac{F}{G}\right)$$

$$= \int_{F>G} F \ln\left(\frac{F}{G}\right) + \int_{F<G} G \ln\left(\frac{G}{G}\right)$$

$$= \int_{F>G} (F \vee G) \ln\left(\frac{F \vee G}{G}\right) + \int_{F<G} (F \vee G) \ln\left(\frac{F \vee G}{G}\right)$$

$$= \int (F \vee G) \ln\left(\frac{F \vee G}{G}\right) = \left|\int L \ln\left(\frac{L}{G}\right)\right|.$$

Similarly it can be shown that $\left|\int F \ln\left(\frac{F}{L}\right)\right| = \left|\int R \ln\left(\frac{R}{G}\right)\right|$, $\left|\int R \ln\left(\frac{R}{F}\right)\right| = \left|\int G \ln\left(\frac{G}{L}\right)\right|$, and $\left|\int L \ln\left(\frac{L}{F}\right)\right| = \left|\int G \ln\left(\frac{G}{R}\right)\right|$.

$\left|\int G \ln\left(\frac{G}{R}\right)\right| + \left|\int R \ln\left(\frac{R}{G}\right)\right|$ is a measure for the mutual amount of lost information between G and R. Thus, by dividing $\left|\int G \ln\left(\frac{G}{R}\right)\right| + \left|\int R \ln\left(\frac{R}{G}\right)\right|$ over the summation of itself and $\left|\int F \ln\left(\frac{F}{R}\right)\right| + \left|\int L \ln\left(\frac{L}{F}\right)\right|$ we can obtain a measure which is comparable with probability that obtained from p-AFSD. That is,

$\frac{\left|\int G \ln\left(\frac{G}{R}\right)\right|+\left|\int R \ln\left(\frac{R}{G}\right)\right|}{\left|\int F \ln\left(\frac{F}{R}\right)\right|+\left|\int R \ln\left(\frac{R}{F}\right)\right|+\left|\int G \ln\left(\frac{G}{R}\right)\right|+\left|\int R \ln\left(\frac{R}{G}\right)\right|}$ measures the percentage that F dominates G which can also be rewritten by $\frac{\left|\int G \ln\left(\frac{G}{R}\right)\right|+\left|\int F \ln\left(\frac{F}{L}\right)\right|}{\left|\int F \ln\left(\frac{F}{R}\right)\right|+\left|\int G \ln\left(\frac{G}{R}\right)\right|+\left|\int G \ln\left(\frac{G}{R}\right)\right|+\left|\int F \ln\left(\frac{F}{L}\right)\right|}$.

For three distribution sets, probabilities associated with p-AFSD and ESD are compared with each other. The results are illustrated in Tables Table 5-7, Table 5-8, and Table 5-9.

Table 5-7. Comparison between probabilities associated with p-AFSD and ESD for distributions set #1

	Probability Associated With p-AFSD	Probability Associated With ESD
$Beta(1,1)\ vs\ Beta\ (10,3)$	p=99.33%	p=99.24%
$Beta(1,1)\ vs\ Beta\ (10,4)$	p=96.82%	p=96.46%
$Beta(1,1)\ vs\ Beta\ (10,5)$	p=91.90%	p=91.32%
$Beta(1,1)\ vs\ Beta\ (10,6)$	p=84.78%	p=83.77%
$Beta(1,1)\ vs\ Beta\ (10,7)$	p=76.21%	p=75.23%
$Beta(1,1)\ vs\ Beta\ (10,8)$	p=67.08%	p=66.66%
$Beta(1,1)\ vs\ Beta\ (10,9)$	p=53.98%	p=54.08%
$Beta(1,1)\ vs\ Beta\ (10,9.8)$	p=51.56%	p=52.11%
$Beta(1,1)\ vs\ Beta\ (10,10)$	p=50.00%	p=50.20%

Table 5-8. Comparison between probabilities associated with p-AFSD and ESD for distributions set #2

	Probability Associated With p-AFSD	Probability Associated With ESD
$Beta(1,1)\ vs\ Beta\ (0.8,0.9)$	p=98.41%	p=98.40%
$Beta(1,1)\ vs\ Beta\ (0.7,0.8)$	p=92.12%	p=92.05%
$Beta(1,1)\ vs\ Beta\ (0.3,0.6)$	p=99.18%	p=98.99%
$Beta(1,1)\ vs\ Beta\ (0.5,0.7)$	p=97.24%	p=97.09%
$Beta(1,1)\ vs\ Beta\ (0.4,0.6)$	p=95.68%	p=95.26%
$Beta(1,1)\ vs\ Beta\ (0.6,0.7)$	p=86.46%	p=86.36%
$Beta(1,1)\ vs\ Beta\ (0.3,0.5)$	p=94.95%	p=94.05%
$Beta(1,1)\ vs\ Beta\ (0.3,0.4)$	p=80.49%	p=78.73%
$Beta(1,1)\ vs\ Beta\ (0.3,0.35)$	p=67.02%	p=65.77%
$Bleta(1,1)\ vs\ Beta\ (0.3,0.3)$	p=50.00%	p=50.38%

Table 5-9. Comparison between probabilities associated with p-AFSD and ESD for distributions set #3

	Probability Associated With p-AFSD	Probability Associated With ESD
$Beta(1,1)\ vs\ Beta\ (3,10)$	p=99.33%	p=99.30%
$Beta(1,1)\ vs\ Beta\ (4,10)$	p=96.82%	p=96.75%
$Beta(1,1)\ vs\ Bet\ (5,10)$	p=91.90%	p=91.83%
$Beta(1,1)\ vs\ Beta\ (6,10)$	p=84.78%	p=84.39%
$Beta(1,1)\ vs\ Beta\ (7,10)$	p=76.21%	p=75.45%
$Beta(1,1)\ vs\ Beta\ (8,10)$	p=67.08%	p=66.28%
$Beta(1,1)\ vs\ Beta\ (9,10)$	p=60.95%	p=60.67%
$Beta(1,1)\ vs\ Beta\ (9.5,10)$	p=53.98%	p=54.07%
$Beta(1,1)\ vs\ Beta\ (9.8,10)$	p=51.56%	p=51.87%

5-4. Conclusion

Using Kullback-Leibler information between two cumulative distributions, we present an entropy-based criterion for stochastic dominance, ESD. Since our criterion is weaker than FSD and SSD, its advantage is to reduce the possibility of obtaining a nondominated set of alternatives. In addition, under the new stochastic dominance criterion, it is not required to calculate the surrendered area between two distributions as it is in p-AFSD.

We investigate the relation between ESD and p-AFSD, SSD, and FSD. We show that if a distribution is dominant based on ESD, then that distribution would not be dominated based on SSD. Also we show that the probability associated with our new measure is too close to the probability obtained from p-AFSD.

Appendix I

A simple numerical example

Suppose a discrete part manufacturing system produces two parts, part I and part II using four work centers**.** Parts enter from work center 1 with the mean interarrival time of 1.25 minutes and exit from work center 4. The pickup and drop-off points are assumed to be at the same location for each work center. Table AI.1 shows the operation sequence, production volume percentage of each part type, and the processing time of each part type in each work center. The vehicle speed is 100 feet per minute and the distance between work centers is shown in Table AI.2. Loading and unloading operations are assumed to be negligible.

Table AI-1. Operation sequences, processing times and production volume percentage of part types

Part Type	Operation Sequence (processing time)	Production Volume
I	1(1 minute)→2(2 minutes)→4(1 minute)	%55
II	1(1 minute)→3(2 minutes)→4(2 minutes)	%45

Table AI-2. The distance between work centers (feet)

	Work Center 1	Work Center 2	Work Center 3	Work Center 4
Work Center 1	-	150	300	450
Work Center 2	150	-	150	300
Work Center 3	300	150	-	150
Work Center 4	450	300	150	-

Parts are being moved by a single AGV. Suppose that the vehicle is idle at work center 2 at time t_0 and at this moment parts are waiting in the input queues and output queues of each work center as illustrated in Figure AI.1.

Input queue		Output queue
$P_I \rightarrow P_{II} \rightarrow P_I \rightarrow P_{II} \rightarrow P_{II}$	Work Center 1	P_{II}
P_I	Work Center 2	P_I
$P_{II} \rightarrow P_{II}$	Work Center 3	EMPTY
P_{II}	Work Center 4	P_I, P_{II}, P_{II}

Figure AI-1. Input and output queues of work centers at time t_0
P_I : part I; P_{II} : part II

According to the Figure AI.1, two move requests are available for the vehicle at time t_0:

1) Pickup one unit of part I from work center 2 and drop it off at work center 4 at time $t_0 + 3$.

   After the completion of this move request the input and output queues of work centers are estimated to be as Figure AI.2 :

Input queue		Output queue
$P_I \rightarrow P_{II} \rightarrow P_I \rightarrow P_{II}$	Work Center 1	$P_I, P_{II}, P_{II}, P_{II}$
EMPTY	Work Center 2	P_I
P_{II}	Work Center 3	P_{II}
P_I	Work Center 4	$P_I, P_{II}, P_{II}, P_{II}$

Figure AI-2. Estimated input and output queues of work centers at time $t_0 + 3$
P_I : part I; P_{II} : part II

2) Pickup one unit of part II from work center 1 and drop it off at work center 3 at time $t_0 + 4.5$. After the completion of this move request the input and output queues of work centers are estimated to be as Figure AI.3:

Input queue		Output queue
$P_I \rightarrow P_I \rightarrow P_{II} \rightarrow P_{II}$	Work Center 1	$P_I, P_{II}, P_{II}, P_{II}$
EMPTY	Work Center 2	P_I, P_I
P_{II}	Work Center 3	P_{II}, P_{II}
EMPTY	Work Center 4	$P_I, P_{II}, P_{II}, P_{II}$

Figure AI-3. Estimated input and output queues of work centers at time $t_0 + 4.5$

P_I : part I; P_{II} : part II

The kullback-leibler divergence measure is calculated based on each of the proposed algorithms and the results are shown in Table AI.3:

Table AI-3. Kullback-Leibler Divergence measure calculation for each proposed algorithm

	Kullback-Leibler Divergence measure	
	First move request	Second move request
WCBA	0.9557	0.6365
OQBA	1.3768	1.3297
IOQBA	1.7329	1.6915

Since all three algorithms select the move request with the largest kullback-leibler divergence measure, the first move request is chosen based WCBA, OQBA, and IOQBA.

Appendix II

Waiting time data for different configuration

Table AII-1. Average waiting time (in minutes) for different configurations of model 1

(U: AGV capacity ; λ: Mean interarrival time)

	WCBA	OQBA	IOQBA	STD	MOQS	MFCFS	WCBA	OQBA	IOQBA	STD	MOQS	MFCFS
(U, λ)	Number of AGVs = 4						Number of AGVs = 5					
(5,4.6)	46.23	48.58	47.21	63.51	58.65	604.69	29.50	30.33	29.15	31.94	40.31	508.80
(5,4.8)	41.24	43.83	42.76	50.75	56.33	684.78	28.45	29.20	29.02	31.28	37.26	434.25
(5,5.0)	38.74	39.29	39.18	47.40	53.46	535.61	26.98	28.38	28.97	29.66	35.70	398.84
(5,5.2)	34.54	37.63	36.28	43.45	51.32	536.74	26.38	26.32	26.93	28.14	33.76	384.64
(5,5.4)	33.06	36.67	34.57	39.95	48.08	512.74	25.78	26.19	26.36	25.68	31.42	367.23
(5,5.6)	31.93	32.70	33.52	38.62	46.15	459.26	25.72	25.99	26.51	24.84	32.02	372.50
(5,5.8)	25.14	30.51	28.44	35.39	44.36	474.21	25.70	26.45	25.49	24.13	29.67	365.43
(5,6.0)	25.01	28.15	26.13	32.43	41.03	444.98	24.57	24.61	25.03	22.49	26.74	333.71
(20,4.6)	42.27	46.65	45.60	55.47	58.95	470.24	30.01	30.26	29.69	33.26	41.08	419.63
(20,4.8)	39.07	42.58	41.74	49.03	57.10	400.92	28.53	28.73	28.71	30.45	38.85	361.73
(20,5.0)	36.09	38.55	39.00	45.41	52.83	399.41	27.72	27.60	28.01	29.12	37.43	371.66
(20,5.2)	34.21	35.90	35.65	41.96	51.30	392.21	26.58	27.21	26.78	28.45	34.52	332.93
(20,5.4)	32.69	34.30	34.03	39.75	49.62	394.98	26.13	26.26	26.84	27.31	33.11	318.46
(20,5.6)	30.72	31.93	32.63	36.07	45.82	365.04	25.67	25.75	25.66	24.51	31.52	290.62
(20,5.8)	29.79	31.82	30.97	34.72	43.61	369.24	25.66	25.16	25.14	23.97	28.35	273.36
(20,6.0)	28.62	29.73	30.72	33.46	42.01	344.35	24.61	25.00	25.09	22.88	28.06	278.25
(10000,4.6)	41.40	44.84	45.34	53.64	58.84	449.84	29.92	30.47	30.80	31.04	41.46	373.77
(10000,4.8)	41.98	43.27	40.72	50.03	56.98	466.85	28.00	29.07	29.29	30.95	38.19	373.89
(10000,5.0)	35.48	39.80	39.02	41.82	54.05	406.80	27.26	27.52	27.13	27.12	38.77	320.68
(10000,5.2)	36.26	34.47	35.67	39.52	51.83	416.55	27.78	27.02	26.81	27.45	34.74	296.80
(10000,5.4)	33.11	35.15	34.63	38.61	49.19	404.32	25.94	26.23	26.78	25.54	31.95	303.07
(10000,5.6)	31.98	32.66	32.44	35.06	45.55	364.17	25.75	26.08	25.87	25.75	30.54	290.93
(10000,5.8)	30.31	30.77	31.31	34.83	43.89	351.75	25.05	25.10	25.10	23.80	29.08	290.87
(10000,6.0)	29.26	29.54	28.95	33.17	42.68	339.13	24.10	24.80	24.84	23.32	27.31	290.17

Table AII-2. Average waiting time (in minutes) for different configurations of model 2

(U: AGV capacity ; λ: Mean interarrival time)

	WCBA	OQBA	IOQBA	STD	MOQS	MFCFS	WCBA	OQBA	IOQBA	STD	MOQS	MFCFS
(U, λ)	Number of AGVs=3						Number of AGVs=5					
(5,4.0)	35.65	40.06	46.61	51.62	60.60	268.31	135.02	543.10	1126.02	486.86	116.50	414.62
(5,4.4)	30.69	33.30	34.56	43.87	53.84	257.35	111.97	111.97	435.67	297.54	107.00	413.79
(5,4.8)	28.59	29.28	29.77	40.06	48.64	253.99	98.37	98.37	253.82	133.94	96.77	408.69
(5,5.2)	26.80	26.83	26.91	35.80	44.40	254.73	82.18	86.68	157.22	101.65	91.94	413.70
(5,5.6)	24.96	25.09	25.27	34.00	40.12	245.63	76.52	81.06	147.63	96.55	83.06	398.91
(5,6.0)	24.40	24.99	24.90	32.31	36.27	238.89	74.81	80.72	145.48	91.74	75.11	387.97
(20,4.0)	34.56	39.89	1287.62	49.09	59.73	261.95	134.08	998.10	772.68	221.16	114.78	422.93
(20,4.4)	31.54	33.75	33.97	44.67	52.39	265.77	112.75	510.81	305.61	180.92	106.49	407.31
(20,4.8)	27.78	28.97	29.05	40.74	47.47	257.36	100.02	230.56	233.07	129.85	102.53	415.54
(20,5.2)	26.30	27.20	27.80	37.44	43.87	260.00	81.43	123.44	129.85	106.76	95.76	410.66
(20,5.6)	25.60	25.87	26.14	34.10	40.57	255.66	69.45	92.83	96.39	84.97	87.26	409.61
(20,6.0)	24.30	24.80	24.79	32.23	36.98	255.81	55.61	68.51	94.76	74.86	82.58	398.14
(10000,4.0)	39.42	44.37	35.83	49.95	60.30	261.40	144.57	326.74	325.74	851.32	117.05	495.90
(10000,4.4)	33.91	32.92	31.39	43.78	54.84	257.35	124.04	261.33	262.33	635.23	110.45	475.32
(10000,4.8)	30.11	29.99	27.82	38.62	49.29	260.02	108.27	181.17	180.07	464.52	98.91	453.62
(10000,5.2)	28.19	28.27	26.83	37.58	43.19	260.16	85.54	107.78	108.76	263.04	94.58	406.70
(10000,5.6)	25.92	26.90	24.88	34.47	40.68	260.62	67.84	98.34	97.40	135.00	90.67	430.07
(10000,6.0)	24.26	25.47	24.18	32.99	37.09	254.01	55.05	74.34	75.44	77.37	84.39	409.70

Appendix III

Proof of theorems 4.1-4.3

Proof of Theorem 4.1.

By the central limit theorem it can be shown that $\sqrt{n}(\bar{a}-\mu)\sim N(0,\frac{\sum_{i=1}^{n}(a_i-\bar{a})^2}{n-1})$. On the other hand $p_i^{*\mu}=e^{-\lambda_0-\lambda_1 x_i-1}-1$. This function is continuous and has first partial derivative with respect to the component μ at $\mu=\bar{a}$. Hence, according the multivariate delta method:

$\sqrt{n}(P^{*\bar{a}}-P^{*\mu})\sim MVN(0,A^T\frac{\sum_{i=1}^{n}(a_i-\bar{a})^2}{n-1}A)$ where $A_i=\left.\frac{\partial p_i^{*\mu}}{\partial \mu}\right|_{\mu=\bar{a}}$

$$\frac{\partial p_i^{*\mu}}{\partial \mu}=\frac{\partial p_i^{*\mu}}{\partial \lambda_0}\frac{\partial \lambda_0}{\partial \mu}+\frac{\partial p_i^{*\mu}}{\partial \lambda_1}\frac{\partial \lambda_1}{\partial \mu} \qquad \text{AIII-1}$$

By replacing $p_i^{*\mu}=e^{-\lambda_0-\lambda_1 x_i-1}-1$ in the first and second constraints of model (I) we will have:

$$\sum_{i=1}^{B}\left(\mathrm{e}^{-\lambda_0-\lambda_1 x_i-1}-1\right)=1\Longrightarrow(B+1)e^{\lambda_0+1}=\sum_{i=1}^{B}\mathrm{e}^{-\lambda_1 x_i} \qquad \text{AIII-2}$$

$$\sum_{i=1}^{B}x_i\left(\mathrm{e}^{-\lambda_0-\lambda_1 x_i-1}-1\right)=\mu\ \Longrightarrow e^{\lambda_0+1}=\frac{\sum_{i=1}^{B}x_i\mathrm{e}^{-\lambda_1 x_i}}{\mu+\sum_{i=1}^{B}x_i} \qquad \text{AIII-3}$$

By combining equations (AIII-2) and (AIII-3) we will get:

$$\mu=\frac{(B+1)\sum_{i=1}^{B}x_i e^{-\lambda_1 x_i}}{\sum_{i=1}^{B}e^{-\lambda_1 x_i}}-\sum_{i=1}^{B}x_i \qquad \text{AIII-4}$$

- Calculating $\frac{\partial \lambda_1}{\partial \mu}$: from equation (AIII-4)

$$\frac{\partial \mu}{\partial \lambda_1} = \frac{-(B+1)\left(\sum_{i=1}^{B} x_i{}^2 e^{-\lambda_1 x_i}\right)\left(\sum_{i=1}^{B} e^{-\lambda_1 x_i}\right) + (B+1)\left(\sum_{i=1}^{B} x_i e^{-\lambda_1 x_i}\right)^2}{\left(\sum_{i=1}^{B} e^{-\lambda_1 x_i}\right)^2} \quad \text{AIII-5}$$

Hence:

$$\frac{\partial \lambda_1}{\partial \mu} = \frac{\left(\sum_{i=1}^{B} e^{-\lambda_1 x_i}\right)^2}{-(B+1)\left(\sum_{i=1}^{B} x_i{}^2 e^{-\lambda_1 x_i}\right)\left(\sum_{i=1}^{B} e^{-\lambda_1 x_i}\right) + (B+1)\left(\sum_{i=1}^{B} x_i e^{-\lambda_1 x_i}\right)^2} \quad \text{AIII-6}$$

- Calculating $\frac{\partial \lambda_0}{\partial \mu}$: from equation (AIII-2)

$$(B+1)e^{\lambda_0}\frac{\partial \lambda_0}{\partial \lambda_1} = -\sum_{i=1}^{B} x_i \mathrm{e}^{-\lambda_1 x_i - 1} \implies \frac{\partial \lambda_0}{\partial \lambda_1} = -\frac{\sum_{i=1}^{B} x_i \mathrm{e}^{-\lambda_1 x_i - \lambda_0 - 1}}{B+1}$$

$$\Longrightarrow \frac{\partial \lambda_0}{\partial \lambda_1} = \frac{-\bar{a} - \sum_{i=1}^{B} x_i}{B+1} \quad \text{AIII-7}$$

$$\frac{\partial \lambda_0}{\partial \mu} = \frac{\partial \lambda_0}{\partial \lambda_1}\frac{\partial \lambda_1}{\partial \mu}$$

$$= -\frac{\bar{a} + \sum_{i=1}^{B} x_i}{(B+1)^2} * \frac{\left(\sum_{i=1}^{B} e^{-\lambda_1 x_i}\right)^2}{\left(-\sum_{i=1}^{B} x_i{}^2 e^{-\lambda_1 x_i}\right)\left(\sum_{i=1}^{B} e^{-\lambda_1 x_i}\right) + \left(\sum_{i=1}^{B} x_i e^{-\lambda_1 x_i}\right)^2} \quad \text{AIII-8}$$

- Calculating $\frac{\partial p^{*\mu}}{\partial \lambda_1}$ and $\frac{\partial p_i^{*\mu}}{\partial \lambda_0}$: Since $p_i^{*\mu} = e^{-\lambda_0 - \lambda_1 x_i - 1} - 1$, we have:

$$\frac{\partial p_i^{*\mu}}{\partial \lambda_1} = -x_i e^{-\lambda_0 - \lambda_1 x_i - 1} \quad \text{AIII-9}$$

$$\frac{\partial p_i^{*\mu}}{\partial \lambda_0} = -e^{-\lambda_0 - \lambda_1 x_i - 1} \quad \text{AIII-10}$$

Therefore by replacing $\frac{\partial \lambda_1}{\partial \mu}$, $\frac{\partial \lambda_0}{\partial \mu}$, $\frac{\partial p_i^{*\mu}}{\partial \lambda_1}$ and $\frac{\partial p_i^{*\mu}}{\partial \lambda_0}$ equations (AIII-6), (AIII-8), (AIII-9) and (AIII-10) into equation (AIII-1), we obtain:

$$A_i = \left(\frac{(B+1)^{-1}\left(\sum_{i=1}^{B} e^{-\lambda_1 x_i}\right)^2}{(-\sum_{i=1}^{B} {x_i}^2 e^{-\lambda_1 x_i})(\sum_{i=1}^{B} e^{-\lambda_1 x_i}) + (\sum_{i=1}^{B} x_i e^{-\lambda_1 x_i})^2} \right).$$

$$e^{-\lambda_0 - \lambda_1 x_i - 1} \left(\frac{\bar{a} + \sum_{i=1}^{B} x_i}{B+1} - x_i \right) \qquad \text{AIII-11}$$

■

Proof of Theorem 4.2.

Since $\sqrt{n}(P^{*\bar{a}} - P^{*\mu}) \xrightarrow{\mathcal{L}} MVN(0, \Sigma)$, Using Taylor series expansion we obtain:

$U(P^{*\bar{a}}) = U(P^{*\mu}) + h^T(P^{*\bar{a}} - P^{*\mu}) + R$ Where $h = (h_1, h_2, \dots, h_B)$ are the values given above and $R \longrightarrow 0 \; as \; n \longrightarrow \infty$. Thus it follows that $\sqrt{n}\big(U(P^{*\bar{a}}) - U(P^{*\mu})\big)$ $\xrightarrow{\mathcal{L}} N(0, \sigma^2)$ and estimated divergence variance, $\sigma^2 = h\Sigma h'$. Hence $U(P^{*\bar{a}}) \pm z_{\alpha/2}\sigma/\sqrt{n}$ is an approximate $100(1-\alpha)\%$ confidence interval for $U(P^{*\mu})$ where $z_{\alpha/2}$ is the upper $\alpha/2$ percentage point from the normal distribution.

■

Proof of Theorem 4.3.

Assume that $P^{**\prime} = (p_1^{**\prime}, p_2^{**\prime}, \dots, p_B^{**\prime})$ is the optimal solution of Model (I) if the unknown mean was $\mu' = \mu + \Delta\mu$. Define $\Delta KL = KL(P^{**\prime}) - KL(P^{**})$, $\Delta H = H(P^{**\prime}) - H(P^{*})$ and $\Delta P = P^{**\prime} - P^{**} = (\Delta p_1, \dots, \Delta p_B)$ where:

$$KL(P) = \sum_{i=1}^{B-\left\lfloor \frac{B}{K} \right\rfloor} (1+p_i)\ln(1+p_i) + \sum_{i=B-\left\lfloor \frac{B}{K} \right\rfloor+1}^{B} (1+p_i)\ln\left(\frac{1+p_i}{1+\frac{1}{\left\lfloor \frac{B}{K} \right\rfloor}}\right) \qquad \text{AIII-12}$$

We have:

$$\Delta KL = \left[\sum_{i=1}^{B-\left\lfloor \frac{B}{K} \right\rfloor} [1+p_i^{**\prime}]\ln[1+p_i^{**\prime}] + \sum_{i=B-\left\lfloor \frac{B}{K} \right\rfloor+1}^{B} [1+p_i^{**\prime}]\ln\left[\frac{1+p_i^{**\prime}}{1+\frac{1}{\left\lfloor \frac{B}{K} \right\rfloor}}\right] - \right.$$

$$\left. \sum_{i=1}^{B-\left\lfloor \frac{B}{K} \right\rfloor} (1+p_i^{**})\ln(1+p_i^{**}) + \sum_{i=B-\left\lfloor \frac{B}{K} \right\rfloor+1}^{B} (1+p_i^{**})\ln\left(\frac{1+p_i^{**}}{1+\frac{1}{\left\lfloor \frac{B}{K} \right\rfloor}}\right) \right]$$

$$= \left[\sum_{i=1}^{B-\left\lfloor \frac{B}{K} \right\rfloor} (1+p_i^{**\prime})\ln(1+p_i^{**\prime}) + \sum_{i=B-\left\lfloor \frac{B}{K} \right\rfloor+1}^{B} (1+p_i^{**\prime})\ln(1+p_i^{**\prime}) \right] -$$

$$\left[\sum_{i=1}^{B-\left\lfloor \frac{B}{K} \right\rfloor} (1+p_i^{**})\ln(1+p_i^{**}) + \sum_{i=B-\left\lfloor \frac{B}{K} \right\rfloor+1}^{B} (1+p_i^{**})\ln(1+p_i^{**}) \right] -$$

$$\left[\sum_{i=B-\left\lfloor \frac{B}{K} \right\rfloor+1}^{B} (1+p_i^{**\prime})\ln\left(1+\frac{1}{\left\lfloor \frac{B}{K} \right\rfloor}\right) - \sum_{i=B-\left\lfloor \frac{B}{K} \right\rfloor+1}^{B} (1+p_i^{**})\ln\left(1+\frac{1}{\left\lfloor \frac{B}{K} \right\rfloor}\right) \right]$$

$$= H(P^{**}) - H(P^{**\prime}) - \sum_{i=B-\left\lfloor\frac{B}{K}\right\rfloor+1}^{B} (p_i^{**\prime} - p_i^{**}) \ln\left(1 + \frac{1}{\left\lfloor\frac{B}{K}\right\rfloor}\right)$$

$$= -\Delta H - \sum_{i=B-\left\lfloor\frac{B}{K}\right\rfloor+1}^{B} \Delta p_i^{**} \ln\left(1 + \frac{1}{\left\lfloor\frac{B}{K}\right\rfloor}\right)$$

$$\Longrightarrow \frac{\Delta U}{\Delta \mu} = -\frac{\Delta H}{\Delta \mu} - \sum_{i=B-\left\lfloor\frac{B}{K}\right\rfloor+1}^{B} \frac{p_i^{**}}{\Delta \mu} \ln\left(1 + \frac{1}{\left\lfloor\frac{B}{K}\right\rfloor}\right)$$

$$= \lambda_1 + \sum_{i=B-\left\lfloor\frac{B}{K}\right\rfloor+1}^{B} \frac{1}{x_i} \ln\left(1 + \frac{1}{\left\lfloor\frac{B}{K}\right\rfloor}\right) \qquad \text{AIII-13}$$

Now suppose that $\Delta\mu > \delta$; we are going to find δ_I such that: $P(\Delta KL > \delta_I | \Delta\mu > \delta) \geq 1 - \alpha$.

Replacing ΔU from (AIII-13) we will get:

$$P(\Delta KL > \delta_I | \Delta\mu > \delta) = P\left(\Delta\mu\left[\lambda_1 + \sum_{i=B-\left\lfloor\frac{B}{K}\right\rfloor+1}^{B} \frac{1}{x_i} \ln\left(1 + \frac{1}{\left\lfloor\frac{B}{K}\right\rfloor}\right)\right] > \delta_I | \Delta\mu > \delta\right)$$

$$= P\left(\lambda_1 + \sum_{i=B-\left\lfloor\frac{B}{K}\right\rfloor+1}^{B} \frac{1}{x_i} \ln\left(1 + \frac{1}{\left\lfloor\frac{B}{K}\right\rfloor}\right) > \frac{\delta_I}{\Delta\mu} | \Delta\mu > \delta\right)$$

$$\geq P\left(\lambda_1 + \sum_{i=B-\left\lfloor\frac{B}{K}\right\rfloor+1}^{B} \frac{1}{x_i} \ln\left(1 + \frac{1}{\left\lfloor\frac{B}{K}\right\rfloor}\right) > \frac{\delta_I}{\delta}\right)$$

$$\geq P\left(\lambda_1{}^{*\mu} + \frac{\left\lfloor\frac{B}{K}\right\rfloor}{x_{[B]}}\ln(1+\frac{1}{\left\lfloor\frac{B}{K}\right\rfloor}) > \frac{\delta_I}{\delta}\right)$$

$$= P\left(\frac{\left\lfloor\frac{B}{K}\right\rfloor}{x_{[B]}}\ln(1+\frac{1}{\left\lfloor\frac{B}{K}\right\rfloor}) > \frac{\delta_I}{\delta} - \lambda_1\right)$$

$$= P\left(Y > \frac{\delta_I}{\delta} - \lambda_1\right) \text{ where } Y = \frac{\left\lfloor\frac{B}{K}\right\rfloor}{X_{[B]}}\ln(1+\frac{1}{\left\lfloor\frac{B}{K}\right\rfloor}) \qquad \text{AIII-14}$$

The cumulative probability distribution of x_i is $F_X(t) = \sum_{i=1}^{B}\left(\mathrm{e}^{-\lambda_0-\lambda_1 x_i - 1} - 1\right)$; So

$$F_{X_{[B]}}(t) = \left(F_X(t)\right)^B = \left(\sum_{i=1}^{B}\left(\mathrm{e}^{-\lambda_0-\lambda_1 x_i - 1} - 1\right)\right)^B \qquad \text{AIII-15}$$

On the other hand since $Y = \frac{\left\lfloor\frac{B}{K}\right\rfloor}{X_{[B]}}\ln(1+\frac{1}{\left\lfloor\frac{B}{K}\right\rfloor})$:

$$F_Y(y) = 1 - F_{X_{[B]}}\left(\frac{\left\lfloor\frac{B}{K}\right\rfloor}{y}\ln(1+\frac{1}{\left\lfloor\frac{B}{K}\right\rfloor})\right)$$

$$= 1 - \left(\sum_{i=1}^{\frac{\left\lfloor\frac{B}{K}\right\rfloor}{y}\ln\left[1+\frac{1}{\left\lfloor\frac{B}{K}\right\rfloor}\right]}\left(\mathrm{e}^{-\lambda_0-\lambda_1 x_i - 1} - 1\right)\right)^B \qquad \text{AIII-16}$$

From equations (AIII-14) and (AIV-16) we have:

$$1 - F_Y\left(\frac{\delta_I}{\delta} - \lambda_1\right) = \left(\sum_{i=1}^{\frac{\lfloor \frac{B}{K} \rfloor}{\frac{\delta_I}{\delta} - \lambda_1} \ln\left[1 + \frac{1}{\lfloor \frac{B}{K} \rfloor}\right]} \left(\mathrm{e}^{-\lambda_0 - \lambda_1 x_i - 1} - 1\right) \right)^B \geq 1 - \alpha$$

$$\Longrightarrow \frac{\delta_I}{\delta} - \lambda_1 \geq {F_Y}^{-1}(1 - \alpha) \qquad \text{AIII-17}$$

So we will choose δ_I such that it satisfies (AIV-17).

References

Abbas., A.E. (2006). Entropy. Wiley Encyclopedia of Biomedical Engineering, John Wiley & Sons, Inc.

Ahmadi-Javid, A. (2012). Entropic Value-at-Risk: A New Coherent Risk Measure, Journal Optimization Theory and Applications,155: 1105–1123.

Asadi, M., Ebrahimi, N., Hamedani, G.G., Soofi, E.S. (2004). Maximum dynamic entropy models. Journal of Applied Probability, 41: 379–390.

Aczel, J.D., Daroczy, Z. (1975), On Measures of Information and Their Generalizations. Academic Press, New York.

Bali, T.G., Brown, S.J., Demirtas, K.O. (2011). Investing in stock market anomalies. Working paper. Georgetown University, Washington, DC.

Bali, T.G., Demirtas, K.O., Levy, H., Wolf, A. (2009). Bonds versus stocks: Investors' age and risk taking. Journal of Monetary Economics, 56(6): 817–830.

Barapour, S., Rad, A.H. (2012). Testing goodness-of-fit for exponential distribution based on cumulative residual entropy. Communications in Statistics: Theory and Methods, 41: 1387-1396.

Bartholdi, J.J., Platzman, L.K. (1989). Decentralized control of automated guided vehicles on a simple loop. IIE Transactions, 21(1), 76-81.

Bechhofer, R.E., Santner, T. J. (1995). Goldsman, D. M., Design and Analysis of experiments for Statistical Selection. Screening, and Multiple Comparisions, John Wiley & Sons, Inc., NewYork.

Becker, C., Scholl, A. (2006). A survey on problems and methods in generalized assembly line balancing. European Journal of Operational Research, 168: 694-715.

Bera, A.K., Park, S.Y. (2005). Optimal portfolio diversification using the maximum entropy principle. In: Proceedings of the Second Conference on Recent Developments in the Theory, Method, and Applications of Information and Entropy Econometrics, The American University Washington, DC, USA.

Bera, A.K., Park, S.Y. (2008). Optimal portfolio diversification using the maximum entropy principle. Econometric Review, 27: 484–512.

Bhattacharyya, R., Ahmed Hossain, S., Kar, S. (2013). Fuzzy cross-entropy, mean, variance, skewness models for portfolio selection. Journal of King Saud University-Computer and Information Sciences, Article in press.

Boltzmann, L. (1871a). Analytischer Beweis des zweiten Haubtsatzes der mechanischen Wärmetheorie aus den Sätzen über das Gleichgewicht der lebendigen Kraft. Wiener Berichte, 63, 712–732. In Boltzmann (1909), Volume 1, Paper 20. Cited in Uffink (2004)

Boltzmann, L. (1871b). Einige allgemeine Sätze über wärmegleichgewicht. Wiener Berichte, 63, 679–711. In Boltzmann (1909), Volume 1, Paper 19. Cited in Uffink (2004)

Boltzmann, L. (1871c). Über das Wärmegleichgewicht zwischen mehratomigen Gasmolekülen. Wiener Berichte, 63, 397–418. In Boltzmann (1909), Volume 1, Paper 18. Cited in Uffink (2004)

Bozer, Y.A., Yen, C.K. (1996). Intelligent dispatching rules for trip-based material handling systems. Journal of Manufacturing Systems, 15(4): 226-239.

Calinescu, A., Efstathiou, J., Schirn, J., Bermejo, J. (1998). Applying and assessing two methods for measuring complexity in manufacturing. Journal of the Operational Research Society, 49: 723-733.

Chang, A.Y., Whitehouse, D.J., Chang, S.L., Hsieh, Y.C. (2001). An approach to the measurement of single-machine flexibility. International Journal of Production Research, 39 (8): 1589-1601.

Chang, A.Y. (2004). On the measurement of labor flexibility. Proceeding of the IEEE International Engineering Management Conference IEMC-2004, October, Singapore: 18-22.

Chang, A.Y. (2007). On the measurement of routing flexibility: A multiple attribute approach. International Journal of Production Economics, 109: 122-136.

Chang, A.Y. (2009). An attribute approach to the measurement of machine-group flexibility. European Journal of Operational Research, 194: 774-786.

Chen, H.C., Chen. C.H., Yucesan, E. (2000). Computing efforts allocation for ordinal optimization and discrete event simulation. IEEE Transactions on Automatic Control, 45: 960-964.

Chengli, Z., Yan, Y. (2012). Coherent Risk Measure Based on Relative Entropy. Applied Mathematics & Information Sciences, 6(2): 233-238.

Chick, S. E., Inoue, K. (2001). New two-stage and sequential procedures for selecting the best simulated system. Operations Research, 49 (5): 732-743.

Csiszer, I. (1972). A Class of Measures of Informativity of Observation Channels, Periodica Mathematica Hungarica, 2: 191-213.

Deshmukh, A.V. (1993). Complexity and chaos in manufacturing systems, PhD Thesis, Purdue University, USA.

Deshmukh, A.V., Talavage, J.J., Barash, M.M. (1998). Complexity in manufacturing systems. Part 1: Analysis of static complexity. IIE Transactions, 30(4): 35-44.

Egbelu, P.J., Tanchoco, J.M.A. (1984). Characterization of automatic guided vehicle dispatching rules. International Journal of Production Research, 22(3): 359-374.

Föllmer, H., Knispel, T. (2011). Entropic risk measures: coherence vs. convexity, model ambiguity, and robust large deviations. stochastic and dynamics, 11(333).

Frizelle, G., Woodcock, E., (1995). Measuring complexity as an aid to developing operational strategy. International Journal of Operation and Production Management, 15(5): 26-39.

Frizelle, G. (1996). An entropic measurement of complexity in manufacturing operations Internal Report of the Institute for Manufacturing. Cambridge University, Cambridge, UK.

Frizelle, G., Suhov, Y.M. (2001). An entropic measurement of queueing behavior in a class of manufacturing operations. Proceedings of the Royal Society, London, 457: 1579–1601.

Fujimoto, H., Ahmed, A., Iida, Y., Hanai, M. (2003). Assembly process design for managing manufacturing complexities because of product varieties. International Journal of Flexible Manufacturing Systems, 15: 283-307.

Gaur, D.R., Gupta, A., Krishnamurti, R. (2003). An approximation algorithm for scheduling vehicles on a path with release and handling times. Information Processing Letters, 86: 87-91.

Gibbs, J. W. (1902). Elementary Principles in Statistical Mechanics, Developed with Especial Reference to the Rational Foundations of Thermodynamics. C. Scribner's sons, New York.

Guan, X. and Dai, X. (2009). Deadlock-free multi-attribute dispatching method for AGV systems. International Journal of Advanced Manufacturing Technology, 45 (5–6), 603–615.

Gupta, S.S., Miescke, K.J. (1994). Bayesian look ahead one stage sampling allocations for selecting the largest normal mean. Statistical Papers, 35: 169-177.

Gupta, S. S., Miescke, K.J. (1996). Bayesian look ahead one-stage sampling allocations for selection of the best population. Journal of Statistical Planning and Inference, 54: 229-244.

Havrada, J. H., Charvat, F. (1967). Quantification methods of classification processes: Concepts of structural entropy. Kybernetika, 3: 30-35.

Ho, Y.C., Chein, S.H. (2006). A simulation study on the performance of task determination rules and delivery dispatching rules for multi-load AGVs. International Journal of Production Research, 44(20): 4193-4222.

Hong, L.J., Nelson, B.L. (2005). The tradeoff between sampling and switching: New sequential procedures for indifference-zone selection. IIE Transactions, 37: 623-634.

Hong, L.J., Nelson, B.L. (2007). Selecting the best system when systems are revealed sequentially. IIE Transactions, 39: 723-734.

Huang, X. (2008). Mean-entropy models for fuzzy portfolio selection. IEEE Transactions on Fuzzy Systems, 16: 1096-1101.

Hwang, H., Kim, S.H. (1998). Development of dispatching rules for automated guided vehicles, journal of manufacturing systems. journal of manufacturing systems, 17(2): 137-143.

Jana, P., Roy, T.K., Mazumder, S.K. (2009). Multi-objective possibilistic model for portfolio selection with transaction cost. Journal of Computational and Applied Mathematics, 1: 188–196.

Jaynes, E. T. (1957a). Information theory and statistical mechanics. Physical review, 106(4): 620-630.

Jaynes, E. T. (1957b). Information theory and statistical mechanics II, Physical review, 108(2): 171-190.

Jeong, B.H., Randhawa, S.U. (2001). A multi-attribute dispatching rule for automated guided vehicle systems. International Journal of Production Research, 39(13): 2817-2832.

Jones, B.L., Zitikis, R. (2007). Risk measures, distortion parameters, and their empirical estimation. Insurance: Mathematics and Economics, 41: 279–297.

Kapur, J. N., Kesavan, H. K. (1992). Entropy optimization principles with applications. Academic press, Inc., NewYork.

Kapur, J. N. (1993). Maximum Entropy Models in Science and Engineering. John Wiley & Sons, Inc., NewYork.

Kim, S.H., Nelson, B. L. (2001). A fully sequential procedure for indifference-zone selection in simulation. ACM Transactions on Modeling and Computer Simulation, 11: 251-273.

Kim, J., Klein, C. M. (1996). AGV dispatching. International Journal of Production Research, 34(1): 95-110.

Kim, C.W., Tanchoco, J.M.A. (1991). Conflict-free shortest-time bidirectional AGV routing. International Journal of Production Research, 29(12): 2377-2391.

Kim, C.W., Tanchoco, J.M.A., Koo, P.H. (1999). AGV dispatching based on workload balancing. International Journal of Production Research, 37 (17): 4053-4066.

Kim, S.H., Hwang, H. (1999). An adaptive dispatching algorithm for automated guided vehicles based on an evolutionary process. International Journal of Production Economics, 60–61: 465-472.

Kim, K.S., Chung, B.D., Park, Y.S. (2004). Minimum empty distance dispatching rules for a single loop-single vehicle AGV system, International Journal of Industrial Engineering, 11(3): 252-260.

Kullback, S. (1959). Information Theory and Statistics, Wiley, New York.

Kullback, S., Leibler, R. A. (1951). On information and sufficiency. Annals of mathematical statistics, 22(1): 79-86.

Kumar, V. (1987). Entropic measurement of manufacturing flexibility. International Journal of Production Research, 25 (7): 957-966.

Kumar, V. (1988). Measurement of loading and operations flexibility in flexible manufacturing systems: An information-theoretic approach. Annals of Operations Research, 15: 65-80.

Leshno, M., Levy, H. (2002). Preferred by “all” and preferred by “most” decision makers: Almost stochastic dominance. Management Science, 48(8):1074–1085.

Levy, H. (1992). Stochastic dominance and expected utility: Survey and analysis. Management Science, 38(4): 555–593.

Levy, H. (1998). Stochastic Dominance: Investment Decision Making Under Uncertainty. Kluwer Academic, Boston.

Levy, H., Leshno, M., Leibovitch, B. (2010). Economically relevant preferences for all observed epsilon. Annals of Operations Research. 176(1): 153–178.

Lexa, M. (2004). Useful Facts About the Kullback-Leibler Discrimination Distance. Rice Univ. Houston, TX, Tech. Rep.

Mandelbaum, M. (1990). Flexibility and decision making. European Journal of Operational Research, 44: 17–27.

Martinez-Olvera, C. (2008). Entropy as an assessment tool of supply chain information sharing. European Journal of Operations Research, 185, 405-417.

Mathai, A.M., Rathie, P.N. (1975), Basic Concepts in Information Theory and Statistics. Wiley Eastern Ltd., New Delhi.

Mukhopadhyay, S.K., Midha, S., Krishna, V.M. (1992). A heuristic procedure for loading problems in manufacturing systems. International Journal of Production Research, 30, 2213-2228.

Naso, D., Turchiano, B. (2005). Multi-criteria Meta-Heuristics for AGV Dispatching Control Based on Computational Intelligence. IEEE Transaction Systems, Man, and Cybernetics-PART B: Cybernetics, 35(2): 208-226.

Nelson, B. L., Swann, J., Goldsman, D., Song, W. (2001). Simple procedures for selecting the best simulated system when the number of alternatives in large. Operations Research, 49: 950-963.

Nelson, B. L., Staum, J. (2006). Control variates for screening, selection, and estimation of the best. ACM transactions on Modeling and Computer Simulation, 16: 52-57.

Nishi, T., Shimatani, K., Inuiguchi, M. (2009). Decomposition of Petri nets and Lagrangian relaxation for solving routing problems for AGVs. International Journal of Production Research, 47(14), 3957-3977.

Nishi, T., Ando, M., Konishi, M. (2005). Experimental studies on a local rescheduling procedure for dynamic routing of autonomous decentralized AGV systems. Robotics and Computer-Integrated Manufacturing, 22(2): 154-165.

Olivella, J., Corominas, A., Pastor, R. (2010). An entropy-based measurement of working time flexibility. European Journal of Operational Research, 200: 253-260.

Park, S., Rao, M., Shin, D.W. (2012). On cumulative residual Kullback-Leibler information. Statistics and Probability Letters, 82: 2025–2032.

Pereira, J., Paulre, B. (2001). Flexibility in manufacturing systems: A relational and a dynamic approach, European Journal of Operational Research, 130: 70–82.

Pichitlamken, J., Nelson, B. L., Hong. L. (2006). A sequential procedure for neighborhood selection-of-the-best in optimization via simulation, European Journal of Operational Research, 173 283-298.

Piplani, R., Wetjens, D. (2007). Evaluation of entropy-based dispatching in flexible manufacturing systems. European Journal of Operational Research, 176: 317-331.

Rao, H.A., Gu, P. (1994). Flexibility in cellular manufacturing systems: An entropic measure. Proceedings of the 10th ISPE/IFAC International Conference on CAD/CAM, Robotics and Factories of the Future, 411-416.

Rinott, Y. (1978). On two-stage selection procedures and related probability inequalities. Communications in Statistics, A7: 799-811.

Rödder, W., Gartner, I.R., Rudolph, S. (2010). An entropy-driven expert system shell applied to portfolio selection. Expert Systems with Applications, 37(12): 7509–7520.

Sabuncuoglu, I. (1998). A study of scheduling rules of flexible manufacturing systems: A simulation approach. International Journal of Production Research, 36(2): 527-546.

Scuricini, G.B. (1987). Complexity in large technological systems, Measures of Complexity. Rome, Italy, 83–101.

Sethi, A.K., Sethi, S.P. (1990). Flexibility in manufacturing: A survey. International Journal of Flexible Manufacturing Systems, 2: 289-328.

Shannon, C. E. (1948). A mathematical theory of communication, Bell System Technical Journal, 27:379-423, 623-656.

Sharma, B. D., Mittal, D. P. (1975). New non-additive measures of entropy for discrete probability distributions. Journal of mathematical sciences, 10:28-40.

Shuiabi, E., Thomson, V., Bhuiyan, N. (2005). Entropy as a measure of operational flexibility. European Journal of Operational Research, 165: 696-707.

Sivadasan, S., Efstathiou, J., Frizelle, G., Shirazi, R., Calinescu,A. (2002). An information-theoretic methodology for measuring the operational complexity of supplier–customer systems, International Journal of Operations and Production Management, 22(1): 80-102.

Sivadasan, S., Efstathiou, J., Calinescu, A., Huaccho, H. L. (2006). Advances on measuring the operational complexity of supplier– customer systems. European Journal of Operations Research, 171(1): 208-226.

Sivadasan, S., Smart, J., Huaccho, H. L., Calinescu, A. (2010). Operational complexity and supplier–customer integration: case study insights and complexity rebound. Journal of the Operational Research Society, 61: 1709 -1718.

Sriboonchita, S., Wong, W.K., Dhompongsa, S., Nguyen, H.T. (2009). Stochastic Dominance and Applications to Finance, Risk and Economics. Chapman and Hall/CRC, Taylor &Francis Group, Boca Raton, FL, USA.

Stecke, K.E. (1983). Formulation and solution of nonlinear integer production planning problems for flexible manufacturing systems, Management Science, 29: 273-288.

Tan, K.K., Tang, K.Z. (2001). Vehicle dispatching system based on Taguchi-tuned fuzzy rules. European Journal of Operational Research, 128: 545–557.

Tsai, S.C., Nelson, B.L., Staum, J. (2009). Combined screening and selection of the best with control variates. In: Advancing the Frontiers of simulation: A Festschrift in Honor of George S. Fishman, Kluwer, 263-289.

Tzeng, L.Y., Huang, R.J., Shih, P.T. (2012) Revisiting Almost Second-Degree Stochastic Dominance. Management Science, Articles in Advance, 1-5.

Qina, Z., Lia, X., Ji, X. (2009). Portfolio selection based on fuzzy cross-entropy. Journal of Computational and Applied Mathematics, 228: 139-149.

Veeravalli, B., Rajesh, G., Viswanadham, N. (2002). Design and analysis of optimal material distribution policies in flexible manufacturing systems using a single AGV. International Journal of Production Research, 40(12): 2937–2954.

Wu, Y., Frizelle, G., Efstathiou, J. (2007). A study on the cost of operational complexity in customer–supplier systems. International Journal of Production Economy, 106(1): 217–229.

Wu, Z.W., Song, X.F., Xu, Y.Y., Liu, K. (2009). A note on a minimax rule for portfolio selection and equilibrium price system. Applied Mathematics and Computation, 208: 49–57.

Yamashita, H. (2001). Analysis of dispatching rules of AGV systems with multiple vehicles. IIE Transactions, 33: 889–895.

Yang, J., Qiu, W. (2005). A measure of risk and a decision-making model based on expected utility and entropy. European Journal of Operational Research, 164: 792–799.

Yao, D. D. (1985). Material and information flows in flexible manufacturing systems. Material Flows, 2: 143–149.

Yao, D. D., Pei, F.F. (1990). Flexible parts routing in manufacturing systems. IIE Transactions, 1990, 22: 48–55.

Yu, S. B., Efstathiou, J. (2006). Complexity in rework cells: Theory, analysis and comparison. Journal of the Operational Research Society, 57(5): 593–602.

Zaremba, M.B., Obuchowicz, A., Banaszak, Z.A., Jedrzejek, K.J. (1997). A max-algebra approach to the robust distributed control of repetitive AGV systems. International Journal of Production Research, 35(10): 2667–2687.

Zhang, W., Liu, Y.J., Wei-Jun Xu. (2012). A possibilistic mean-semivariance-entropy model for multi-period portfolio selection with transaction costs. European Journal of Operational Research, 222: 341–349.