HÖGSKOLAN I BORÅS
INSTITUTIONEN INGENJÖRSHÖGSKOLAN

Applications of the Law of Large Numbers in Logistics

Navid Bazzazian

Master Thesis
Borås, Sweden 2007

This thesis comprises of 20 credits and is a compulsory art in the Master of Science Degree with Major in Industrial Management with Specialization in Logistics
Nr 1/2007

Applications of the Law of Large Numbers in Logistics

Navid Bazzazian, 2007

Master Thesis
Subject Category: Technology - Industrial Engineering and Logistics
Series Number: 1/2007

University College of Borås
School of Engineering
SE-501 90 BORÅS
Telephone +46 033 435 4640

Examiner: Maria Fredriksson
Supervisor: Kenth Lumsden; Department of Logistics and Transportation, Chalmers University of Technology

Abstract

One of the most remarkable theories in probability and statistics is the *law of large numbers*. Law of large numbers describes the behavior of random phenomena when they are reiterated infinitely or in very large trials. Apart from the mathematical exposition of the law of large numbers, its theory and applications have been widely used in gambling houses, financial sectors, and healthcare insurance where uncertainties deteriorate prediction and financial strength. However, the applications of the law of large numbers are not confined to the referred sectors and could be widely applied to industrial organizations and service provider companies in which large number of stochastic phenomena incorporate in their planning. In this thesis, the applications of the law of large numbers are studied in relation to logistics and transportation under conditions of operating in large networks. The results of this study assert that transportation companies can benefit from operating in large networks to increase the filling performance of their vehicles, fleet, etc. Equivalently, according to the law of large numbers the inferior capacity utilization in unit loads, containers, etc. converges to 0 with probability 1 as the size of the network grows.

Keywords: Law of Large Numbers, Logistics, Filling Performance, Knapsack Problem

Chapter 1

Introduction

1.1 Motivation

One of the most remarkable theories in probability and statistics is the *law of large numbers.* Law of large numbers describes the behavior of random phenomena when they are reiterated infinitely or in very large trials.

It has been ascertained that determinist phenomena have a very small part in surrounding nature. The vast majority of phenomena from nature and society are stochastic (random). An example of stochastic phenomena in industry is the weekly demands of customers for a product or the orders coming into the company over time. However when these stochastic phenomena reiterate infinitely over time or happen in very large trials, some of their characteristics such as their mean and true distribution can be well predicted. This encouraged the evolution of the law of large numbers in probability and statistics which is able to predict with uncanny precision, the overall result of a large number of individual events, each of which, in itself, is totally unpredictable. In other words, we are faced with a large number of uncertainties producing a certainty.

Apart from the mathematical exposition of the law of large numbers, its theory and applications have been widely used in gambling houses, financial sectors, and healthcare insurance where uncertainties deteriorate prediction and financial strength. When relating this concept to finance, it suggests that as a company grows, its chances of sustaining a large percentage in growth diminish. This is because as a company continues to expand, it must grow more and more just to maintain a constant percentage of growth. As an example, assume that company X has a market capitalization of $400 billion and company Y has a market capitalization of $5 billion. In order for company X to grow by 50%, it must increase its market capitalization by $200 billion, while company Y would only have to increase its market capitalization by $2.5 billion. The law of large numbers suggests that it is much more likely that company Y will be able to expand by 50% than company X [Investopedia, 2007].

Insurance relies heavily on the law of large numbers. In large homogeneous populations it is possible to estimate the normal frequency of common events such as deaths and accidents. Losses can be predicted with reasonable accuracy, and this accuracy increases as the size of the group expands. From a theoretical standpoint, it is possible to eliminate all pure risk if an infinitely large group is selected [Encyclopædia, 2007].

The applications of the law of large numbers are not confined to the referred sectors and could be widely applied to industrial organizations and service provider companies in which large number of stochastic phenomena incorporate in their planning. Within this context, the globalization of markets and increasing the market share of companies in recent years include large number of stochastic data. It can be perceived that this huge number of stochastic data will not add to complexity of the whole system yet it will give accurate prediction of the characteristics of stochastic data. So, further research on the applications of the law of large numbers is necessary.

In this thesis, the applications of the law of large numbers are studied in relation to logistics and transportation under conditions of operating in large markets or networks. According to the law of large numbers, it is conceived that operating in large markets will increase the filling rate of unit loads such as containers in despite of variations of each individual element in the large market. The concept of knapsack problem is also introduced in this thesis to analyze the results and provide a mathematical framework for the problem.

1.2 Problem Description

Transportation companies and haulers operate between nodes in logistics network. Figure 1.2.1 represents a typical logistics network and operating environment of transportation companies and haulers.

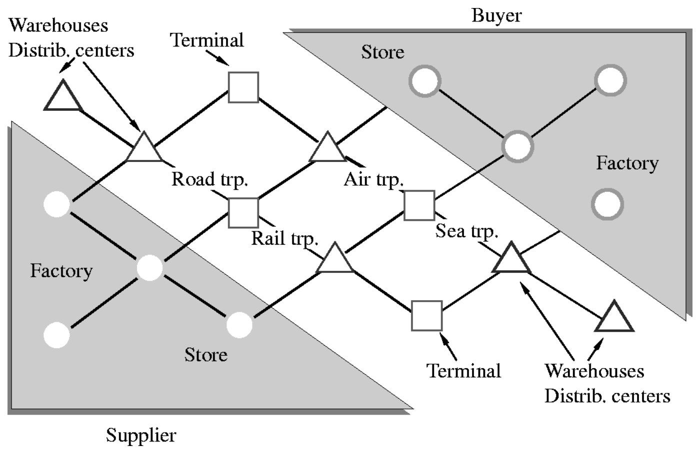


Figure 1.2.1: Logistics network [Stefansson, 2006]

Transportation companies and haulers transporting goods and consignments between nodes in logistics network are often confronted by unexploited space in unit loads and resources used to transfer goods and consignments, figure 1.2.2. These unused capacities deteriorate the capacity utilization of carrying assets such as trucks, containers, etc. The major consequence of inferior capacity utilization can be seen as excessive transportation costs and poor resource utilization.

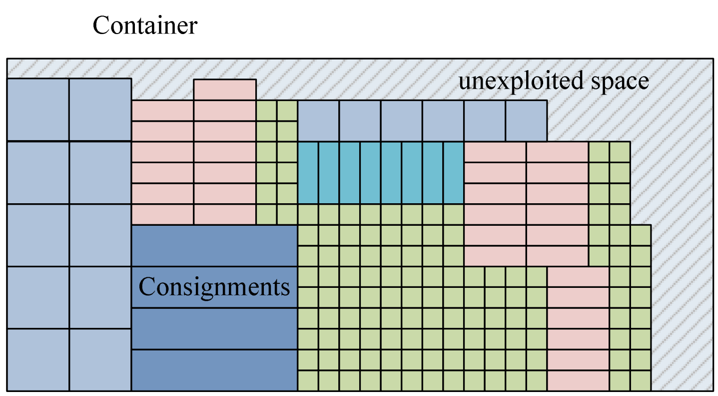


Figure 1.2.2: unexploited space in a container

Combination of consignments comprises a cargo. The closer the size of the cargo to the capacity of the career the lower unexploited space would be. The optimum capacity utilization is achieved when the size of the cargo is equal to the capacity of the career or it is in a very small neighborhood of the capacity. However, since the sizes of consignments vary due to the volatile and uncertain nature of demands of individual customers, the size of the cargo would fluctuate leading to imbalances between its size and the capacity of the career. Variation in the size of the consignments according to a normal distribution is shown in figure 1.2.3.

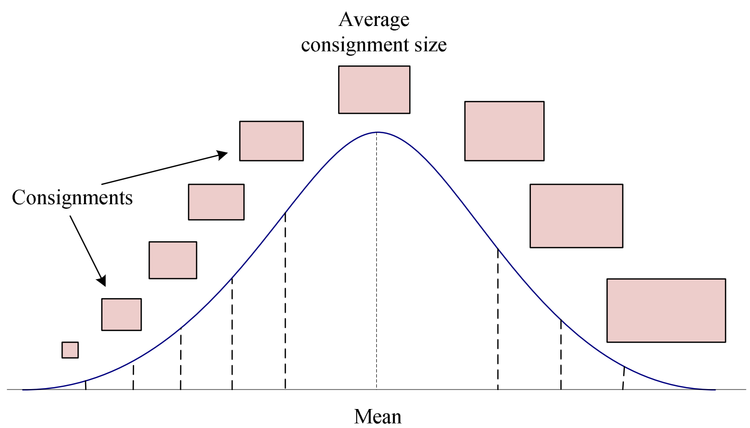
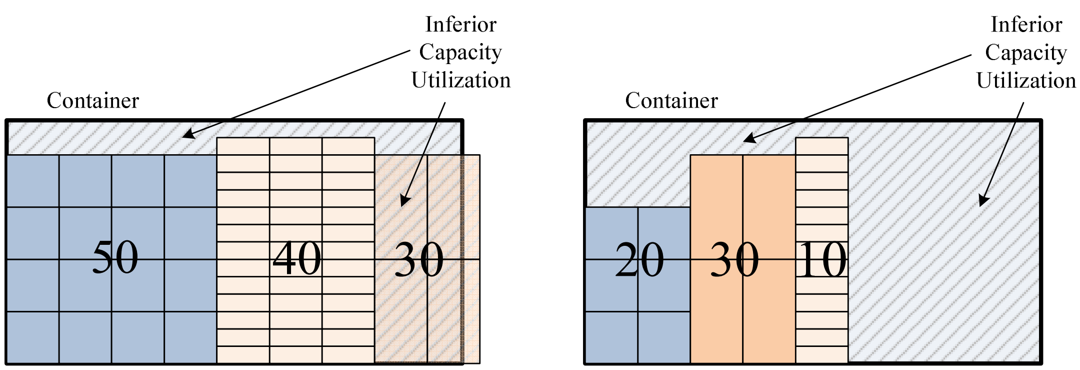


Figure 1.2.3: variation in the size of the consignments

The inferior capacity utilization occurs either when the combination of consignments exceeds the actual capacity of the career or it is much lower than the actual capacity of the career. Combinations exceeding the actual capacity of a career should be split for fitting purpose and consequently result in unexploited space.

To clarify the problem, suppose that a transportation company operates in a market with 3 customers whose demands are uncertain and follow normal distribution with mean 50 units and variance ± 30 units. This means that the actual demands of customers can appear to be any number between 20 and 80 units with respect to their probability of occurrences. The capacity of the career is assumed to be 100 units. When actual demands are appeared, if double or triple combination of them exceeds the capacity of 100 units, combinations should be reduced to fit the capacity which consequently leads to unexploited space. On the other hand if the combinations are much lower than the capacity, poor capacity utilization is perceived. This phenomena leads to fluctuations in capacity utilization due to uncertain nature of customer demands and consequently deteriorate planning and transportation costs. An example of 3 random consignment sizes and their subsequent effect on capacity utilization is illustrated in figure 1.2.4.



(a) Inferior capacity utilization due to large consignment sizes

(b) Inferior capacity utilization due to small consignment sizes

Figure 1.2.4: Inferior capacity utilization in a container with the capacity of 100 units

To increase the capacity utilization, it is perceived that operating in large markets would reduce the unexploited space and increase the filling rate in vehicles. However, one may argue that operating in large markets not only would increase the complexity of the system but also would not generate combinations close enough to the capacity. In this thesis, it is being argued that this will not happen and *n*-tuple combinations very close to the capacity will be generated when the market size grows.

Chapter 2

Methodology

Hereinafter follows a description of how the study was conducted. It includes approaches, methods employed, and literature sources and the general argumentation for performing the study.

2.1 Research Approach

In theory of scientific thinking two different research paradigms have been defined: *Positivism* and *Hermeneutic.* Developed by Auguste Comte, widely regarded as the first true sociologist, positivism is a philosophy stating that the only authentic knowledge is scientific knowledge, and such knowledge can only come from positive affirmation of theories through strict scientific method [Mill, 1961]. The positivistic ideal toward research is based on experiments, quantitative measures and logical reasoning [Wiedersheim & Eriksson, 1991]. Positivists strive to find casual relations in explaining reality. Positivists look for a testable objective truth and the approach is mainly deductive [Mårtensson & Nilstun, 1988]. While positivism describes and explains, hermeneutics emphasize the qualitative ideals of research [Wiedersheim & Eriksson, 1991]. The advocates for this research ideal stress the importance of subjective elements in research saying that the researcher's individual experiences affect and help to understand the research question.

The research problem in this thesis is investigated using a research ideal positioned closer to the positivism ideal since it carries a technological and scientific context. Hypothesis and conclusions are drawn and analyzed by using earlier studies on scientific theories.

2.2 Research Method

Deductive and *inductive* reasoning are two methods of logic used to arrive at a conclusion based on information assumed to be true. Both are used in research to establish hypothesis. Deductive reasoning arrives at a specific conclusion based on generalizations and involves a hierarchy of statements or truths. Starting with a limited number of simple statements or assumptions, more complex statements can be built up from the more basic ones [Trochim, 1999]. According to figure 2.2.1, deductive reasoning process begins with thinking up a theory about a topic of interest. Narrowing down from the theory a more specific, testable hypothesis is created in the next step. Afterwards, observations can be used to narrowing down further addressing the hypothesis. This process ultimately leads to confirmation of hypothesis using specific data.

Inductive reasoning, on the other hand, is essentially the opposite of deductive reasoning. It involves trying to create general principles by starting with many specific instances. This is the kind of reasoning used if you have gradually built up an understanding of how something works, rather than starting with laws and principles and making deductions [Trochim, 1999]. As it is shown in figure 2.2.2, inductive reasoning begins with specific observations and measures. Then attempts to detect patterns and regularities, formulate some tentative hypotheses that can be explored, and finally end up developing some general conclusions or theories [Trochim, 1999]. In essence, inductive reasoning progresses from observations of individual cases to the development of a generality.

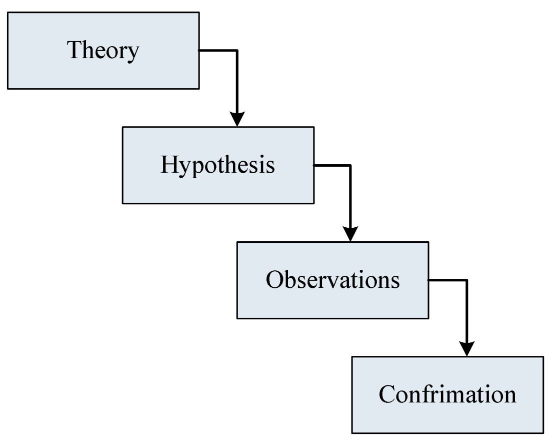


Figure 2.2.1: deductive reasoning [Trochim, 1999]

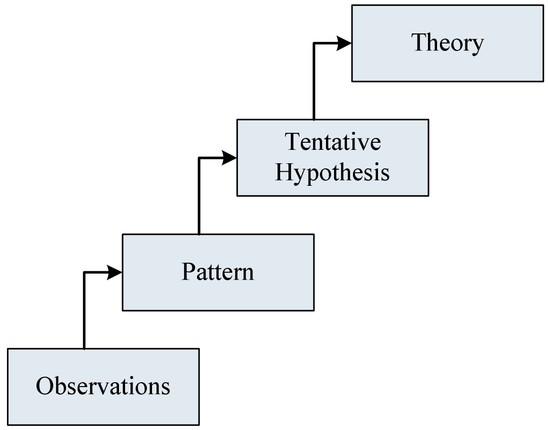


Figure 2.2.2: inductive reasoning [Trochim, 1999]

The problem dealt with in the thesis calls for deductive method since its objective is to explore a hypothesis drawn from theories in literature. Starting the research process, the problem is extracted from theories of transport economy in logistics and empirical phenomena in reality. Following the patterns and theories in statistics and operations research a tentative hypothesis is generated in connection with the problem. The hypothesis is then confirmed by deduction and proofs using theories in statistics and operations research. Following the confirmation, the hypothesis is further analyzed and conclusions are drawn.

2.3 Literature

Throughout different research stages in the present thesis, available literatures on the topic have been the main source of information and required data. These literatures also served as foundation of the theoretical frame of reference in the next chapter. To make the project feasible it was important to accomplish an extensive literature study as early as possible as it increased the understanding of the character of the problem which facilitated the data collection and the defining of the design of the master thesis. Literature of importance was major course references regarding logistics and transportation, operations research, and probability and statistics. Additional information on each topic also gathered from articles and published papers.

From the literature and references in connection with logistics and transportation, parts highlighting resource utilization in transportation, transport economy, and customer service level have been extensively studied. These parts serve as sources for problem description.

Topics on the law of large numbers were the major focus of study in probability and statistics literature and relevant articles. The law of large numbers together with the problem definition constructs the hypothesis. The hypothesis was further analyzed using available literature on operations research specially topics regarding knapsack problem.

Chapter 3

Theoretical Foundation

This chapter is based on theoretical textbooks, articles in scientific journals, and conference papers. The sources discussed in this chapter, mainly of a theoretic nature, to a large extend refer to fundamental aspects of the problem in probability theory, statistics, and operations research. This chapter covers two imperative laws in probability theory and a fundamental operations research problem which altogether constitute the theoretical foundation that this thesis rests on.

3.1 Introduction

The most remarkable theoretical results in probability theory are limit theorems. Of these, the most important are those that are classified either under the heading laws of large numbers or under the heading central limit theorems [Ross, 2000]. Usually, theorems are considered to be laws of large numbers if they are concerned with stating conditions under which the average of a sequence of random variables converges (in some sense) to the expected average. The laws of large numbers have been studied in two perspectives: the weak law of large numbers and the strong law of large numbers which are discussed extensively in the first section of this chapter. Knapsack problem is also introduced in this chapter. Knapsack problem serves as an operations research model for many industrial situations such as capital budgeting, cargo loading, cutting stock, etc. The law of large numbers combined by knapsack problem would represent a mathematical model for the meant hypothesis in this thesis.

In research, theories can be used differently. Theories can be viewed as being true story yet they can also be used as ideas about how reality works and to generate hypothesis. In this thesis theories are used instrumentally rather than to, by themselves, generate hypothesis. Theories used in this thesis serve as an instrument to formulate, analyze a problem, and draw conclusions.

3.2 The Law of Large Numbers

The most important theoretical results in probability theory are limit theorems. Probably the best reason why the study of limiting behavior of sequences of random variables is important is that when we use random samples to estimate distributional parameters, we would like to know that as the sample size gets larger, the estimates are likely to be close to the parameters that they are estimating [Hastings, 1997]. Recalling from probability theory, an intuitive way to view the probability of a certain outcome is as the frequency with which that outcome occurs in the long run, when the experiment is repeated a large number of times. However, to justify the use of empirical frequencies to estimate actual probabilities, proving the convergence of a sequence of sample means to the population mean is necessary. Therefore, the law of large numbers, which is a theorem proved about the mathematical model of probability, was introduced to show that this model is consistent with the frequency interpretation of probability.

The law of large numbers is twofold: the weak law of large numbers and the strong law of large numbers. Both the weak and strong laws refer to convergence of sample mean to population mean as the sample size grows. However, they differ in the way of describing the convergence of the sample mean with the population mean.

The weak law of large numbers is drawn from Chebyshev's inequality which is a corollary of Markov's inequality. Hence, this section starts by representing a result known as Markov's inequality. Sections 3.2.1, 3.2.2, and 3.2.3 are for a large part based on the book [Ross, 2000].

3.2.1 Markov's inequality

If X is a random variable that takes only nonnegative values, then for any value $a > 0$,

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

Proof: for a> 0, let

$$I = \begin{cases} 1 & if X \geq a \\ 0 & otherwise \end{cases}$$

Since $X \geq 0$,

$$I \leq \frac{X}{a}$$

Taking expectations of the preceding yields that

$$E[I] \leq \frac{E[X]}{a}$$

Since $E[I] = 1 \cdot P\{X \geq a\} + 0 \cdot P\{X < a\} = P\{X \geq a\}$, proves the result.

3.2.2 Chebyshev's inequality

If X is a random variable with finite mean μ and variance σ^2, then for any value K> 0,

$$P\{|X-\mu| \geq K\} \leq \frac{\sigma^2}{K^2}$$

Proof: Since $(X-\mu)^2$ is a nonnegative random variable, Markov's inequality with $a = K^2$ can be applied to obtain

$$P\left\{(X-\mu)^2 \geq K^2\right\} \leq \frac{E\left[(X-\mu)^2\right]}{K^2}$$

But since $(X-\mu)^2 \geq K^2$ if and only if $|X-\mu| \geq K$, the above equation is equivalent to

$$P\left\{|X-\mu| \geq K\right\} \leq \frac{E\left[(X-\mu)^2\right]}{K^2} = \frac{\sigma^2}{K^2}$$

and the proof is complete.

The importance of Markov's and Chebyshev's inequalities is that they make possible to derive bounds on probabilities when only the mean, or both the mean and the variance, of the probability distribution are known.

3.2.3 The weak law of large numbers

The weak law of large numbers is derived from Chebyshev's inequality. If $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean $E[X_i] = \mu$, then, for any $\varepsilon > 0$,

$$P\left\{\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right\} \to 0 \quad as \quad n \to \infty$$

Proof: the result is proven only under the additional assumption that the random variables have a finite variance σ^2. Now, as

$$E\left[\frac{X_1 + \ldots + X_n}{n}\right] = \mu \quad \text{and} \quad Var\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

It follows from Chebyshev's inequality that

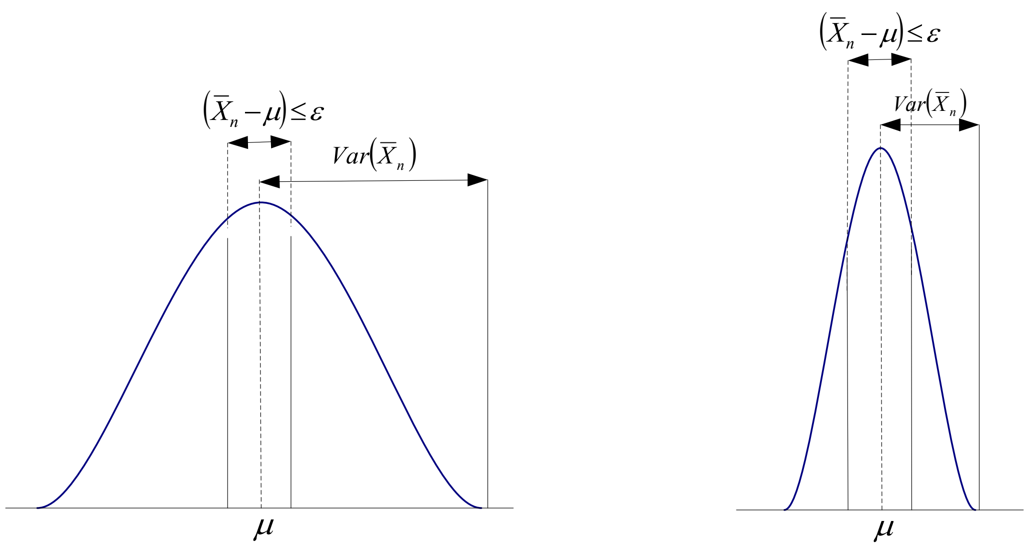
$$P\left\{\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right\} \leq \frac{\sigma^2}{n\varepsilon^2}$$

The above inequality clearly shows that when n tends to infinity, $\frac{\sigma^2}{n\varepsilon^2}$ tends to 0, and the result is proved. It should be noted that $\overline{X} = \frac{X_1 + ... + X_n}{n}$ is an average of the individual outcomes, which converges to population's expected value (mean) as the sample size grows. It is therefore a striking fact that one can start with a random experiment about which little can be predicted and, by taking averages, obtain an experiment in which the outcome can be predicted with a high degree of certainty.

As an instance, consider the important special case of Bernoulli trials with probability p for success. Let $X_i = 1$ if the ith outcome is a success and 0 if it is a failure. Then $S_n = X_1 + X_2 + ... + X_n$ is the number of successes in n trials and $\mu = E(X_i) = p \quad \forall \quad i = 1,...,n$. The weak law of large numbers states that for any $\varepsilon > 0$

$$P\left\{\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right\} \to 0 \qquad \text{or} \qquad P\left\{\left|\frac{S_n}{n} - \mu\right| < \varepsilon\right\} \to 1 \qquad as \quad n \to \infty$$

The above statement says that, in a large number of repetitions of a Bernoulli experiment, it can be expected that the proportion of times the event will occur to be near p. This shows that the mathematical model of probability agrees with the frequency interpretation of probability. Figure 3.2.1 depicts the graphical representation of the weak law based on the normal curve for small sample sizes and large sample sizes respectively.



(a) small sample sizes

(b) large sample sizes

Figure 3.2.1: behavior of the sample mean in connection with small and large sample sizes

It can be inferred from figure 3.2.1(a) that having small sample sizes, the variation of sample mean from population mean is large. Hence, the probability that the absolute difference of sample mean and population mean be less than or equal to ε is small. Contrastingly, in the case where sample sizes are sufficiently large, figure 3.2.1(b), the variation of sample mean from population mean becomes so small that the so called probability increases.

3.2.4 The strong law of large numbers

The strong law of large numbers is probably the best-known result in probability theory. It states that the average of a sequence of independent random variables having a common distribution will, with probability 1, converge to the mean of that distribution.

In another account, if $X_1, X_2,...$ be a sequence of independent and identically distributed random variables, each having finite mean $E[X_i] = \mu$, then, with probability 1,

$$\frac{X_1 + X_2 + ... + X_n}{n} \rightarrow \mu \qquad as \quad n \rightarrow \infty$$

This theorem will not be proved here, yet its statement is well worth knowing in proceeding sections.

3.2.5 Weak laws versus strong laws

In this section $X_1, X_2,...$ is a sequence of independent and identically distributed random variables with finite expectation $E[X_i] = \mu$. The associated sequence $\overline{X}_i$ of partial sample means by

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

The Laws of Large Numbers make statements about the convergence of $\overline{X}_n$ to μ..
Both laws relate bounds on sample size, accuracy of approximation, and degree of confidence. The Weak Laws deal with limits of probabilities involving $\overline{X}_n$. The Strong Laws deal with probabilities involving limits of $\overline{X}_n$.
The weak law shows the convergence of $\overline{X}_n$ to μ. as

$$\forall \varepsilon > 0 \quad \lim_{n \to \infty} pr\left(\left| \overline{X}_n - \mu \right| \leq \varepsilon \right) = 1$$

This is often abbreviated to

$$\overline{X}_n \xrightarrow{P} \mu \qquad as \quad n \rightarrow \infty$$

Or in words: $\overline{X}_n$ converges in probability to μ. as $n \rightarrow \infty$.
However, the strong law shows the convergence of $\overline{X}_n$ to μ. as

$$\Pr\left(\lim_{n\to\infty} \overline{X}_n = \mu\right) = 1$$

This is often abbreviated to

$$\overline{X}_n \xrightarrow{a.s.} \mu \qquad as \quad n \to \infty$$

Or in words: $\overline{X}_n$ converges almost surely to μ. as $n \to \infty$.

In essence, the strong law and the weak law do not describe different actual laws but instead refer to different ways of describing the convergence of the sample mean with the population mean. The reason for the choice of adjectives, weak and strong, is that the almost sure convergence of a sequence of random variables implies the convergence in probability of the sequence, but not vice versa, in general. This means that the weak law can be inferred from the strong law.

The graphical representation of the strong law of large numbers is illustrated in figure 3.2.2. 400 observations from the N (0, 1) distribution were simulated and progressively, as each ten new observations went by, the sample mean for all sample values to date were recomputed [Hastings, 1997].

That $\overline{X}_{10} = \frac{(X_1 + ... + X_{10})}{10}$, $\overline{X}_{20} = \frac{(X_1 + ... + X_{20})}{20}$, ..., $\overline{X}_{400} = \frac{(X_1 + ... + X_{400})}{400}$ were computed. The figure shows the running means wandering a bit but generally tending toward 0, which is the mean of the distribution being sampled from.

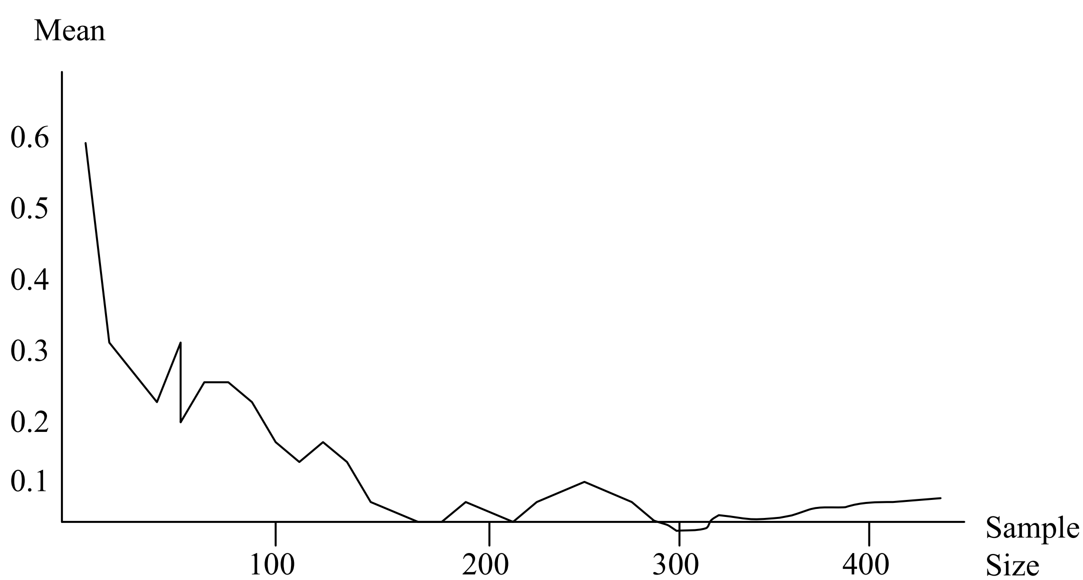


Figure 3.2.2: strong law of large numbers

According to the mathematical definition of the weak law of large numbers, its graphical representation could be as figure 3.2.3, in which the convergence of probabilities is illustrated.

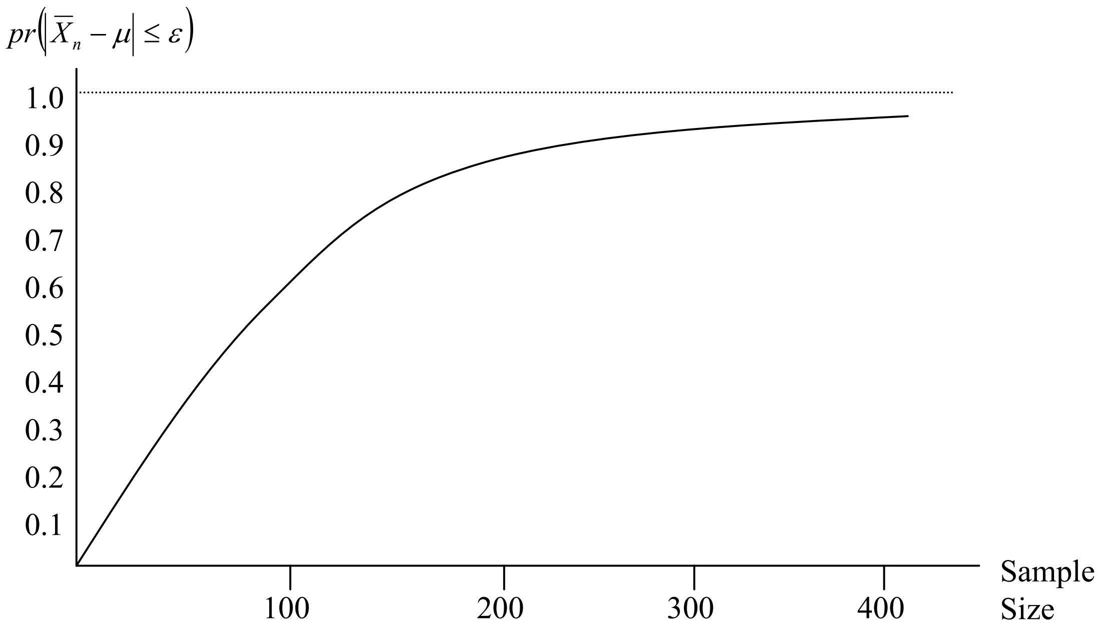


Figure 3.2.3: weak law of large numbers

3.3 Knapsack Problems

The knapsack problem is a problem in combinatorial optimization. It derives its name from the maximization problem of choosing possible essentials that can fit into one bag (of maximum weight) to be carried on a trip. Suppose a hitch-hiker has to fill up his knapsack by selecting from among various possible objects those which will give him maximum comfort [Martello & Toth, 1990]. This knapsack problem can be mathematically formulated by numbering the objects from 1 to n and introducing a vector of binary variables $x_j \quad \forall \quad j = 1,...,n$ having the following meaning:

$$x_j = \begin{cases} 1 & \textit{if object } j \textit{ is selected} \\ 0 & \textit{otherwise} \end{cases}$$

Then, if p_j is a measure of the comfort given by object j, w_j its size and c the size of the knapsack, the problem will be to select, from among all binary vectors x satisfying the following constraint:

$$\sum_{j=1}^{n} w_j x_j \leq c$$

the one which maximizes the objective function:

$$\sum_{j=1}^{n} p_j x_j$$

This problem is representative of a variety of knapsack-type problems in which a set of entities are given, each having an associated value and size, and it is desired to select one or more disjoint subsets so that the sum of the size in each subset does not exceed (or equals) a given bound and the sum of the selected values is maximized. Knapsack problems have been intensively studied, especially in the last decade, attracting both theorists and practicians. The theoretical interest arises mainly from their simple structure which, on the one hand allows exploitation of a number of combinatorial properties and, on the other, more complex optimization problems to be solved through a series of knapsack-type sub problems. From the practical point of view, these problems can model many industrial situations such as capital budgeting, cargo loading, cutting stock, to mention the most classical applications.

Chapter 4

Propagation

The intention of this chapter is to disseminate the theory of the law of large numbers to logistics and transportation and authenticate the claimed hypothesis with mathematical perspective. The hypothesis is interpreted mathematically in connection with theory of the law of large numbers and the concept of knapsack problem in operations research. Mathematical results will be used for further analysis of hypothesis in logistics and transportation in the next chapter.

4.1 Filling Performance

In retrospect to the first chapter, it is conceived that transportation companies can benefit from conditions in large markets to increase the filling rate of their fleet.

Statistical characteristic of large markets is very similar to conditions in the theory of the law of large numbers. Individual demands of customers can be considered as independent and identically distributed random variables; meanwhile number of customers in the market creates the sample size. Moreover, fleet size can be seen as a knapsack of size *C* which should be filled with the sum (combination) of random variables. Without loss of generality, the hypothesis is whether or not the sum of i random variables would be in a given neighborhood of *C* when the sample size grows. This implies the following inequality:

$$\left(C - \sum_{1}^{i} X_i\right) \leq \varepsilon$$

in which $X_1, X_2, ..., X_i$ represent i random variables drawn from a large sample of random variables. Intuitively, it is surmised that the frequency of the sum of i random variables in a given neighborhood of *C* would be increased as the sample size grows. Generalizing this statement to $i+1, i+2, ...$ etc. combinations, would suggest that with probability 1 there exist a combination of random variables which is in a given neighborhood of *C*.

4.1.1 Combinatorics in cargo loading

Let's assume that $X_1, X_2, ..., X_n$ is a sequence of independent and identically distributed random variables representing the demand of individual customers in a given transportation market. Also, each random variable has finite mean $E[X_i] = \mu$.

Distribution of the sum (combination) of i random variables arbitrarily selected from the above sample would have:

$$E[X_1 + X_2 + ... + X_i] = i \cdot E[X_i] = i \cdot \mu \quad \text{and} \quad Var(X_1 + X_2 + ... + X_i) = i \cdot Var(X_i) = i \cdot \sigma^2$$

Figure 4.1.1 depicts the distribution of the sum of i random variables assuming to have normal distribution.

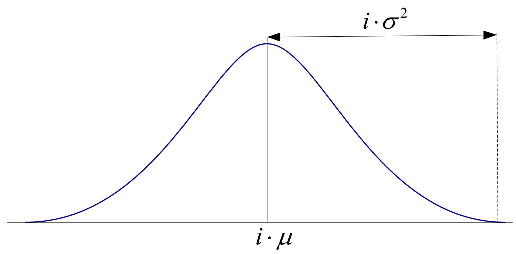


Figure 4.1.1: distribution of the sum of *i* random variables

This is equal to cargo loading logic where the combination of consignments from a given collection is used to fill up the capacity of a cargo or knapsack. However, since the size of the consignments is random with finite mean and variance, the sum of the consignments would also be random with respective mean and variance shown in figure 4.1.1. A typical combination of consignments used to fill up a container is shown in figure 4.1.2.

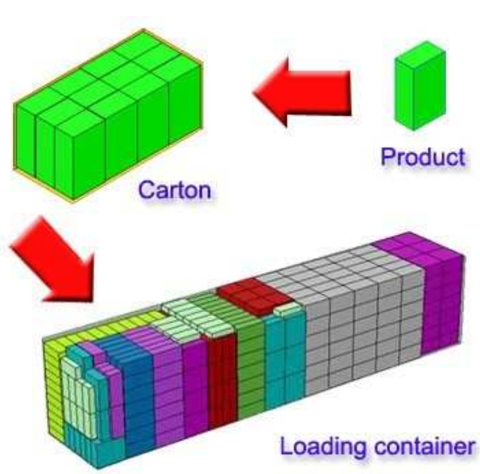
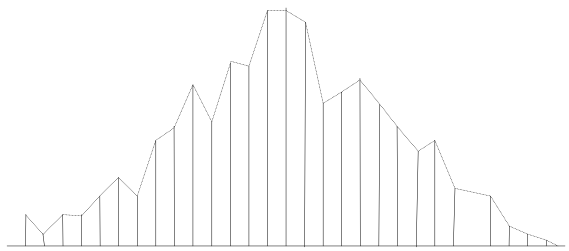


Figure 4.1.2: cargo loading

Moreover, in cargo loading or knapsack problem, all the possible combinations of actual consignments create a discrete distribution as it is shown in figure 4.1.3. In this figure, longer bars denote those combinations that have occurred more frequently. Also, the variable S_j denotes the sum of i consignments for each combination j.



$$Average = \frac{1}{j}\sum_{1}^{j} S_j$$

Figure 4.1.3: discrete distribution of the sums

All the possible combination of i consignments drawn from collection of n consignments is equal to number of subsets each containing i elements out of a set of n elements.

$$\binom{n}{i} = \frac{n!}{i!(n-i)!} = \frac{(n-i+1)!}{i!}$$

Therefore

$$S_1, S_2, S_3, ..., S_j \qquad j = 1,2,..., \frac{(n-i+1)!}{i!}$$

would be a sequence of independent and identically distributed random variables denoting the sum of i consignments drawn from a collection of n consignments, $X_1, X_2, ..., X_n$. The mean and variance of each S_j is finite and equals $i \cdot \mu$ and $i \cdot \sigma^2$ respectively.

4.1.2 Hypothesis

As mentioned earlier in the chapter, the hypothesis is whether or not better filling performance of a knapsack would be achieved as the sample size, number of the consignments, grows. To prove the hypothesis it is sufficient to show that the frequency of the sum of i consignments, drawn from a collection of n consignments, in a given neighborhood of the knapsack capacity, *C*, converges to a positive probability *p* as the number of the consignments grows.

Distribution function of the sum of i consignments creates a twofold situation. It will either include the knapsack capacity or not. If it includes the knapsack capacity then there exists a positive probability, *p* that the sum of i consignments is in a given neighborhood of *C*, otherwise capacity is larger than the maximum variance of the distribution function and the number of the elements in the combination should be increased beyond i. These situations are depicted in figure 4.1.4.

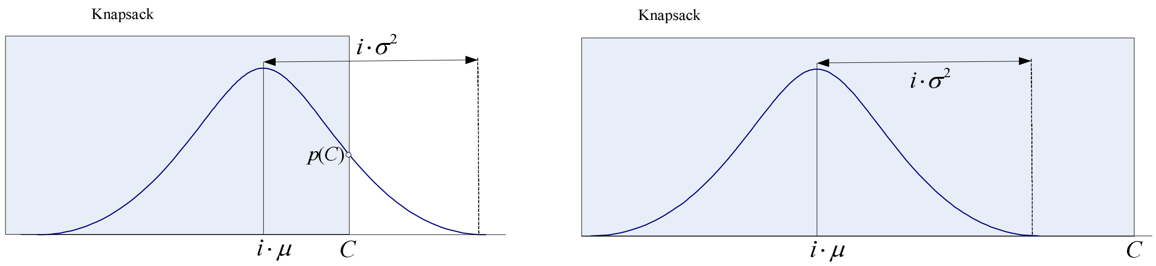


Figure 4.1.4: (a) distribution function includes capacity; (b) capacity exceeds distribution function

Definition: Indicator function

An indicator function or a characteristic function is a function defined on a set S that indicates membership of an element in a subset A of S. considering this notion, the indicator function defined on a set of random variables, S and subset $A = \{S_k | \quad (S_k - C) \leq \varepsilon; \quad \forall \quad k = 1,2,...j \quad and \quad \varepsilon \geq 0 \}$ of S would be:

$$I_A(S) = \begin{cases} 1 & if\ (S_k - C) \leq \varepsilon \\ 0 & otherwise \end{cases}$$

The indicator function, $I_A(S)$ takes value 1 if the difference of random variables and the capacity is equal or less than ε , otherwise 0. Figure 4.1.5 depicts subset A over the set S. The interesting fact is that the average of $I_A(S)$ represents the frequency that the event $|S_k - C| \leq \varepsilon$ occurs.

$$\bar{I}_A(S) = \frac{1}{j}\sum_{1}^{j} I_A(S_j)$$

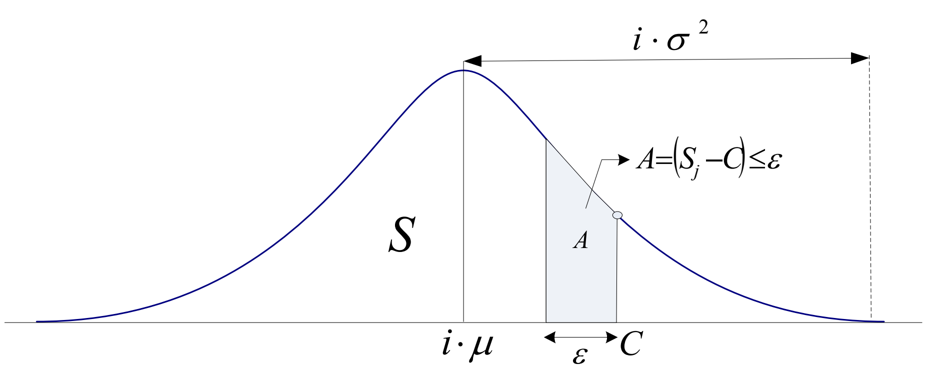


Figure 4.1.5: set *A* over the set *S*

Theorem 4.1.1: let $S_1, S_2, S_3, ..., S_j$ be a collection of the independent and identically distributed random variables, each having finite mean $E\lfloor S_j \rfloor = i \cdot \mu$, $\forall \quad i = 1,2,...,n$ and $j = 1,2,..., \frac{(n-i+1)!}{i!}$, then, for any $\varepsilon > 0$,

$$P\left\{ \left| \frac{1}{j} \sum_{1}^{j} I_A(S_j) - p(A) \right| < \varepsilon \right\} \to 1 \qquad as \quad n \to \infty$$

Proof: the mean and average of the indicator function $I_A(S)$ are:

$$E[I_A(S)] = 1 \cdot P\{(S_j - C) < \varepsilon\} + 0 \cdot (1 - P\{(S_j - C) < \varepsilon\}) = P\{(S_j - C) < \varepsilon\} = P(A)$$

$$\bar{I}_A(S) = \frac{1}{j} \sum_{1}^{j} I_A(S_j)$$

Applying the weak law of large numbers to the indicator function $I_A(S)$ would result in:

$$P\{\bar{I}_A(S) - E[I_A(S)] < \varepsilon\} \to 1 \qquad as \quad j \to \infty$$

This is equivalent to:

$$P\left\{ \left| \frac{1}{j} \sum_{1}^{j} I_A(S_j) - p(A) \right| < \varepsilon \right\} \to 1 \qquad as \quad j \to \infty$$

However, since $j = 1,2,..., \frac{(n-i+1)!}{i!}$, $j \to \infty$ only if $n \to \infty$. This implies:

$$P\left\{ \left| \frac{1}{j} \sum_{1}^{j} I_A(S_j) - p(A) \right| < \varepsilon \right\} \to 1 \qquad as \quad n \to \infty$$

and the proof is complete.

Theorem 4.1.1 states that with probability 1 the frequency of the sum of i consignments being in a given neighborhood of C converges to the true probability of that event as the number of the consignments tends to infinity. This implies that in very large samples if the distribution of the sum of i consignments includes the capacity of the container then with probability 1 there exists at least one cargo consisting of i consignments whose size is very close to the capacity C.

This convergence in probability would look like figure 4.1.6 for a cargo desired to be filled with i consignments. It can be inferred from the figure that when the number of the consignments increases, with probability 1 there exists at least one cargo consisting of i consignments whose size is closer to the capacity than any given ε . For instance,

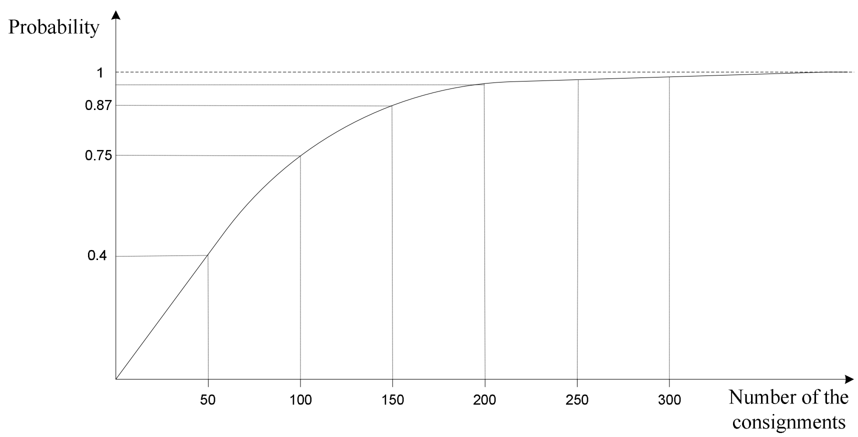


Figure 4.1.6: convergence in probability as number of the consignments grows

Chapter 5

Analysis

In this chapter, mathematical results achieved in the previous chapter are further analyzed. Different scenarios related to cargo loading are proposed and explored in connection with the law of large numbers.

5.1 Introduction

In cargo loading and knapsack problems, many assumptions can be made that affect the outcome of a given problem. These assumptions could be related to the size of the knapsack or container, size of the consignments, number of the consignments etc. These assumptions create different scenarios in cargo loading and knapsack problems while operating in large markets. In the following sections some of these assumptions will be introduced and their effect on cargo loading and knapsack problems will be analyzed in connection with the law of large numbers.

5.2 Large Knapsack and Small Consignments

It is perceived that the size of the individual consignments compared to the size the knapsack would have contrasting impacts on filling performance of the knapsack. In situations where the size of individual consignment is very small compared to the large size of the knapsack, it is conceived that higher filling rates will be achieved. However, when the consignments are bulk and their size is large, lower filling rates will be achieved. In logistics and transportation, envelops, parcels, rolls, and tubes are categorized under small consignments group. Figure 5.1.1 depicts some examples of small consignments used by transportation companies. On the other hand, sizes of knapsacks, which could be containers, truck capacities, etc., are usually large and internationally standardized.

Figure 5.2.1: small consignments [DropShipInfo, 2005]

To analyze the situation where the sizes of the consignments are very small compared to the size of the container, let X_i be the size of the consignment i, and C be the size of the knapsack. This situation, $X_i << C$, is shown in figure 5.1.2.

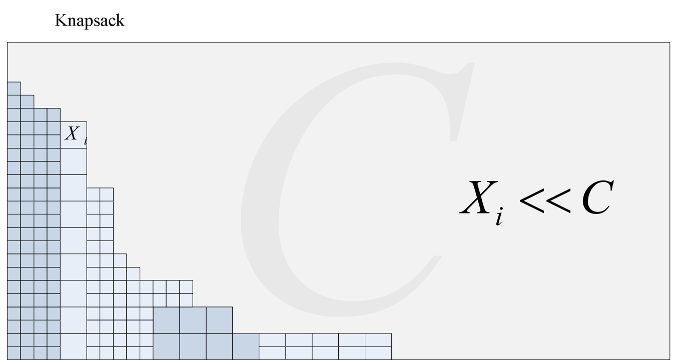


Figure 5.2.2: small consignments and large knapsack

It is clear that in order to fill the entire capacity with small consignments, many consignments must be gathered. This consequently augments the number of the possible combinations of consignments required to create a cargo. Looking mathematically to this situation, the combination of small consignments would have an interesting effect on the aggregated distribution of the consignments. The combination of small consignments would increase the density of the aggregated distribution function in contrast to the same situation with large consignment sizes that create sparse aggregated distribution function. These situations are illustrated in figure 5.1.3.

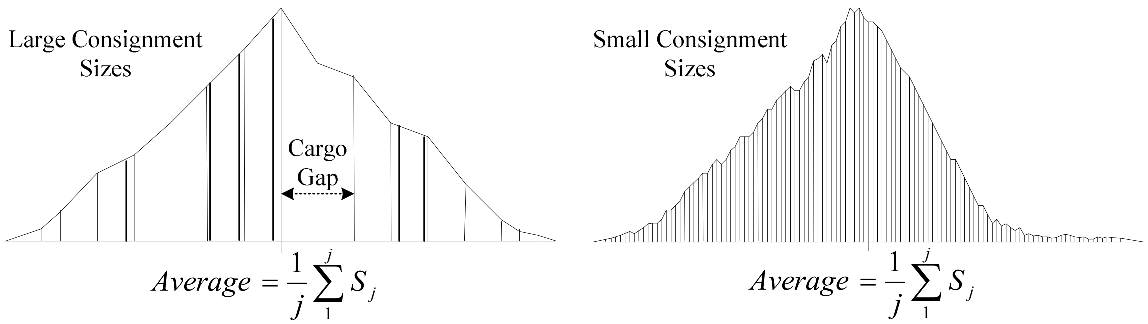


Figure 5.2.3: aggregated distribution of large and small consignment sizes

Considering a knapsack capacity, *C*, it can be inferred from figure 5.1.3 that in a given neighborhood of the capacity, ε more combination of consignments is available and the gap between the cargo and capacity decreases. However having large consignment sizes, larger gap between cargo and capacity can occur and less combination of consignments is available, figure 5.1.4.

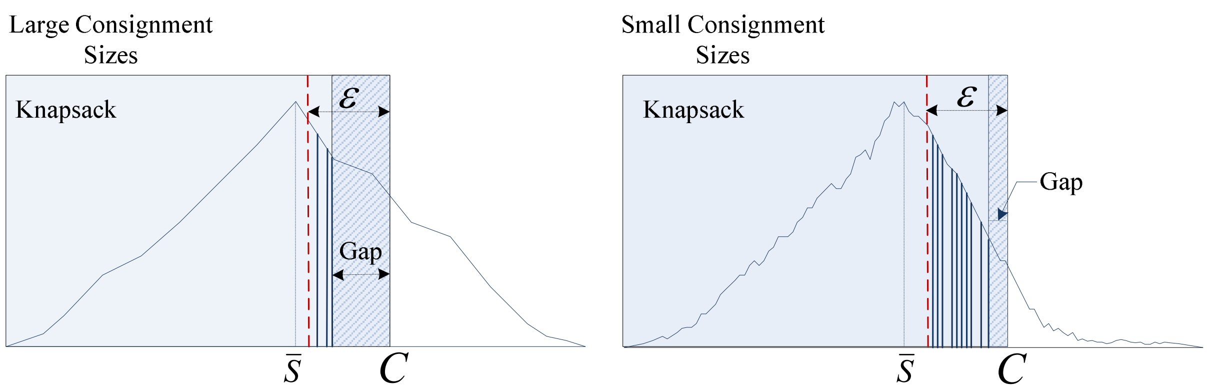


Figure 5.2.4: capacity gap in large and small consignment sizes

As an example, consider a knapsack which is intended to be filled once with large consignments and once with small consignments, figure 5.1.5. When the container is intended to be filled with large consignments, only two consignments are filled and addition of the third one falls out of the capacity. Therefore a large gap between the capacity and the cargo is created. However, when the container is filled with small consignments, smaller gap between the cargo and capacity is perceived.

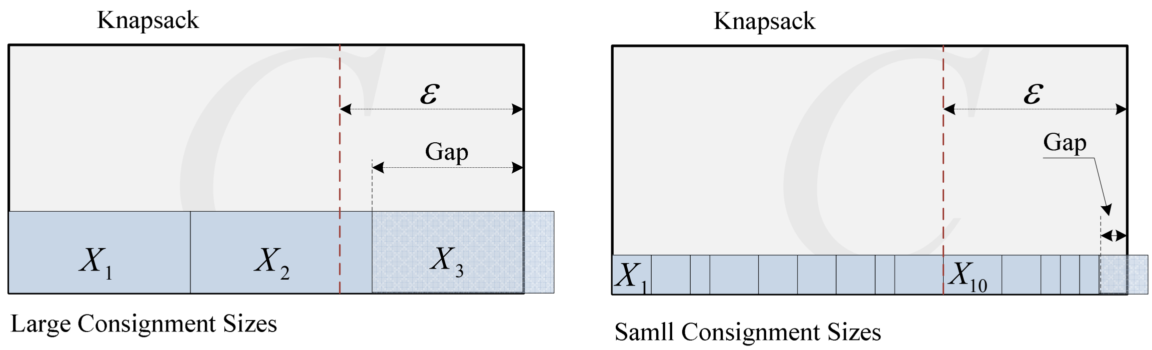


Figure 5.2.5: capacity gap

Therefore, by the law of large numbers, small consignment sizes combined by increased number of consignments would increase the filling performance of the knapsack.

5.3 Standardized Consignments

Another assumption that would affect the filling performance of a knapsack is standardized consignments. Standardized consignments mean that all the consignments in a cargo have the same size. This is because the dimensions of all consignments are standardized by regulations. For instance, Euro pallets that are used by European freight transport in retail business have standard dimensions of 800 x 1200 mm. Standardized boxes, and packages, etc. are also considered as standardized consignments. Figure 5.2.1 depicts the definition of standardized consignments in a knapsack.

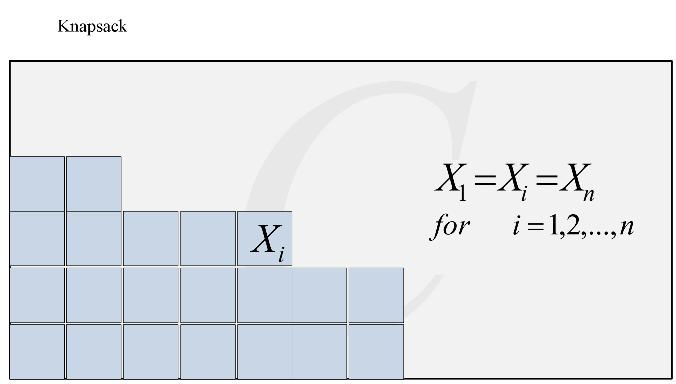


Figure 5.3.1: standardized consignments in a knapsack

Since the sizes of the consignments are equal, the variation in size is zero. This means that with probability 1 the size of the consignments equals a constant X_i and the probability that sizes deviate from X_i would be zero. Therefore, the sum of the

i consignments would also be constant resulting in a discrete constant aggregate distribution function, figure 5.2.2. As it is shown in the figure, the expected value of the sum of i consignments equals the sum of i consignments and the deviation from the mean is zero.

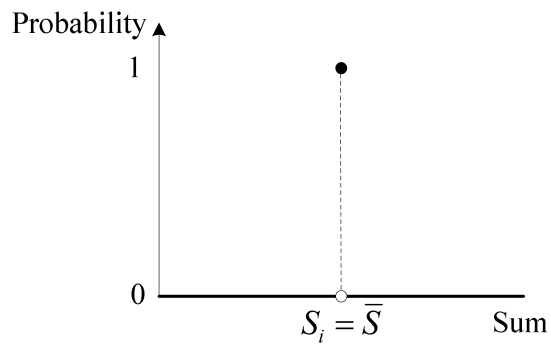


Figure 5.3.2: aggregate distribution for the sum of standardized consignments

Since the size of the cargo increases linearly with the addition of consignments, there would be only one optimum combination of consignments whose size is equal or very close to the capacity of the knapsack. This means that addition of another consignment to the optimum number of consignments, i^*, would deteriorate the capacity constraint. This situation is depicted in figure 5.2.3.

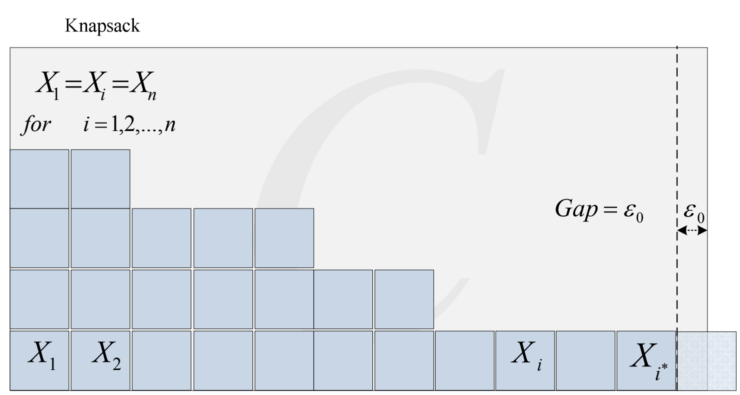


Figure 5.3.3: capacity gap with standardized consignments

It can be inferred from the figure that depending on the size of the consignments a capacity gap of ε_0 would be created between the cargo and knapsack. This capacity gap can not be reduced further since the sizes of the consignments are standardized and addition of smaller consignments is not allowed. As a result capacity gap is always smaller that the size of the consignments. Applying the law of large numbers to this scenario suggests that the capacity gap between the cargo and knapsack would not be reduced further than ε_0 as the number of the consignments grows, figure 5.2.4.

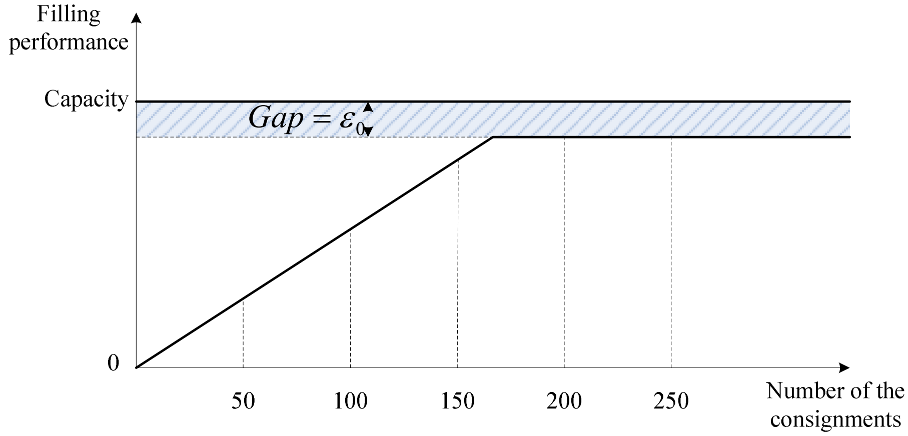


Figure 5.3.4: Filling performance as the number of the consignments grows

Chapter 6

Conclusions and Recommendations

This chapter will present the results of the study proceeding by some conclusions about the thesis. The chapter ends with some important issues that the analysis in this study did not manage to capture.

6.1 Conclusions

The main purpose in this thesis was to prove that the filling performance of vehicles, fleet, etc. operating in the logistics network can be increased while operating in large networks including many customers into the system. The mathematical model of the hypothesis was created using the concept of knapsack problem in operations research. The validity of the hypothesis was then proved using the theory of the law of large numbers. This study showed that when the sample size grows the inferior capacity utilization in the knapsack converges to 0.

The results of the analysis chapter in this thesis showed that incorporating small consignments into the system has the advantage of reducing capacity gap in the knapsack, in contrast to large consignments which increase the capacity gap in the knapsack. In addition, the consequence of incorporating standardized consignment sizes into the system is having a capacity gap equal to ε_0 in the knapsack.

6.2 Recommendations for Future Work

In reality there are many limitations and events that deteriorate the results of this study in real life problems. However, the results of this study from the theoretical point of view is significantly important.

I perceive that simulating real life problems and eliminating assumptions which is used in developing theory would be challenging for future works in this area. Simulating real life scenarios would contribute to importance of the law of large numbers in future logistics operations and reliability of the theory in real life problems. In addition, the theoretical study can be further developed by analyzing the effects of different consignment dimensions in filling performance of vehicles, fleet, etc. while operating in large markets. By this I mean considering the difference in height, width, and length of consignments in the analysis of the filling performance of the knapsack rather that considering the total volume of the consignments in the analysis.

References

DropShipInfo, 2005. *Small Parcels.* [online]. Available from: www.dropshipinfo.co.uk/
[Cited 10 May 2007].

Encyclopædia Britannica. 2007. *Insurance*. [Online]. Available from: http://www.britannica.com/eb/article-9106314
[Cited 20 March 2007].

Engström, R., 2004. *Competition in the Freight Transport Sector – a Channel Perspective*. Göteborg: Bokförlaget BAS.

Grinstead, C. M. and Snell, J. L., 1997. *Introduction to Probability*. 2nd ed. Providence, R.I.: American Mathematical Society.

Hastings, K. J., 1997. *Probability and Statistics*. USA: Addison Wesley Longman, Inc.

Investopedia ULC. 2007. *Law of Large Numbers*. [Online]. Available from: http://www.investopedia.com/terms/l/lawoflargenumbers.asp
[Cited 17 March 2007].

Lambert, D. M. and Stock, J. R., 1993. *Strategic Logistics Management*. 3rd ed. USA: Richard D. Irwin, Inc.

Lumsden, K., 2004. *Logistikens Grunder*. Lund: Studentlitteratur.

Mårtensson, B. and Nilstun, T., 1988. *Praktisk Vetenskapsteori*. Lund: Studentlitteratur.

Mill, J. S., 1961. *Auguste Comte and Positivism*. Ann Arbor, MI: Michigan U.P.

Martello, S. and Toth, P., 1990. *Knapsack Problems: Algorithms and Computer Implementations*. West Sussex, England: John Wiley & Sons, Ltd.

Ross, S. M., 2000. *Introduction to Probability and Statistics for Engineers and Scientists*. 2^{nd} ed. Burlington, MA: Harcourt Academic Press.

Stefansson, G., 2006. *Collaborative Logistics Management and the Role of Third-Party Service Providers.* International Journal of Physical Distribution & Logistics Management, Vol. 36 No. 2, PP. 76-92.

Trochim, W.M.K. 1999. *The Research Methods Knowledge Base*. Ithaca: Cornell University.

Wiedersheim, P. and Eriksson, L. T., 1991. *Att Utreda, Forska och Raportera*. Malmö: Liber Economi.

Yin, R. K., 1994. *Case Study Research – Design and Methods*. London: Sage Publications, Inc.