GALATASARAY UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

LEARNING ENABLED NEWSVENDOR PROBLEM

Gözde KÖYBAŞI

Jul 2020

LEARNING ENABLED NEWSVENDOR PROBLEM

(ÖĞRENME TABANLI GAZETE SATICISI PROBLEMİ)

by

Gözde KÖYBAŞI, B.S.

Thesis

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

in

INDUSTRIAL ENGINEERING

in the

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

of

GALATASARAY UNIVERSITY

Jul 2020

This is to certify that the thesis entitled

LEARNING-ENABLED NEWSVENDOR PROBLEM

prepared by **Gözde KÖYBAŞI** in partial fulfillment of the requirements for the degree of **Master of Science in Industrial Engineering** at the **Galatasaray University** is approved by the

Examining Committee:

Doç. Dr. Sezi ÇEVİK ONAR
Department of Industrial Engineering
İstanbul Technical University -------------------------

Prof. Dr. Temel ÖNCAN
Department of Industrial Engineering
Galatasaray University -------------------------

Doç. Dr. M. Ebru ANGÜN (Supervisor)
Department of Industrial Engineering
Galatasaray University -------------------------

Date: **----------------------**

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Doc. Dr Ebru ANGÜN who has helped tremendously in carrying out this research. Her exceptional guidance, extensive knowledge, and high motivation allow me to keep progressing throughout the study.

I would also like to take this opportunity to declare my profound appreciation to my family for providing full support and inspiration in each and every moment of my life.

Jul 2020
Gözde KÖYBAŞI

TABLE OF CONTENTS

LIST OF SYMBOLS

NV **:** Newsvendor
SAA **:** Sample Average Approximation
MLE **:** Maximum Likelihood Estimator

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The newsvendor model is one of the most popular analytical models in decision science and operations management. The standard newsvendor problem is a single period inventory management problem in which the newsvendor has to decide the optimal stocking quantity for a single product. With the consideration of overstocking and understocking costs, the optimal quantity can be found that minimizes the newsvendor's total expected cost. Because of its simple structure, the newsvendor model has been widely applied to analyze several issues in supply chain systems involving perishable and seasonal products since the mid-1980s.

The standard newsvendor problem assumes the knowledge of the demand distribution so that the optimal stocking quantity is given by the critical fractile. In practice, however, this distribution is unknown; yet, there usually exist enormous historical demand data and demand related data. Then, the complexity of optimal ordering decision not only comes from coping with the ambiguity of demand, but also from incorporating a vast range of demand related information available into the decision process in order to make an enhanced decision.

This research considers a data-driven newsvendor problem for a single product and a single period, where historical data of aggregated demands and attributes that can be used to leverage the demand distribution are available. Distribution of demand is assumed to be unknown except for its support. By modelling probability of buying the product through binary logit model, the demand process is approximated by a binomial process, and the resulting newsvendor problem is solved as a linear programming problem. Within the framework of this research, the robust approach is adopted as a benchmark problem, along with the Sample Average Approximation (SAA) approach. The separate numerical experiments for normal, gamma, and lognormal distributions show similarities with the well-known SAA-based optimization method.

ÖZET

Gazete satıcısı problemi, oldukça geniş bir uygulama alanına sahip olmakla birlikte temelde kolay bozulabilen ürünler için bir envanter yönetimi problemidir. Belirlenecek sipariş miktarı; müşteri talebini karşılayacak derecede yüksek, aynı zamanda gereğinden fazla stok tutulmasının önüne geçecek derecede düşük olmalıdır. Bu durumda elde bulundurmama maliyeti ile envanter maliyeti eş zamanlı olarak dikkate alınmalıdır. Gazete satıcısı problemi de bu iki maliyet türünü dengeleyen bir sipariş miktarı belirlemeyi amaçlar. Problemin literatürdeki çözümü, genellikle talep dağılımının bilindiği varsayımına dayanır. Ancak pratikte karar verici, talep dağılımı bilgisi yerine geçmişe dönük talep verilerini elinde bulundurur. Bu tür bir sipariş miktarı kararının karmaşıklığı yalnızca talep belirsizliğiyle başa çıkmaktan değil, talebe ilişkin mevcut bilgilerin karar sürecine olabildiğince dahil edilmeye çalışılmasından da kaynaklanır. Bu çalışma tek bir ürün ve tek bir dönem için gazete satıcısı problemini ele almakta, talep dağılımını tahmin ederken geçmiş talep ve niteleyici faktör verilerini dikkate almaktadır. Talebin alabileceği maksimum değer dışında, talep dağılımı ile ilgili herhangi bir bilgi olmadığı varsayılmıştır. Ürünü satın alma olasılığının ikili logit modeli kullanılarak modellenmesiyle, talep binom dağılımına yaklaştırılmış ve ortaya çıkan gazete satıcısı problemi doğrusal bir programlama modeli olarak çözülmüştür. Bu çalışma çerçevesinde, denektaşı problemi olarak Örneklem Ortalaması Yakınsama (SAA) yaklaşımının yanı sıra, robust yöntemi de benimsenmiştir. Normal, gama ve lognormal dağılımları için ayrı ayrı yapılan sayısal deneyler, elde edilen sonuçların Örneklem Ortalaması Yakınsama (SAA) yaklaşımı sonuçlarıyla benzerlik taşıdığını göstermiştir.

1. INTRODUCTION

The newsvendor problem, otherwise known as the newsboy problem, is a widely acknowledged problem of stochastic inventory management. Initially introduced by Whitin (1955) to literature, the problem focuses on determining the order quantity of a perishable product that maximizes the expected profit and by all means, minimizes the expected cost simultaneously. Since the ordering is deemed to have completed prior to the actual selling period, the customer demand is uncertain during the ordering process. Considering the fact that an unsold perishable product loses its value in whole or in part over time, the vendor confronts a situation in which there should be a balance between disposal of an unsold product for a salvage value and miss of a sales opportunity. Facing demand uncertainty, the trade-off is defined to be between the risks of overstocking and understocking (Gallego & Moon, 1993). Therefore, the order quantity needs to be not only high enough to meet the entire demand but also low enough to avoid having excessive unsold inventory.

1.1 Classical Newsvendor Problem and Its Extensions

In the classical newsvendor problem, a decision maker plans to sell a product over a single period to meet a stochastic demand D (Zipkin, 2000). Thus, the scope of the problem at its simplest form consists of a random demand of a single perishable product for a single selling period. The decision maker now needs to commit to a non-negative order quantity q before observing the actual demand at the end of the sales period.

The basic assumption of the newsvendor problem is that the demand distribution F is known. It is assumed that uncertain demand D is a random variable with a known probability distribution function. The goal of the decision maker is to choose an order quantity q^* that minimizes the expected total cost, which can be formulated as in Eq. (1):

$$\min_{q \geq 0} E[C(q, D)] \tag{1}$$

The decision maker incurs an under-stocking cost c_u for each unit of unsatisfied demand, and an overstocking cost c_o for each unsold unit of products. The expected cost can be reformulated as the sum of the expected overage (overstocking) cost and the expected underage (understocking) cost. Considering that the overage cost c_o arises when the demand is less than the order quantity, and the underage cost c_u arises if not so, the problem can be represented as in Eq. (2). The following random quantities $(D - q)^+$ and $(q - D)^+$ are the total units of the unsatisfied demand and units of unsold product at the end of the period, respectively shown in Eq. (3.a) and Eq. (3.b).

$$\min_{q \geq 0} E[C(q, D)] = \min_{q \geq 0} c_o\, E[(q - D)^+] + \; c_u\, E[(D - q)^+] \tag{2}$$

where

$$(q - D)^+ = max\, \{q - D, 0\} \tag{3.a}$$

$$(D - q)^+ = max\, \{D - q, 0\} \tag{3.b}$$

The optimal order quantity solution q^* of the classic newsvendor problem is the formulated as in Eq. (4) and Eq. (5), where Eq. (5) is called as critical fractile solution, that are equivalent to each one another.

$$F(q^*) = \frac{c_u}{c_u + c_o} \tag{4}$$

$$q^* = F^{-1}\left(\frac{c_u}{c_u + c_o}\right) \tag{5}$$

where F and F^{-1} represent the cumulative distribution function of demand and its inverse respectively. According to the probability theory, on the other hand, a cumulative distribution function is non-decreasing and right continuous by definition (Limaye, 2004). Due to the fact that a function is not necessarily strictly increasing, not every cumulative distribution function is invertible. Hence, a pseudo-inverse function that is valid under all circumstances is defined (Santambrogio, 2015). In that case, the optimal order quantity q^* is the smallest q that satisfies the condition in Eq. (4), which leads us to obtain Eq. (6).

$$q^* = inf\left\{q: F(q) \geq \frac{c_u}{c_u + c_o}\right\}. \tag{6}$$

The primary problem with the classical newsvendor problem is that the probability distribution of demand is assumed to be known. Nevertheless, F is typically unknown in real-life cases. Besides the absence of sufficient information, it is considered hard to identify the true distribution of the customer demand in an accurate manner. In fact, even if the demand data are presumed to conform to a certain probability distribution, reliable results cannot be acquired since the demand for perishable products is inherently fluctuating most of the time. Owing to this high degree of unforeseeability, data coming from a specific demand distribution can easily be misleading for perishable products (Bertsimas & Thiele, 2005). In order to overcome such complication, distribution free newsvendor models, in which the distribution of the demand remains unknown, are established.

As it is stated previously, in most of the practical cases, F is unknown; yet, there can be some prior knowledge $\mathcal{I}$ about the demand and its distribution such as population mean, variance, range, median, skewness, unimodality, etc., and/or there can exist N demand observations $\mathcal{T}_N = \{d_1, \dots, d_N\}$. As one of the distribution free approaches, the data-driven newsvendor problem is based on the notion that demand data comes from an unknown distribution. In this problem, actual past demand data coming from N different demand observations are used instead as input.

The objective remains unchanged, but the expected cost from the classical newsvendor problem turns into a formulation of an estimated expected cost as in Eq. (7). The new sets of information $\mathcal{I}$ and $\mathcal{T}_N$ are introduced to Eq. (2) to obtain and solve a reasonable substitute

$$\min_{q\geq 0} E[C(q,d) \mid \mathcal{I}] \text{ or } \min_{q\geq 0} E[C(q,d) \mid \mathcal{T}_N] \quad (7)$$

where $\mathcal{T}_N$ is called the training set and the dependence on $\mathcal{T}_N$ is called data-driven.

The estimated expected cost $\widehat{EC}(q,D)$ can be simply redefined as sample average approximation of the traditional expected cost $EC(q)$ as indicated below in Eq. (8).

$$\widehat{EC}(q,d) = \frac{1}{n}\sum_{i=1}^{n} [c_o\,(q-d_i)^+ + c_u\,(d_i-q)^+] \quad (8)$$

where

$$(q-d)^+ = \max\{q-d, 0\} \quad (9.a)$$

$$(d-q)^+ = \max\{d-q, 0\} \quad (9.b)$$

where each of the realization d_n is given an empirical probability of occurrence $1/\mathrm{N}$. The optimal order quantity decision based on sample average approximation of the data-driven newsvendor problem is given in Eq. (10).

$$\hat{q}_n = \inf\left(q\colon \hat{F}_n(q) \geq \frac{c_u}{c_u + c_o}\right) \quad (10)$$

where $\hat{F}_n(q)$ represents the empirical cumulative distribution function of past demand data obtained from N observations.

As a matter of fact, the aforementioned data-driven newsvendor model mostly falls short of reflecting more complex, real-world situations since multiple forms of data regarding demand can be gathered alongside with their value. The decision maker may have the opportunity to benefit from a wider range of information during the ordering decision-making process (Rudin and Vahn, 2019). Demand-related information may include price of the good, season at the time, sales region of the good, sales channel of the good, and even consumer profile regarding both demographic and socioeconomic traits such as age, gender, education, occupation, income.

The featurized newsvendor problem assumes to have a data set of the form $\mathcal{T}_N = \{(d_1, \boldsymbol{x}_1), \dots, (d_N, \boldsymbol{x}_N)\}$, where d_n denotes a realization of the random demand and $\boldsymbol{x}_n$ is the vector of attributes represented as $x_n = [x_n^1, x_n^2, \dots, x_n^K]$ with each containing K attributes, which can be used to leverage demand distribution or order quantity for $n = 1, \dots, N$.

In this study, a newsvendor problem has been constructed by taking features (in other words attributes) of demand into account, similarly to the abovementioned featurized newsvendor problem. For the demand, its conditional probability is modelled by using binary choice modelling, which is explained in the next section.

1.2 Binary Choice Modelling

Discrete choice analysis concerns with the modelling of an individual's selection process, which is required to be made among mutually exclusive alternatives. The analysis is based on the principle that the individual as the decision-maker chooses the alternative that provides her maximum utility. As the utility of each alternative is treated as a random variable, it is assumed that the probability of an alternative to be chosen corresponds to the probability that the relevant alternative has the highest utility compared to remaining alternatives.

In case that a discrete choice model is composed of merely two available alternatives, the model is called as a binary choice model. In the circumstances, the probability of any alternative i to be preferred to any alternative j, both included in the choice set A_n, by the individual n is demonstrated as in Eq. (11).

$$P(i \mid A_n) = P\left(U_{in} \geq U_{jn},\ \forall j \in A_n\right) \tag{11}$$

The probability of choosing alternative j instead can be therefore demonstrated as in Eq. (12).

$$(j \mid A_n) = 1 - P(i \mid A_n) \tag{12}$$

Reminding that U_{in} and U_{jn} are random variables, each utility function can be divided into two main components: deterministic component (V) and random component (ε) as the following.

$$U_{in} = V_{in} + \varepsilon_{in} \tag{13.a}$$

$$U_{jn} = V_{jn} + \varepsilon_{jn} \tag{13.b}$$

V_{in} and V_{jn} are named as systematic, non-random components possibly consisting of deterministic functions, which yield the same results as long as inputs remain the same. ε_{in} and ε_{jn} , on the other hand, are considered random.

Component-wise, the probability of selecting alternative i over alternative j can be reformulated as in Eq. (14).

$$\begin{aligned} P\left(U_{ij} \geq U_{jn}\right) &= P\left(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}\right) \\ &= P\left(\varepsilon_{jn} - \varepsilon_{in} \leq V_{in} - V_{jn}\right) \end{aligned} \tag{14}$$

By investigating the new version of the probability formulation of alternative i selection from above, it can be deducted that the significant point is whether the difference between random components is smaller than the one between deterministic components or not, rather than the numerical values of these components. Hence, binary choice models can be set up by designating purely the differences and neglecting separate component values.

Deterministic components can be identified by a vector of attributes x and a vector of M unknown regression coefficients, namely β, where $\beta^T = [\beta_1, \beta_2, \dots, \beta_K]$. Most of the time, deterministic functions are established in the manner that they can be linear in vector of parameters β in order to ease the estimation of parameters.

Given the sub-components, deterministic components can be specified as follows.

$$V_{in} = V(x_{in}) = \beta^T x_{in} = \beta^1 x_{in1} + \beta^2 x_{in2} + \dots + \beta^K x_{inK} \qquad (15.a)$$

$$V_{jn} = V(x_{jn}) = \beta^T x_{jn} = \beta^1 x_{jn1} + \beta^2 x_{jn2} + \dots + \beta^K x_{jnK} \qquad (15.b)$$

where β^T stands for transpose of vector of parameters while x_{in} and x_{jn} represent vectors of attributes for alternatives i and j respectively, under the acceptance that non-random components are emphatically linear in variables and parameters.

As regards to random components ε_{in} and ε_{jn} that represent the combination of all unobserved attributes, different distribution assumptions result in different binary choice models. Widely used binary choice models include the linear probability model, the binary probit, and the binary logit models.

Before examining the most frequently used distribution assumptions, let us define a random variable $\varepsilon_n = \varepsilon_{jn} - \varepsilon_{in}$, which brings about a reformulation of the aforesaid probability function of choosing alternative i over j.

$$P\left(U_{ij} \geq U_{jn}\right) = P\left(\varepsilon_n \leq V_{in} - V_{jn}\right) \qquad (16)$$

In the linear probability model, it is simply assumed that the difference of random components $\varepsilon_n = \varepsilon_{jn} - \varepsilon_{in}$ is uniformly distributed along with the random components themselves. Its density function $f(\varepsilon_n)$ is defined within a range from $-M$ to $+M$, with $M > 0$. The probability function is obtained by the cumulative distribution function of ε_n as indicated in Eq. (17).

$$P\left(U_{ij} \geq U_{jn}\right) = \begin{cases} 0, \quad if\ V_{in} - V_{jn} < -M \\ \int_{-M}^{V_{in}-V_{jn}} f(\varepsilon_n) d\varepsilon_n = \dfrac{V_{in} - V_{jn} + M}{2M}, if -M \leq V_{in} - V_{jn} \leq M \\ 1, \quad if\ V_{in} - V_{jn} > M \end{cases} \tag{17}$$

At this point, it is important to mention that numerical value of the parameter M solely has influence on the scale of the utility function, so it can be arbitrarily determined.

The binary probit model is based on the assumption that distributions of random components approach a normal distribution by using the central limit theorem. Both components of ε_{in} and ε_{jn} conform to a normal distribution with means zero and standard deviations of σ_i and σ_j, respectively. In that case, the difference ε_n follows a normal distribution with mean zero and a variance of σ with $\sigma > 0$, as shown in Eq. (18).

$$\varepsilon_n \approx N(0, \sigma^2) \approx N\left(0, \sigma_i^2 + \sigma_j^2 - 2\sigma_{ij}\right) \tag{18}$$

where σ_{ij} represents covariance between alternatives I and j.

Recalling the probability density function of normal distribution by Eq. (19.a), the probability function $P\left(U_{ij} \geq U_{jn}\right)$ is acquired by ε_n's cumulative distribution function of standard normal distribution Φ as simplified in Eq. (19.b).

$$P\left(U_{ij} \geq U_{jn}\right) = \int_{-\infty}^{V_{in}-V_{jn}} \frac{1}{\sqrt{2\pi\sigma}}\, e^{\frac{-1}{2}\left(\frac{\varepsilon}{\sigma}\right)^2} d\epsilon \tag{19.a}$$

$$= \Phi\left(\frac{V_{in} - V_{jn}}{\sigma}\right) = \Phi\left(\frac{\beta^T(x_{in} - x_{jn})}{\sigma}\right) \tag{19.b}$$

It is well-known that Eq. (19.a) has no explicit mathematical formulation. This time, the utility function is scaled according to reciprocal of variance $\frac{1}{\sigma}$, which can take an arbitrary value.

Lastly, in the binary logit model, it is accepted that ε_n follows a logistic distribution. In addition to the logistic cumulative distribution function $F(\varepsilon_N)$ in Eq. (20), logistic density function $f(\varepsilon_N)$ is recollected in Eq. (21).

$$F(\varepsilon_n) = \frac{1}{1 + e^{-\mu\varepsilon_n}} \quad \mu > 0, -\infty < \varepsilon_n < +\infty \tag{20}$$

$$f(\varepsilon_n) = \frac{\mu e^{-\mu\varepsilon_n}}{(1 + e^{-\mu\varepsilon_n})^2} \tag{21}$$

Supposing that random components are logistically distributed, the related probability function for preferring alternative i to alternative j can be demonstrated as in Eq. (22).

$$P\left(U_{ij} \geq U_{jn}\right) = \frac{1}{1 + e^{-\mu\left(V_{in} - V_{jn}\right)}} = \frac{e^{\mu V_{in}}}{e^{\mu V_{in}} + e^{\mu V_{jn}}} \tag{22}$$

In the event of linearity of deterministic functions, the choice probability function can be presented in the following.

$$P\left(U_{ij} \geq U_{jn}\right) = \frac{e^{\mu\beta^T x_{in}}}{e^{\mu\beta^T x_{in}} + e^{\mu\beta^T x_{jn}}} \tag{23}$$

Since components' differences are the main concern in the sense of finding probabilities of selection, the selection probabilities remain unchanged in case of an addition of a constant figure to both utility functions U_{in} and U_{jn}. Within this context, any random components' means μ that are not equal to zero are consumed by deterministic components of the utility function. Therefore, there exist no discrepancy between altering the means of random components and adding (or subtracting) the equal quantity from deterministic component. From this point of view, it is worth emphasizing that the mean parameter μ cannot be evaluated separately from the deterministic parameter β in terms of scaling.

Within the scope of this study, the probability of selling the product (therefore the demand) has been modeled using a binary logit model.

1.3 Estimation Methods

Suppose that the data set $\mathcal{T}_N$ is available. Then, there are two estimation techniques for the unknown vector of regression parameters βin Eq. (15a) and (15b). These techniques are maximum likelihood estimation and and least squares estimation. (Akiva & Lerman, 1985).

The likelihood function basically shows how probable for the sample data to be observed. Hence, the maximum likelihood estimator (MLE) corresponds to the parameter value that gives the highest probability of occurrence of the sample data. The general form of a likelihood function L regarding a sample data of size N is in Eq. (24).

$$L(\boldsymbol{\beta}) = \prod_{n=1}^{N} f(\boldsymbol{y}_n | \, \beta, \boldsymbol{x}_n) \tag{24}$$

Afterwards, the likelihood function or its natural logarithm (i.e., log-likelihood function) is maximized over the feasible parameter values.

On the other hand, if a linear regression model is fitted to the sample data, least squares estimation is mostly preferred. Least square estimators are computed by finding the parameter value that gives the minimum sum of squares of differences between the expected values and the sample values. Its general formulation is given in Eq. (25).

$$f = \sum_{n=1}^{N}(y_n - E[(\boldsymbol{y}_n | \beta, \boldsymbol{x}_n)])^2 \quad (25)$$

where $E[(\boldsymbol{y}_n | \beta, \boldsymbol{x}_n)]$ represents the expected value of y_n, which depends on both parameters β and $\boldsymbol{x}_n$. Later on, a solution is found for β by minimizing the function in (25).

In this study, it is preferred to adopt the maximum likelihood estimation function, more specifically log-likelihood function, while estimating the unknown coefficients vector of choice probability function in a manner to maximize the likelihood of the historical demand data.

To sum up, similar to the featurized newsvendor problem, we assume to have $\mathcal{T}_N = \{(d_1, \boldsymbol{x}_1), \dots, (d_N, \boldsymbol{x}_N)\}$ at the beginning, which we use to leverage the demand distribution through a logit model for a single product and a single period newsvendor model. Afterwards, the acquired integer programming problem is solved to identify the optimum ordering decision.

We obtain the estimated order quantities for each level of critical ratio selected and compare these quantities with the ones obtained from various benchmark problems, namely the robust newsvendor as in Gallego & Moon (1993) and the SAA solutions. We find that our procedure gives very similar results to those of the SAA, and the order quantities are usually higher than the true order quantities.

The remainder of this paper can be summarized as follows. The literature review, regarding the treatment of different newsvendor problem types and different solution approaches proposed for them, is introduced in Section 2. Section 3 is divided into two subsections, where Subsection 1 introduces the demand learning process and Subsection 2 demonstrates the optimization problem. Section 4 presents the two benchmark problems, and the numerical results are given for normally, gamma, and lognormally distributed demand data separately in Section 5. The last section summarizes the findings of the study.

2. LITERATURE REVIEW

In recognition of perishable inventory management, the newsvendor problem is still being worked on intensely even though the widely recognized problem has been enquired for such a long time. Over the course of many years, the simple version has been expanded using various forms, including a data-driven approach (Janssen et al., 2016). Various approaches to solve the problem defined in Eq. (7) have been adopted in the literature; the prominent studies are summarized below.

Robust optimization addresses the distribution uncertainty by allowing the distribution to belong to a specified class of distributions, and applies a min-max approach to minimize the worst-case (i.e., maximum) expected cost over the allowed class of distributions. It is usually considered as a highly conservative approach since it takes the most unfavorable distribution of the demand into account while determining the order quantity. Scarf (1958), Gallego & Moon (1993), and Natarajan et al. (2008) all derive the min-max order quantity, which maximizes the worst-case expected newsvendor profit over a class of distributions for which population mean and variance of the demand are assumed to be known in the works of Scarf (1958) and Gallego & Moon (1993), and semi-variance (used as a measure of asymmetry) in addition to population mean and variance is assumed to be known in the paper of Natarajan et al. (2008). Gallego & Moon (1993) start by identifying a lower bound for the expected cost with the intention of pointing out the effect of parameters on the worst-case expected cost. They also expand Scarf's study by adding a multitude of distinct cases. A recourse case, in which making an order for the second time subsequent to an observation of the true demand becomes an option, a fixed ordering cost case, in which both fixed cost and initial level of inventory considerations are introduced, a random yield case, in which a notion of ordering q units of product randomly brings out $G(q)$ good units is promoted, and finally a multi-product case, in which more than one product is to be ordered under budget constraints are all included in the study. By combining a measure of asymmetry to demand distribution, the work of Natarajan et al. (2008) differentiates itself from those of Scarf (1958) and Gallego & Moon (1993). Another expression of a lower bound for the expected cost is established, and the optimal solution is revealed to be less conservative than the solution of Scarf approach.

Furthermore, Rahimian et al. (2019) considers the minimization of the worst-case expected cost over a class of distributions formed by those which are within a certain variation distance from a nominal distribution that is chosen as a reference distribution. Another robust approach is to minimize the worst-case regret over a class of distributions, where the regret is formulated as the opportunity cost from not making the optimal decision in Perakis and Roels (2008). The last paper assumes the presence of knowledge of some properties regarding the distribution's shape such as range, unimodality, skewness to define the class of distributions.

Another line of research which considers a data set $\mathcal{T}_N$ is given in Levi et al. (2007) and Levi et al. (2015). Levi et al. (2007) derives sampling-based optimal policies based on $\mathcal{T}_N$, and also establishes bounds on the number of samples required to guarantee that with high probability, the expected cost of the sample-based policies is sufficiently close to the expected cost of the true optimal policies. To solve the problem that is studied over both classical single period and multi period, SAA approach is adopted. Moreover, another study of Levi et al. (2015) improves the bounds of the study of Levi et al. (2007). These improved bounds allow to indicate the influence of the distribution's mean range on the SAA result's quality of approaching the true demand value.

A further related line of research is called the "featurized newsvendor problem" treated in the works of He et al. (2012), Green et al. (2013), and Rudin & Vahn (2019). He et al. considered a nurse staffing decision in which number of staff needed to be assigned in each specialty is modelled using linear regression under the assumption that it is normally distributed. Green et al. (2013), however, approximates the absenteeism probability through a logit regression model while dealing with the nurse staffing problem. Hence, both demand and supply ambiguity are integrated within the study, and additionally it involves a recourse option allowing to reregulate the supply after observation of the demand as in reordering. Rudin & Vahn (2019) approximates the order quantity through linear decision rule (i.e., a linear regression model), and solves the resulting problem through the Sample Average Approximation (SAA) method. A very similar problem to Rudin & Vahn (2019), namely quantile regression is dealt with in Huber et al. (2019) for the combination of demand estimation and inventory optimization.

Literature review of methodology is presented in Table 2.1 with the inclusion of authors, dates on which stated articles are published, considered uncertain parameter, related methodologies, and known parameter(s).

Table 2.1: Literature comparison

Author	NV Problem Type	Uncertain Factor	Given Factor(s)	Methodology
Scarf (1958)	Distribution Free NV	Demand Distribution	Distribution Mean and Variance	Min-max Approach
Gallego & Moon (1993)	Distribution Free NV	Demand Distribution	Distribution Mean and Variance	Min-max Approach
Levi et al. (2007)	Data-driven NV	Demand Distribution	Data Samples from the Distribution	SAA
Natarajan et al. (2008)	Distribution Free NV	Demand Distribution	Mean, Variance, and Semi-variance	Min-max Approach
Perakis & Roels (2008)	Distribution Free NV	Demand Distribution	Distribution Mean, Variance, Range, Unimodality, Skewness	Min-max Regret Approach
He et al. (2012)	Featurized NV	Demand Distribution	Feature based Demand	Linear Regression

			Data	
Green et al. (2013)	Featurized NV	Supply and Demand Distribution	Feature based Demand Data	Logit Regression
Levi et al. (2015)	Data-driven NV	Demand Distribution	Data Samples from the Distribution	SAA
Huber et al. (2019)	Featurized NV	Demand Distribution	Feature based Demand Data	Quantile Regression
Rahimian et al. (2019)	Distribution Free NV	Demand Distribution	Variation Distance from a Nominal Distribution	Min-max Approach
Rudin & Vahn (2019)	Featurized NV	Demand Distribution	Feature based Demand Data	SAA

3. PROBLEM FORMULATION

In the following two subsections, we consider a data-driven newsvendor problem, where the demand distribution is unknown except its support given by $\{0, \dots, D_{max}\}$. However, there are historical data of aggregated demands and attribute vectors that can be used to leverage the demand distribution. Let $\mathcal{T}_N = \{(d_1, \boldsymbol{x}_1), \dots, (d_N, \boldsymbol{x}_N)\}$ denote this set of training data, where d_n denotes the aggregated demand corresponding to the attribute vector $\boldsymbol{x}_n$ for $n = 1, \dots, N$. Then, our newsvendor formulation initially defined in Eq. (2) becomes

$$\min_{q \geq 0} E[c_u(D - q)^+ + c_o(q - D)^+] \cong \\ \min_{\hat{q}_M \geq 0} \sum_{m=1}^{M} \widehat{w}_m(\hat{\beta}_N^m, \boldsymbol{x}') \, [c_u(d_m - q)^+ + c_o(q - d_m)^+] \tag{26}$$

where $\mathcal{V}_M$ is a set of sampled data which does not overlap with $\mathcal{T}_N$, $\boldsymbol{x}'$ is the current vector of attribute available before the ordering decision q is made, and $\widehat{w}_m(\hat{\beta}_N^m, \boldsymbol{x}')$ is an estimate of the conditional distribution of demands given $\boldsymbol{x}'$. The remaining notation in Eq.(26) is the same as in Eq.(2). Similar formulations to Eq. (26) for general stochastic optimization problems are given in Bertsimas and Kallus (2019), where the conditional probabilities are estimated through different machine learning techniques such as local regression, regression trees, and random forests. Moreover, the formulation on the right-hand-side of Eq.(26) is similar to the Sample Average Approximation solution of the newsvendor problem in, e.g., Ban and Rudin (2019) except that $\widehat{w}_m(\hat{\beta}_N^m, \boldsymbol{x}')$ is not $1/M$ for every realization m.

Visual illustration of our open-loop integrated proposed methodology is indicated in Fig. 3.1.

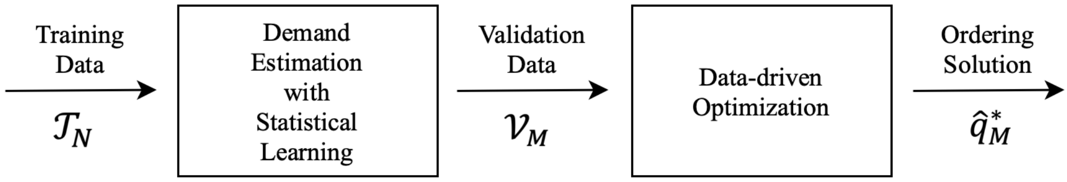

Fig. 3.1: Proposed methodology process

The initial step is the generation of demand related training data $\mathcal{T}_N$ by using normal, gamma, and lognormal distributions, as it can be seen in each subsection of *Numerical Examples.* In the subsection named Statistical Learning, we detail how we obtain the random sample $\mathcal{V}_M$ and how we estimate $\hat{w}_m(\hat{\beta}_N^m, \boldsymbol{x}')$. Finally, in the subsequent subsection named Optimization, the formulation Eq. (26) is minimized to find an estimate $\hat{q}_M$ of the ordering decision.

3.1 Statistical Learning

For simplicity, we assume that there are two alternatives substitutable alternatives g_0 and g_1, where g_0 and g_1 correspond to "buy-nothing" and "buy-product" alternatives, respectively. Each customer, independent of the others, maximizes her utility while selecting between g_0 and g_1; i.e., if her utility of g_1 exceeds the one of g_0, then she prefers g_1, and vice versa. In particular, it is assumed that the customer's demand for g_1 is given by the following logit model

$$p(g_1|\boldsymbol{x}_n) = \frac{e^{\beta^T x_n}}{(1+e^{\beta^T x_n})} \quad n = 1, \ldots, N \tag{27}$$

where $p(g_1|\boldsymbol{x}_n)$ is the conditional probability of selecting g_1 given the attribute vector $\boldsymbol{x}_n = (x_{n1}, \ldots, x_{nK})^T$, T denotes the transpose of a vector or a matrix, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)^T$ is the vector of unknown coefficients. Furthermore, the terms $\boldsymbol{\beta}^T \boldsymbol{x}_n$ and the zero in $e^0 = 1$ in Eq. (27) are the linear-in-parameters utilities of selecting g_1 and g_0, respectively. The conditional probability of selecting g_0 given $\boldsymbol{x}_n$ is simply given by $p(g_0|\boldsymbol{x}_n) = 1 - p(g_1|\boldsymbol{x}_n)$. Akiva & Lerman (1985) is a general reference for discrete choice models such as the one in Eq. (27).

It is intended to estimate the $\boldsymbol{\beta}$ so as to maximize the likelihood of the training set $\mathcal{T}_N$. Note that because of the independence of the customers and Eq.(27), the demand process is approximated by a binomial process for a given $\boldsymbol{x}_n$. Hence, the likelihood function is given by

$$L(\boldsymbol{\beta}) = \prod_{n=1}^{N} \binom{D_{max}}{d_n} p(g_1|\boldsymbol{x}_n)^{d_n} p(g_0|\boldsymbol{x}_n)^{D_{max}-d_n} \tag{28.a}$$

$$= \prod_{n=1}^{N} \binom{D_{max}}{d_n} \left(\frac{e^{\beta^T x_n}}{1+e^{\beta^T x_n}}\right)^{d_n} \left(\frac{1}{1+e^{\beta^T x_n}}\right)^{D_{max}-d_n} \tag{28.b}$$

and the log-likelihood function $l(\boldsymbol{\beta}) = \ln\,[L(\boldsymbol{\beta})]$, where ln is the natural logarithm, is given by

$$l(\boldsymbol{\beta}) = \sum_{n=1}^{N} \left[\tau + d_n \boldsymbol{\beta}^T \boldsymbol{x}_n - D_{max}\; ln(1 + e^{\beta^T x_n})\right] \tag{29}$$

where τ is the constant given by $\tau = \ln(D_{max}!) - [\ln(d_n!) + \ln\,((D_{max} - d_n)!)]$. Then, the maximum likelihood estimator $\widehat{\boldsymbol{\beta}}_N$ is given by

$$\widehat{\boldsymbol{\beta}}_N = \underset{\boldsymbol{\beta}}{\operatorname{argmax}}\, l(\boldsymbol{\beta}). \tag{30}$$

Corollary 1. The $\widehat{\boldsymbol{\beta}}_N$ in Eq. (30) is estimated by solving the following system of nonlinear equations:

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_k} = \sum_{n=1}^{N} \left[d_n x_{nk} - D_{max} \frac{x_{nk} e^{\beta^T x_n}}{1+e^{\beta^T x_n}}\right] = 0 \;\; for \;\; k = 1, \ldots, K \tag{31}$$

Proof. We prove that the log-likelihood function $l(\boldsymbol{\beta})$ in Eq. (29) is concave so that an optimal $\widehat{\boldsymbol{\beta}}_N$ in Eq. (30) is found by solving the Karush-Kuhn-Tucker first-order necessary conditions in Eq. (31). The function $-\ln(1+e^{y})$ is known to be concave and $\boldsymbol{\beta}^T\boldsymbol{x}_n$ is an affine substitution for y, which preserves concavity. Furthermore, $\boldsymbol{\beta}^T\boldsymbol{x}_n$ is affine, and the summation in Eq. (29) for $d_n > 0$ and $D_{max} > 0$ corresponds to taking the conic combination of concave functions, which again results in a concave function. Hence, $l(\boldsymbol{\beta})$ is concave, and Corollary 1 follows.

To obtain the random sample $\mathcal{V}_M$, we proceed as follows. It is known that the maximum likelihood estimator $\widehat{\boldsymbol{\beta}}_N$ in Eq.(30) has asymptotically a multivariate normal distribution with mean vector $\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{\Sigma}$, which is given by the Cramér-Rao bound (Akiva & Lerman, 1985); i.e., this $\boldsymbol{\Sigma}$ is given by $\boldsymbol{\Sigma} = \boldsymbol{B}^{-1}$, where $\boldsymbol{B}$ is an $K \times K$ matrix with components.

$$B_{kk} = -E\left[\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_k^2}\right] = D_{max}\sum_{n=1}^{N} x_{nk}^2 E[\hat{p}_N(g_1|\boldsymbol{x}_n)\hat{p}_N(g_0|\boldsymbol{x}_n)] \quad for\ k = 1,\dots,K \tag{32.a}$$

$$B_{kj} = -E\left[\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_k \partial \beta_j}\right] = D_{max}\sum_{n=1}^{N} x_{nk}x_{nj} E[\hat{p}_N(g_1|\boldsymbol{x}_n)\hat{p}_N(g_0|\boldsymbol{x}_n)] \quad for\ k,\ j = 1,\dots,K \quad k \neq j \tag{32.b}$$

where the $\hat{p}_N$ are obtained by replacing the $\boldsymbol{\beta}$ by its estimator $\widehat{\boldsymbol{\beta}}_N$ in Eq. (30). In particular, if there is a scalar attribute x_n (i.e., $K = 1$), then the Cramér-Rao bound for the variance σ^2 of $\hat{\beta}_N$ simplifies to

$$\sigma^2 \geq \frac{1}{-E\left[\frac{d^2 l(\beta)}{d\beta^2}\right]} = \frac{1}{D_{max}\sum_{n=1}^{N} x_n^2 E[\hat{p}_N(g_1|\boldsymbol{x}_n)\hat{p}_N(g_0|\boldsymbol{x}_n)]}. \tag{33}$$

In the rest of this paper, we assume a scalar attribute x_n. The maximum likelihood estimators can be biased, and the lower bound on the right-hand-side of Eq. (33) is only achieved as $N \to \infty$. Nevertheless, we estimate the mean by its maximum likelihood estimate and the variance by the right-hand-side of Eq. (33), where we ignore the dependence between $\hat{p}_N(g_1|\boldsymbol{x}_n)$ and $\hat{p}_N(g_0|\boldsymbol{x}_n)$; i.e.,

$$\hat{\mu}_N = \hat{\beta}_N, \hat{\sigma}_N^2 = \frac{1}{D_{max}\sum_{n=1}^{N} x_n^2 \frac{e^{\hat{\beta}_N x_n}}{(1+e^{\hat{\beta}_N x_n})^2}}. \tag{34}$$

Let $\hat{\beta}_N^m$ for $m = 1, \dots, M$ be a sample from the normal distribution with the parameters given in Eq.(34) and $\hat{p}_N^m(g_1|x')$ be its corresponding conditional probability of selecting g_1; i.e.,

$$\hat{p}_N^m(g_1|x') = \frac{e^{\hat{\beta}_N^m x'}}{1+e^{\hat{\beta}_N^m x'}}. \tag{35}$$

To form $\mathcal{V}_M$, a realization d'_m is sampled with replacement from $\{0, 1, \dots, D_{max}\}$ with respect to the binomial distribution with parameters D_{max} and $\hat{p}_N^m(g_1|x')$. Then, the normalized weights $\widehat{w}_m(\hat{\beta}_N^m, \boldsymbol{x}')$ are found by

$$\widehat{w}_m(\hat{\beta}_N^m, \boldsymbol{x}') = \frac{\binom{D_{max}}{d'_m} (\hat{p}_N^m(g_1|x'))^{d'_m} (\hat{p}_N^m(g_0|x'))^{D_{max}-d'_m}}{\sum_{r=1}^{M} \binom{D_{max}}{d'_r} (\hat{p}_N^r(g_1|x'))^{d'_r} (\hat{p}_N^r(g_0|x'))^{D_{max}-d'_r}} \tag{36}$$

for $m = 1, \dots, M$. Therefore, the set $\mathcal{V}_M$ is formed by $\mathcal{V}_M = \{(d'_1, \widehat{w}_1(\hat{\beta}_N^1, \boldsymbol{x}')), \dots, (d'_M, \widehat{w}_M(\hat{\beta}_N^M, \boldsymbol{x}'))\}$, and this set $\mathcal{V}_M$ is used to find an estimate $\hat{q}_M$ of the order quantity in the next subsection.

3.2 Optimization

Given $\mathcal{V}_M$, the problem given in Eq. (26) can be written as the following linear programming problem:

$$\min_{\hat{q}_M, u_m, o_m \geq 0} \sum_{m=1}^{M} \widehat{w}_m(\hat{\beta}_N^m, x')[c_u u_m + c_o o_m] \tag{37}$$

$$\text{s.t. } u_m \geq d'_m - \hat{q}_M \;\; m = 1, \dots, M$$

$$o_m \geq \hat{q}_M - d'_m \;\; m = 1, \dots, M.$$

Since the demands are discrete random variables, the problem in Eq. (37) is in fact an integer programming problem. However, for a big M, solving a linear programming problem instead of an integer one is computationally more efficient. After solving Eq. (37), the estimated optimum order quantity can be rounded upwards to the nearest integer to obtain $\hat{q}_M^* = \lceil \hat{q}_M \rceil$.

4. BENCHMARK PROBLEMS

We consider the following two benchmark problems plus the classic newsvendor problem for which the optimization problem and the optimum order quantity are given in Eq. (2) and Eq. (6), respectively.

4.1 Robust Solution of the Newsvendor Problem

In this problem, it is assumed that the population mean μ and variance σ^2 of the demands are known, and they are the only information on the distribution of demands. Their interpretation of the problem is formulated as in Eq. (38)

$$\min_{q\geq 0}\max_{F\in\mathcal{F}} E[c_u(D-q)^+ + c_o(q-D)^+|\mathcal{J}] \tag{38}$$

where $\mathcal{F}$ is a class of distributions for which the mean and the variance are equal to μ and σ^2, respectively, and $\mathcal{J}$ is the given set of information (i.e., the knowledge of μ and σ^2). By adopting a min-max approach, the decision maker determines an order quantity in a way that minimizes the expected cost for the worst distribution of the class, in other words the distribution with the highest expected cost among the distributions with the mean μ and the variance σ^2. We assume that $c_u + c_o$ and c_o are unit selling price and unit acquisition cost, respectively, and that there is no salvage value. Furthermore, let $s = \left(\frac{c_u+c_o}{c_o}\right) - 1$ be the mark up. Initially introduced by Scarf (1958) to the literature and extended by Gallego & Moon (1993) later on, the definitive Scarf's ordering rule which solves Eq. (38) is given by

$$q^{*S} = \mu + \frac{\sigma}{2}[s^{\frac{1}{2}} - (\frac{1}{s})^{1/2}] \tag{39}$$

provided that $s \geq \left(\frac{\sigma}{\mu}\right)^2$, and $q^{*S} = 0$ otherwise.

4.2 Sample Average Approximation Solution of the Newsvendor Problem

As one of the most inherently prominent data-enabled heuristic techniques, the Sample Average Approximation allows to minimize the expected cost by obtaining a numerical average over the unknown demand distribution, presented by a random sample of size M.

For this problem, we assume that the random sample $\mathcal{V}_M$ in Section 3 is given only by the demand realizations; i.e., $\mathcal{V}_M = \{d_1', \dots, d_M'\}$. Now, the problem in Eq. (26) is rewritten as

$$\min_{q\geq 0} \hat{C}_M\,(q) = \frac{1}{M} \min_{q\geq 0} \sum_{m=1}^{M} [c_u(d_m - q)^+ + c_o(q - d_m)^+] \qquad (40)$$

where the only difference is $\widehat{w}_m(\hat{\beta}_N^m, \boldsymbol{x}') = 1/M$ for all $m = 1, \dots, M$. Under the assumption that each demand has an equal probability of realization, an estimated optimum order quantity $\hat{q}_M^{*SA}$ can be found by solving the linear programing problem in (9) after replacing all normalized weights $\widehat{w}_m(\hat{\beta}_N^m, \boldsymbol{x}')$ by $\frac{1}{M}$. It is worth emphasizing that the solution of the SAA heuristic $\hat{q}_M^{*SA}$ is non-deterministic as the order quantity acquired comes from a random sample of demand (Levi et al., 2015).

5. NUMERICAL EXAMPLES

In this section, we compare the optimal order quantities of the classic newsvendor problem q^* in Eq.(6), our proposed $\hat{q}_M^*$, the Scarf's ordering rule q^{*S}, and the Sample Average Approximation solution $\hat{q}_M^{*SA}$ for different levels of the critical ratio $\frac{c_u}{c_u+c_o}$. We set the overstocking cost to c_o =1 and change the understocking cost c_u such that the critical ratios are from 0.2 to 0.8. For the Scarf's ordering rule q^{*S} in Eq.(39), the mark ups corresponding to the critical ratios are: $s = 0.25$ for $\frac{c_u}{c_u+c_0} = 0.2$, $s = 0.43$ for $\frac{c_u}{c_u+c_0} = 0.3$, $s = 0.67$ for $\frac{c_u}{c_u+c_0} = 0.4$, $s = 1$ for $\frac{c_u}{c_u+c_0} = 0.5$, $s = 1.5$ for $\frac{c_u}{c_u+c_0} = 0.6$, $s = 2.33$ for $\frac{c_u}{c_u+c_0} = 0.7$, and $s = 4$ for $\frac{c_u}{c_u+c_0} = 0.8$.

We use the sample size $M = 1000$, where only $\hat{q}_M^*$ and $\hat{q}_M^{*SA}$ depend on M. The demands are generated through normal, gamma, and lognormal distributions, but the newsvendor observes only the histograms in Fig. 5.1, Fig. 5.3 and Fig. 5.5, where the scalar attribute x is the price. Moreover, the price for the next period is $x' = 115$ monetary units, the maximum demand is $D_{max} = 100$, and $N = 10$.

5.1 Normal Distributed Demands

For this subsection, the demands are generated through a normal distribution with means given by

$$\mu_n = D_{max} - 0.4x_n = 100 - 0.4x_n \quad (41)$$

where the x_n's $n = 1, \ldots, 10$ are the prices corresponding to the observed demands in Fig. 5.1; i.e., as the price increases, less demands should be observed. The variance is $\sigma_n^2 = \sigma^2 = 100$ for all n. The maximum likelihood in Eq.(30) is estimated as $\hat{\beta}_{10}$

$= 0.016$ and its estimated standard deviation (i.e., the positive square root of Eq.(34)) is $\hat{\sigma}_{10} = 0.001$.

For $x' = 115$, the next period demand is normally distributed with parameters $\mu' = 100 - 0.4 * 115 = 54$ and variance $\sigma^2 = 100$. Now, for normally distributed demands, the true optimal order quantity from Eq.(6) is given by $q^* = \mu' + \sigma\Phi^{-1}(c_u/(c_u + c_0))$, where Φ is the cumulative distribution function of the standard normal distribution. Furthermore, the Scarf's ordering rule q^{*S} in Eq. (39) is strictly greater than zero at all levels of the critical ratio because for all mark up values, the following is satisfied: $s \geq \left(\frac{\sigma}{\mu}\right)^2 = 0.0039$.

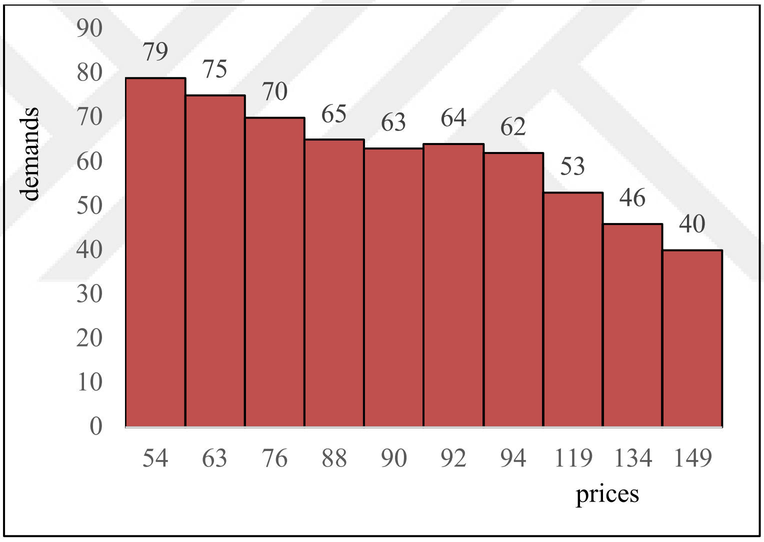

Fig. 5.1: Normal distributed demands and their corresponding prices

Fig. 5.1 shows the prices and demands as observed by the newsvendor. For the prices and demands given in Fig. 5.1, the four different order quantities resulting from our procedure $\hat{q}_M^*$, Sample Average Approximation $\hat{q}_M^{*SA}$, true optimal q^*, and the Scarf's ordering rule q^{*S} are shown in Fig. 5.2, and these ordering quantities are presented in Table 5.1. For normal demands, the optimal ordering quantities $\hat{q}_M^*$ are very similar to $\hat{q}_M^{*SA}$; see Fig. 5.2 or Table 5.1. Both $\hat{q}_M^*$ and $\hat{q}_M^{*SA}$ are below q^* as the critical ratio increases. For normal demands, the best approximation to q^* is given by q^{*S}.

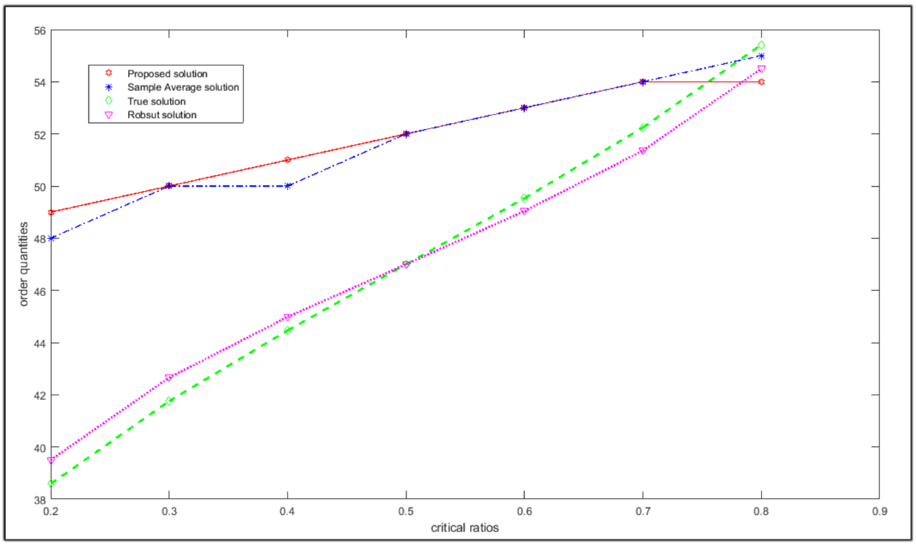

Fig. 5.2: Ordering quantities at different levels of the critical ratio: normal demands

Table 5.1: Order quantities at different critical ratios: normal demands

Critical ratio	Mark up	Order quantities			
		$\hat{q}_M^*$	$\hat{q}_M^{*SA}$	q^*	q^{*S}
0.2	0.25	49	48	45.58	46.50
0.3	0.43	50	50	48.76	49.65
0.4	0.67	51	51	51.47	51.98
0.5	1.0	52	52	54.00	54.00
0.6	1.5	53	53	56.53	56.04
0.7	2.33	53	54	59.24	58.36
0.8	4.0	54	55	62.42	61.50

5.2 Gamma Distributed Demands

Now, we assume that the demands are gamma distributed with shape parameter one and scale parameters are given by

$$\frac{D_{max}}{3} - 0.2x_n \tag{42}$$

for $n = 1, \dots, 10$. The observed demands as well as the prices are shown as a histogram in Fig. 5.3. The maximum likelihood estimate is $\hat{\beta}_{10} = -0.015$ and its estimated standard deviation is $\hat{\sigma}_{10} = 0.001$.

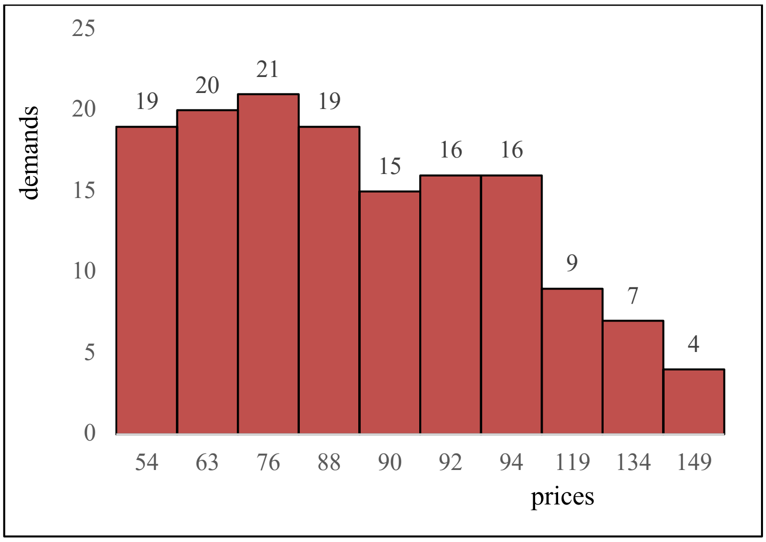

Fig. 5.3: Gamma distributed demands and their corresponding prices

We again consider $x' = 115$ as the next period price so that the next period demand is gamma distributed with shape parameter one and scale parameter 10.33. The true optimal solution q^* is given by Eq. (6) for those values of the critical ratio in Fig. 5.4.

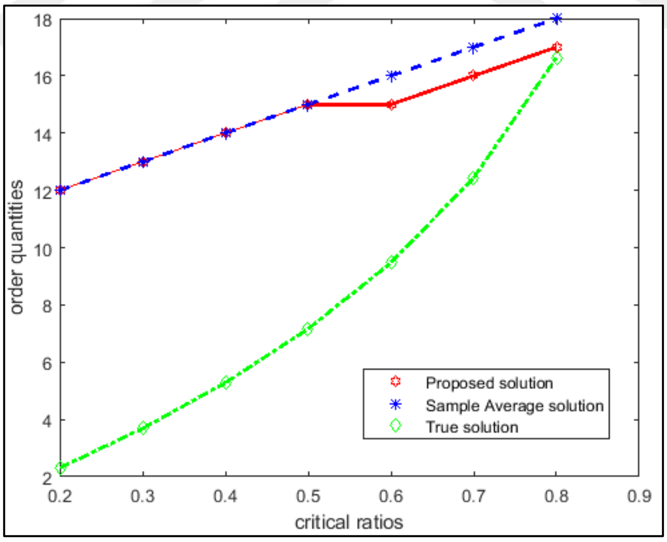

Fig. 5.4: Ordering quantities at different levels of the critical ratio: gamma demands

For this case, we find the order quantities $\hat{q}_M^*$, $\hat{q}_M^{*SA}$, and q^* at different levels of the critical ratio in see Fig. 5.4 for which the actual values of the order quantities are given in Table 5.2. Once again, our order quantities and the order quantities suggested by the Sample Average Approximation are very similar, and are always higher than the true order quantities. Because the newsvendor does not know the true distribution, by estimating the order quantities by either method, the newsvendor will order more than necessary except at the critical ratio 0.8.

Table 5.2: Order quantities at different critical ratios: gamma demands

Critical ratio	Order quantities		
	$\hat{q}_M^*$	$\hat{q}_M^{*SA}$	q^*
0.2	12	12	2.31
0.3	13	13	3.69
0.4	14	14	5.28
0.5	15	15	7.16
0.6	15	16	9.47
0.7	16	17	12.44
0.8	17	18	16.63

5.3 Lognormal Distributed Demands

For this section, the demands are assumed to be lognormal with the means and variance given by

$$\mu_n = D_{max} - 0.4x_n = 100 - 0.4x_n \quad (43)$$

and $\sigma_n^2 = \sigma^2 = 100$ for $n = 1, \ldots, 10$ as in Subsection 1 of this section. Then, the means μ_{normal} and the variances σ_{normal}^2 of the underlying normal demands are given by

$$\mu_{normal} = ln\left(\frac{\mu_n^2}{\sqrt{\mu_n^2 + \sigma^2}}\right), \sigma_{normal}^2 = ln\left(1 + \frac{\sigma^2}{\mu_n^2}\right). \quad (44)$$

The next period price is again $x' = 115$. The maximum likelihood estimate is $\hat{\beta}_{10} = 0.016$ with an estimated standard deviation $\hat{\sigma}_{10} = 0.001$. The demands observed by the newsvendor is given in the histogram in Fig. 5.5.

Again, the true optimal order quantities q^* obtained from the true distribution by Eq. (6) are compared with $\hat{q}_M^*$ and $\hat{q}_M^{*SA}$ in Fig. 5.6 for which the numerical values are given in Table 5.3. For this case, for low values of the critical ratio, the ordering quantities of our procedure and Sample Average Approximation are very big compared to q^*; see Fig. 5.6 and Table 5.3.

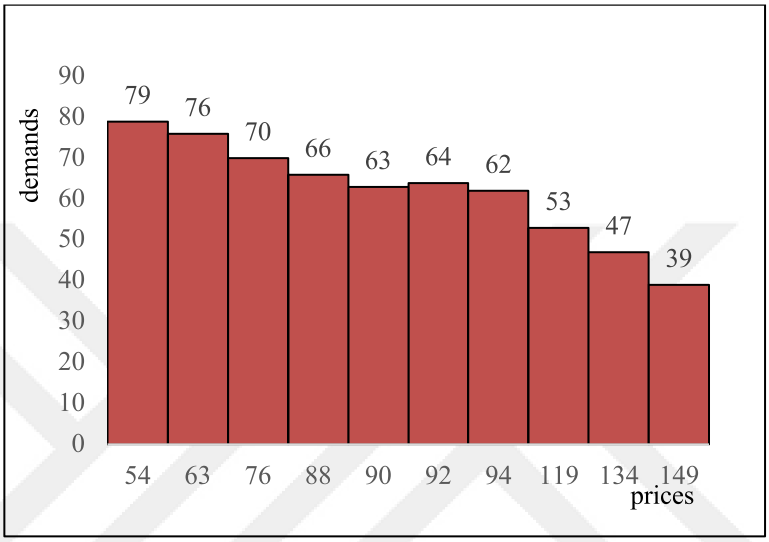

Fig. 5.5: Lognormal distributed demands and their corresponding prices

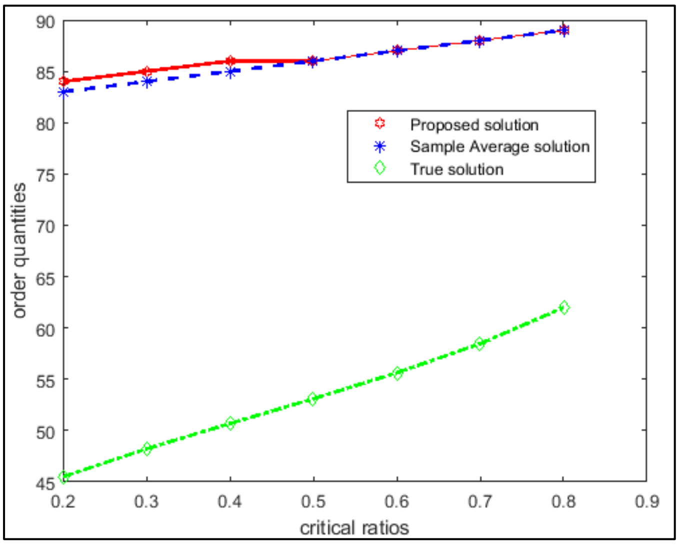

Fig. 5.6: Ordering quantities at different levels of the critical ratio: lognormal demands

Table 5.3: Order quantities at different critical ratios: lognormal demands

Critical ratio	Order quantities		
	$\hat{q}_M^*$	$\hat{q}_M^{*SA}$	q^*
0.2	84	83	45.49
0.3	85	84	48.22
0.4	86	85	50.68
0.5	86	86	51.09
0.6	87	87	55.63
0.7	88	88	58.47
0.8	89	89	61.97

6. CONCLUSION

6.1 Thesis Contribution

In this research, we consider the classic newsvendor problem with a single product and a single period. We assume that the demand distribution is unknown except for its support, yet there exist data of aggregated demands and attribute vectors that can be used to leverage the demand distribution. Considering that the probability of buying a product is approximated by a logit model, we approximate the demand by a binomial process, and decide the estimated demand by using this process. Afterwards, we identify the estimated optimal order quantity by minimizing the expected cost function through optimization. Therefore, an integrated methodology is employed by performing both estimation of demand, which is assumed to be depending on a single extraneous determinant named price, and identification of optimal order quantity as an inventory decision.

With the intention of practicing numerical examples, demands based on a single attribute, specifically price, are generated by using normal, gamma, and lognormal distributions in separate cases. We also consider two benchmark problems, namely the robust newsvendor and the Sample Average Approximation (SAA) solutions of the order quantity. We make a comparison between the solutions of the aforementioned approaches and the solution obtained by our proposed integrated method and observe their results for different levels of critical ratio, hence for different levels of mark up. We conclude that our procedure gives very similar results as those of the SAA. Furthermore, we tend to order more than the true optimal order quantity, similarly to the SAA solution.

6.2 Limitations and Future Work

In this study, we provide a learning-enabled stochastic inventory problem for a single product. It can be extended by employing the proposed data-driven methodology to a newsvendor problem with multiple products, in which an ordering policy for a group of products has to be determined rather than a single ordering quantity.

This study can also be improved by associating multiple attributes. In case of the presence of data regarding attributes that have influence on the product's demand other than its price, utilization of those data values of external attributes could allow to have more accurate demand estimate results.

REFERENCES

Akiva, B. M., & Lerman, S.R. (1985). MIT Press Series in Transportation Studies: Discrete Choice Analysis: Theory and Application to Travel Demand (M. L. Manheim, Ed.). Cambridge, MA: The MIT Press.

Bertsimas, D., & Thiele, A. (2005). A Data-Driven Approach to Newsvendor Problems. Technical Report, Massachusetts Institute of Technology, Cambridge, MA.

Bertsimas, D., & Kallus, N. (2020). From predictive to prescriptive analytics, *Management Science*, 66(3): 1025–1044. doi: 10.1287/mnsc.2018.3253

Gallego, G. & Moon, I. (1993). The Distribution Free Newsboy Problem: Review and Extensions. *Journal of the Operational Research Society,* 44(8): 825-834. https://doi.org/10.1057/jors.1993.141

Green, L. V., Savin, S., & Savva, N. (2013). Nursevendor problem: Personnel staffing in the presence of endogenous absenteeism. *Management Science,* 59(10): 2237–2256. doi: 10.1287/mnsc.2013.1713

He, B., Dexter, F., Macario, A., & Zenios, S. (2012). The timing of staffing decisions in hospital operating rooms: Incorporating workload heterogeneity into the newsvendor problem. *Manufacturing & Service Operations Management,* 14(1): 99–114. doi: 10.1287/msom.1110.0350

Huber, J., Müller, S., Fleischmann, M., & Stuckenschmidt, H. (2019). A data-driven newsvendor problem: From data to decision. *European Journal of Operational Research,* 278(3): 904–915. https://doi.org/10.1016/j.ejor.2019.04.043

Janssen, L., Claus, T., & Sauer, J. (2016). Literature review of deteriorating inventory models by key topics from 2012 to 2015. *Journal of Production Economics,* 182(C): 86-112. http://dx.doi.org/10.1016/j.ijpe.2016.08.019

Levi, R., Roundy, R. O., & Shmoys, D. B. (2007). Provably near-optimal sampling-based policies for stochastic inventory control models. *Mathematics of Operations Research,* 32(4): 821–839. doi: 10.1287/moor.1070.0272

Levi, R., Perakis, G., & Uichanco, J. (2015). The data-driven newsvendor problem: New bounds and insights," Operations Research, 63(6): 1294–1306. doi: 10.1287/opre.2015.1422

Limaye, B. V. (2004). Functional Analysis (2nd ed.). New Delhi, India: New Age International (P) Ltd. Publishers.

Natarajan, K., Sim, M., & Uichanco, J. (2008). Asymmetry and ambiguity in newsvendor models. *Management Science,* 64(7): 3146-3167. doi: 10.1287/mnsc.2017.2773

Perakis, G. & Roels, G. (2008). Regret in the newsvendor model with partial information. *Operations Research*, 56(1): 188–203. doi: 10.1287/opre.1070.0486

Rahimian, H. G., Bayraksan, G., & de Mello T. H. (2019). Controlling risk and demand ambiguity in newsvendor models. *European Journal of Operational Research,* 279(3): 854-868. doi: 10.1016/j.ejor.2019.06.036

Rudin, C. & Vahn, G. Y. (2019). The Big Data Newsvendor: Practical Insights from Machine Learning Analysis. *Operations Research,* 67(1): 90-108. https://doi.org/10.1287/opre.2018.1757

Santambrogio, F. (2015). *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling.* doi: 10.1007/978-3-319-20828-2

Scarf, H. (1958). Studies in the Mathematical Theory of Inventory and Production: A min-max solution of an inventory problem (K. Arrow, S. Carlin, and H. Scarf, Eds.). Redwood City, CA: Stanford University Press.

Whitin, T. M. (1955). Inventory Control and Price Theory. *Management Science,* 2(1): 61-68. https://doi.org/10.1287/mnsc.2.1.61

Zipkin, P. H. (2000). Foundations of Inventory Management. New York: The McGraw-Hill Companies, Inc.

BIOGRAPHICAL SKETCH

Gözde Köybaşı was born on 15th September 1993, in Istanbul.

As for her educational background, Köybaşı has successfully completed her bachelor's degree in industrial engineering from Galatasaray University with a score of 3.76/4.0 in 2017. Regarding high school education, Köybaşı has graduated from Darüşşafaka Schools in 2012.

In respect to her work experience, Köybaşı is currently a corporate finance associate at EY Turkey, in the field of Project Finance and Capital and Debt Advisory.