Annals of Operations Research (2019) 275:607–621
https://doi.org/10.1007/s10479-018-3016-7

ORIGINAL RESEARCH

Robust newsvendor problems: effect of discrete demands

Anh Ninh[1] · Honggang Hu[2] · David Allen[1]

Published online: 21 August 2018

Abstract
Distribution-free newsvendor models often assume continuous demand distributions to facilitate analysis and computation. However, in practice, discrete demand is a natural phenomenon. So far, there exists no analytical and computational results in the literature under this setting. Thus, the goal of this paper is to investigate the newsvendor problems with partial information when the demand is discrete and solve them using the so-called *discrete moment problems*. Numerical results are presented to illustrate the value of discrete information.

Keywords Discrete moment · Newsvendor problems · Stop-loss · Discrete demand · Shape information

1 Introduction

The classical newsvendor problem is a fundamental building block of many stochastic inventory control models. Its goal is to help decision making of a retailer who plans to sell a product facing random (continuous or discrete) demand with a known distribution. Before any sale periods, the retailer must choose and commit to an order quantity. Then, the actual demand is observed and satisfied as much as possible with inventory on hand. If stock out occurs, the retailer is penalized with an underage cost for each unit of unsatisfied demand. On the other hand, if there are unsold units, the retailer has to pay an overage cost for each unit. Further, unsold units are salvaged at the end of the sale period. The objective of the retailer is to select an optimal order quantity to maximize her profit or minimize her cost. If the demand distribution is known, the optimal order quantity can be computed in closed-form formula by the critical fractile.

However, in reality, full characterization of demand distributions is almost never available. Still, the retailer must attempt to make good inventory decisions. Generally, the only

✉ Anh Ninh
atninh@wm.edu

Honggang Hu
hhu01@email.wm.edu

David Allen
dwallen@email.wm.edu

1 Department of Mathematics, College of William and Mary, Williamsburg, USA

2 Warrington College of Business, University of Florida, Gainesville, USA

information available is historical data of demand. One approach is to use approximation and fit a continuous distribution over the data to compute the respective order quantity. In certain situations, this approach can become inappropriate since it can either lead to negative demand or fails to capture asymmetry of the true demand distributions. Another approach utilizes the limited information (moments, range, unimodality) of demand distributions and makes the ordering decision accordingly. This class of methods is called the *distribution-free newsvendor problems* in the literature. We refer the readers to Gallego and Moon (1993), Perakis and Guillaume (2008) and references therein for further details.

A classic solution to a distribution-free newsvendor problem is due to Scarf (1958). Scarf's approach aims to maximize the worst-case objective (e.g., expected profit or payoff) over a parametric family of demand distributions, which brings about its name as the *maximin* approach. Scarf was able to derive an analytical formula for the optimal inventory decision. It is well-known that the maximin approach can result in conservative policies. A less conservative approach is developed under the name minimax regret, which minimizes the maximum opportunity cost from not making the optimal decision (see, e.g., Perakis and Guillaume 2008). These models do not include information about the asymmetry of the distribution. Recent literature attempts to address this issue by inclusion of higher-order moments, such as skewness, kurtosis, or semivariance to the optimization problems (see, e.g, Natarajan et al. 2018). Some other researchers advocate the use of maximum entropy to approximate the demand distribution instead (Andersson et al. 2013).

Regardless of modeling choice for ordering decision under uncertainty, the assumption of continuous demand distributions are often utilized in the robust newsvendor literature. Consequently, given a fixed ordering decision, the distribution-free newsvendor problem can be casted as a continuous moment bounding problem whose extremal solutions can be used to facilitate derivation of optimal policies and analytical insights. Even though a continuous and differentiable demand distribution can help simplify analysis and facilitate derivation of closed-form policies, discrete demands are ubiquitous in practical setting. Thus, it is worthwhile to investigate the deviation of the optimal policies corresponding to discrete and continuous demand distributions (see, e.g., Swaminathan and Shanthikumar 1999; Axsäter 2013).

The goal of this paper is to investigate and solve distribution-free newsvendor problems under discrete demand assumption. In this context, we rely on the so-called discrete moment problems (DMPs) to explore the structure of the optimal solutions. Discrete moment problems (DMPs) were introduced and studied by Prékopa (1988, 1990a, b, c). In those papers, the author used linear programming techniques to develop theory and numerical solution of the optimization problems. The optimization methods are of dual type and are in close relationship with the dual method of Lemke (1954) for the solution of the general linear programming problem. They are stable and fast thanks to the discovery of the structures of dual feasible bases. These approaches remain state-of-the-art and has seen application in numerous areas including project management, finance, risk management, etc. (see, e.g., Prékopa et al. 2016).

The outline of the paper is as follows. Section 2 briefly reviews approaches for decision making under uncertainty in the context of newsvendor problems. Section 3 introduces the discrete moment problems to find the closed-form bounds for a certain type of objective function and utilize it to determine the optimal order quantity. Section 4 presents numerical results and Sect. 5 concludes the paper.

2 Modeling and preliminary

In this section, we briefly review the discrete moment problems and present the formulations for the discrete distribution-free newsvendor models from their respective continuous counterparts. In particular, we consider two approaches for decision making under uncertainty: the maximin approach by Scarf (1958) and minimax regret approach by Perakis and Guillaume (2008).

2.1 Discrete moment problems

Discrete moment problems (DMP) came to prominence by the discovery that the classical probability bounds, using some binomial moments, can be found as optimal solutions of some special linear programs. These results were introduced and extensively studied by Prékopa (1988, 1990a, b, c). Since then, DMP has been widely used in many application areas including data mining, biology, finance, engineering, etc. For further information on DMP and their generalizations, we refer the readers to Prékopa (1992, 1995, 1998, 2001, 2009), Subasi et al. (2009), Mádi-Nagy and Prékopa (2004), Mádi-Nagy (2008, 2012) and Ninh and Prékopa (2013).

Let X be a discrete random variable, the possible values of which are known to be the numbers $z_0 < z_1 < \cdots < z_n$ and the point probabilities are denoted by

$$x_i = P(X = z_i), \quad i = 0, 1, \ldots, n. \tag{1}$$

Given the knowledge of some power moments $\mu_k = E(X^k),\ k = 1, \ldots, m$, where $m < n$, the discrete moment (bounding) problem (DMP) provides us with the sharp lower and upper bounds on $E[f(X)]$ when the distribution of X is unknown. DMP can be formulated as the following linear programs

$$\begin{aligned} \min(\max) \sum_{i=0}^{n} f_i x_i \\ \text{subject to} \\ Ax = b \\ x \geq 0, \end{aligned} \tag{2}$$

where

$$A = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ z_0 & z_1 & \ldots & z_n \\ & & \ldots & \\ z_0^m & z_1^m & \ldots & z_n^m \end{pmatrix}, \quad b = \begin{pmatrix} \mu_0 \\ \mu_1 \\ \ldots \\ \mu_m \end{pmatrix}, \tag{3}$$

and $f_i = f(z_i), i = 0, \ldots, n$.

Linear programming techniques are often utilized to develop theory and numerical solution of the optimization problems (2) (see, e.g., Prékopa 1995). However, due to the high numerical instability of the Vandermonde matrix A, directly solving the primal problem is not recommended. Hence, the optimization methods of choice are of dual type and are in close relationship with the dual method of Lemke (1954) for the solution of the general linear programming problem. They are stable and fast thanks to the discovery of the structures of dual feasible bases (see, e.g., Prékopa 1995) for some special objective functions. More importantly, these structures facilitate derivation of closed-form bounds when only a few moments are available.

2.2 Distribution-free newsvendor problem

Consider a newsvendor facing a random demand X with support on a discrete set $\Omega = \{z_0, \ldots, z_n\}$. If the demand distribution is completely known, then for a given order quantity q, the expected profit can be computed by $\Pi(q) = pE[\min(X, q)] - cq$, where c is the unit ordering cost and p denotes the unit selling price. In this setting, the newsvendor can compute the optimal ordering quantity to maximize the expected profit using critical fractile. Specifically, given a profit margin $\beta = c/p$, then the optimal order quantity is the smallest q such that $P(X \leq q) \geq 1 - \beta$. In what follows, we assume that the full distribution of the discrete demand is not available, but the newsvendor can estimate the first few moments and collect some shape information of demand from historical data.

Maximin approach is due to Scarf (1958). The newsvendor makes decision by choosing the right order quantity to maximize the worst-case profit

$$\min_{X \in \mathcal{X}} pE[\min(X, q)] - cq,$$

where $\mathcal{X}$ is a parametric family of demand distributions, obtained with known information. If the first two moments are available, then the preceding optimization can be written as a DMP

$$\begin{aligned} \Pi(q) = \min\nolimits_{X \in \mathcal{X}} \; & pE[\min(X, q)] - cq \\ \text{s.t} \quad & \\ & x_0 + x_1 + \cdots + x_n = 1 \\ & x_0 z_0 + \cdots + x_n z_n = \mu_1 \\ & x_0 z_0^2 + \cdots + x_n z_n^2 = \mu_2 \\ & x_0, \ldots, x_n \geq 0, \end{aligned} \tag{4}$$

where $x_i = P\{X = z_i\}, i = 0, \ldots, n$ are unknown. The optimal ordering quantity is then found by maximizing $\Pi(q)$. By direct computation, it can be easily verified that the objective function in the linear program (4) can be rewritten as $\Pi(q) = \mu_1 p - cq - pE[(X - q)_+]$.

Minimax regret approach is introduced to mitigate the conservativeness of the maximin approach (see, e.g., Perakis and Guillaume 2008). Given an ordering decision q and a probability distribution of demand F, the regret measures the additional profit that could have been obtained with full information about the distribution, that is, $\max_{z \geq 0}\{\Pi_F(z)\}\Pi_F(y)$. If the distribution of the demand is unknown, then the maximum regret, given q, is defined by

$$\rho(q) = \max_{F \in \mathcal{X}} \max_{z \geq 0}\{\Pi_F(z)\}\Pi_F(y), \tag{5}$$

which can be interpreted as the maximum price the newsvendor would pay to know the exact demand distribution. Thus, the decision criterion consists of minimizing $\rho(q)$.

Following the formulation in Perakis and Guillaume (2008), the optimization problem for the newsvendor is

$$\min \rho(q) = \min_{q \geq 0} \max_{z \geq 0} \left\{ \max_{F \in \mathcal{X}} E[\min(X, z)] - E[\min(X, q)] \right\} + \beta(q - z), \tag{6}$$

where $\beta = c/p$. The inner problem, consisting of finding the distribution that maximizes the regret for given q and z, can then be formulated as a DMP:

$$\begin{aligned}
&\max_{X\in\mathcal{X}}\ E[\min(X,z)-\min(X,q)]\\
&\text{s.t}\\
&\qquad x_0+x_1+\cdots+x_n=1\\
&\qquad x_0z_0+\cdots+x_nz_n=\mu_1\\
&\qquad x_0z_0^2+\cdots+x_nz_n^2=\mu_2\\
&\qquad x_0,\ldots,x_n\geq 0.
\end{aligned}\tag{7}$$

3 Decision-making with discrete moments

Both Scarf's maximin and the minimax regret approaches to the distribution-free discrete newsvendor problem requires solving some discrete moment (upper bounding) problem with respect to a stop-loss type objective function. In particular, this function is defined as follows (see Fig. 1)

$$f(x)=\begin{cases}0 & \text{if } x<d\\ x-d & \text{if } d\leq x\leq L\\ L-d & \text{if } L<x\leq n,\end{cases}\tag{8}$$

where $d \leq L$. Observe that the objective function in the linear program (7) has the above form, where z and q play the role of d and L, respectively, while the objective function in the linear program (4) can be considered as a special case with $L = n$. In the risk and insurance literature, the function defined by (8) is refered to as a stop-loss function with multiple retention levels. Note that when $L = n$, it is reduced to the regular stop-loss function.

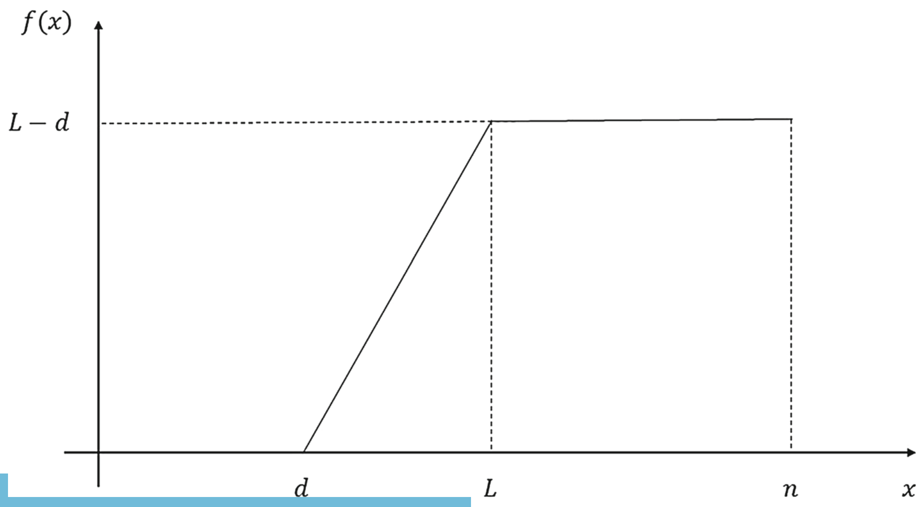

Fig. 1 Stop-loss with multiple retention levels

3.1 Discrete moment bounds

Here, we assume that the first two moments are given and derive the corresponding structures of the dual and optimal basis for both minimization and maximization problems. To ensure feasibility, it can be easily shown that we need the assumptions $0 < \mu_1 < n$, $\mu_1^2 < \mu_2 < n\mu_1$. To this end, let I be the set of subscripts of those columns of the coefficient matrix A which form a basis B.

Lemma 1 *Every dual feasible basis subscript set to the DMPs with the objective function defined by* (8) *has one of the following structures:*

Maximization problem

- $\{0, j, j+1\}$, *if* $2d - 1 \leq j < L$
- $\{i, j, j+1\}$, *if* $2d - 1 \leq i + j \leq 2d$, $0 < i \leq d$ *and* $j < L$
- $\{i, i+1, L\}$, *if* $i \leq 2d - L$
- $\{i, i+1, j\}$, *if* $2d - 1 \leq i + j \leq 2d$, $i < d$ *and* $d < j < L$
- $\{0, L, n\}$
- $I \subset \{L, \ldots, n\}$;

Minimization problem

- $\{d, j, j+1\}$, *if* $2L - 1 \leq d + j$, *and* $L \leq j < n$
- $\{i, j, j+1\}$, *if* $2L - 1 \leq i + j \leq 2L$, $d < i$, *and* $L \leq j < n$
- $\{i, i+1, j\}$, *if* $2L - 1 \leq i + j \leq 2L$, $d \leq i$, *and* $L \leq j < n$
- $\{0, d, n\}$
- $I \subset \{0, \ldots, d\}$,

where in all parentheses the numbers are arranged in increasing order. Those bases for which $I \subset \{0, \ldots, d\}$ ($I \subset \{L, \ldots, n\}$) *are dual degenerate in the minimization (maximization) problem. The bases in all other cases are dual non-degenerate.*

Proof We shall prove the claim for the maximization problem. For every given structure, the values for the dual variables y_k $(k = 0, 1, 2)$ can be explicitly computed as follows.

- If the dual feasible basis structure is $\{0, j, j+1\}$, then
  $y_0 = 0;\ y_1 = \dfrac{j^2 + j - 2dj - d}{j^2 + j};\ y_2 = \dfrac{d}{j + j^2}.$
- If the dual feasible basis structure is $\{0 < i \leq d, j, j+1\}$, then
  $y_0 = \dfrac{i(d - di - j + 2dj - j^2)}{(j-i) + (j-i)^2};\ y_1 = \dfrac{-d + i^2 + j - 2dj + j^2}{(j-i) + (j-i)^2}$
  $y_2 = \dfrac{d - i}{(j-i) + (j-i)^2}.$
- If the dual feasible basis structure is $\{i, i+1, n\}$, then
  $y_0 = \dfrac{i\,(i+1)\,(n-d)}{(i-n) + (i-n)^2};\ y_1 = \dfrac{(d-n)\,(2i+1)}{(i-n) + (i-n)^2};\ y_2 = \dfrac{n-d}{(i-n) + (i-n)^2}.$
- If the dual feasible basis structure is $\{i, i+1, d \leq j < n\}$, then
  $y_0 = \dfrac{i\,(i+1)\,(j-d)}{(i-j) + (i-j)^2};\ y_1 = \dfrac{d + 2di - j - 2ij}{(i-j) + (i-j)^2};\ y_2 = \dfrac{j-d}{(i-j) + (i-j)^2}.$

Plugging these expressions into the constraints of the respective dual problems, we can easily derive the stated conditions. Similarly, the dual feasible basis structures can be established for the minimization problem. □

In what follows, we assume the same conditions on the indices used in Lemma 1. In particular, we ensure that $i < d < j < L$ for maximization problems, and $d < i < L < j$ for minimization problems.

Lemma 2 *The optimal basis to the DMPs with the objective function defined by (8) has the following structures.*

Maximization problem

- *If* $L < \left\lfloor \frac{n\mu_1 - \mu_2}{n - \mu_1} \right\rfloor$*, then the optimal basis is* $I \subset \{L, \ldots, n\}$.
- *If* $\left\lfloor \frac{n\mu_1 - \mu_2}{n - \mu_1} \right\rfloor \leq L \leq \frac{\mu_2}{\mu_1}$*, then the optimal basis is* $\{0, L, n\}$.
- *If* $L > \frac{\mu_2}{\mu_1}$*, we have to consider the below cases.*

  If $\left\lfloor \frac{\mu_2}{\mu_1} \right\rfloor + 1 \geq 2d$*, then the optimal basis is* $\{0, j, j+1\}$*, where* $j = \left\lfloor \frac{\mu_2}{\mu_1} \right\rfloor$.

  If $\left\lfloor \frac{\mu_2}{\mu_1} \right\rfloor + 1 < 2d < L + \left\lfloor \frac{L\mu_1 - \mu_2}{L - \mu_1} \right\rfloor$*, then the optimal basis is* $\{i, j, j+1\}$*, where* $i + \left\lfloor \frac{\mu_2 - i\mu_1}{\mu_1 - i} \right\rfloor \leq 2d \leq i + \left\lfloor \frac{\mu_2 - i\mu_1}{\mu_1 - i} \right\rfloor + 1$ *and* $i \leq \mu_1$.

  If $2d \geq L + \left\lfloor \frac{L\mu_1 - \mu_2}{L - \mu_1} \right\rfloor$*, then the optimal basis is* $\{i, i+1, L\}$*, where* $i = \left\lfloor \frac{L\mu_1 - \mu_2}{L - \mu_1} \right\rfloor$.

Minimization problem

- *If* $d > \frac{\mu_2}{\mu_1}$*, then the optimal basis is* $I \subset \{0, \ldots, d\}$.
- *If* $\left\lfloor \frac{n\mu_1 - \mu_2}{n - \mu_1} \right\rfloor \leq d \leq \frac{\mu_2}{\mu_1}$*, then the optimal basis is* $\{0, d, n\}$.
- *If* $d < \left\lfloor \frac{n\mu_1 - \mu_2}{n - \mu_1} \right\rfloor$*, then we have two cases.*

  If $d + \left\lfloor \frac{d\mu_1 - \mu_2}{d - \mu_1} \right\rfloor \geq 2L - 1$*, then the optimal basis is* $\{d, j, j+1\}$*, where* $j = \left\lfloor \frac{d\mu_1 - \mu_2}{d - \mu_1} \right\rfloor$.

  If $d + \left\lfloor \frac{d\mu_1 - \mu_2}{d - \mu_1} \right\rfloor < 2L - 1$*, then the optimal basis is* $\{i, j, j+1\}$*, where* $i + \left\lfloor \frac{i\mu_1 - \mu_2}{i - \mu_1} \right\rfloor \leq 2L \leq i + \left\lfloor \frac{i\mu_1 - \mu_2}{i - \mu_1} \right\rfloor + 1$ *and* $j = \left\lfloor \frac{i\mu_1 - \mu_2}{i - \mu_1} \right\rfloor$.

Proof The dual feasible bases structures in Lemma 1, in combination with primal feasibility conditions, yields the result. □

Note that when $L = n$, the dual feasible structures are slightly different. For instance, there will be no dual degenerate basis for the maximization problem. Lemma 2 is a generalization to the mean-variance bounds for regular stop-loss function in Courtois and Denuit (2009).

3.2 Closed-form formula for newsvendor application

The results in the previous section give explicit structures of the optimal basis for the objective functions used in newsvendor application. In case of *maximin approach*, a closed-form lower bound on the profit, for a given ordering quantity, can be derived.

Theorem 1 *The closed-form lower bounds for the expected profit $\Pi(q)$, given the first two moments μ_1 and μ_2 are presented below.*

- *If* $2q \leq \left\lfloor \frac{\mu_2}{\mu_1} \right\rfloor + 1$, *then* $\Pi(q) = \frac{pq(\mu_1 + 2\mu_1 j - \mu_2)}{j^2 + j} - cq$, *where* $j = \left\lfloor \frac{\mu_2}{\mu_1} \right\rfloor$.
- *If* $\left\lfloor \frac{\mu_2}{\mu_1} \right\rfloor + 1 < 2q < n + \left\lfloor \frac{n\mu_1 - \mu_2}{n - \mu_1} \right\rfloor$, *and* $h(\lfloor x_2 \rfloor) < 2q - 1$
  - *If* $q^2 - 2q\mu_1 + \mu_2 - 0.25 \leq 0$, *then*
  $$\Pi(q) = \mu_1 p - cq - \frac{p}{6}\left(3\mu_1 - 3q - 2q\mu_1 + \mu_2 + 2 + q^2\right).$$
  Otherwise,
  $$\Pi(q) = (p - c)q - \frac{p\left(j + j^2 - \mu_1 - 2\mu_1 j + \mu_2\right)}{2(2j + 1 - 2q)},$$
  where $j = \left\lfloor q + \sqrt{q^2 - 2q\mu_1 + \mu_2} \right\rfloor$.
- *If* $\left\lfloor \frac{\mu_2}{\mu_1} \right\rfloor + 1 < 2q < n + \left\lfloor \frac{n\mu_1 - \mu_2}{n - \mu_1} \right\rfloor$, *and* $h(\lfloor x_2 \rfloor) \geq 2q - 1$
  - *If* $q^2 - 2q\mu_1 + \mu_2 - 0.25 \leq 0$, *then*
  $$\Pi(q) = \mu_1 p - cq - \frac{p}{6}\left(3\mu_1 - 3q - 2q\mu_1 + \mu_2 + 2 + q^2\right).$$
  Otherwise,
  $$\Pi(q) = \mu_1 p - cq - \frac{p\left(i + i^2 + \mu_2 - \mu_1 - 2i\mu_1\right)}{2(2q - 2i - 1)},$$
  where $i = \left\lfloor q - \sqrt{q^2 - 2q\mu_1 + \mu_2} \right\rfloor$.
- *If* $2q \geq n + \left\lfloor \frac{n\mu_1 - \mu_2}{n - \mu_1} \right\rfloor$, *then we have*
  $$\Pi(q) = \mu_1 p - cq - (n - q)\frac{p\left(i + i^2 - \mu_1 - 2i\mu_1 + \mu_2\right)}{(i - n) + (i - n)^2},$$
  where $i = \left\lfloor \frac{n\mu_1 - \mu_2}{n - \mu_1} \right\rfloor$;

Proof If suffices to show the closed-form upper bound for the expected stop loss $E([X - d]_+)$ given the first two moments μ_1, μ_2. To this end, we introduce the function $h(x)$, defined as

$$h(x) = x + \frac{\mu_2 - x\mu_1}{\mu_1 - x}.$$

Note that there are two distinct solutions x_1 and x_2 to the equation $h(x) = 2d$. Next, use the notation $\sigma^2 = \mu_2 - \mu_1^2$ to compute their values

$$x_1 = d - \sqrt{(d - \mu_1)^2 + \sigma^2}$$
$$x_2 = d + \sqrt{(d - \mu_1)^2 + \sigma^2}.$$

We claim that if $h(\lfloor x_2 \rfloor) < 2d - 1$, there exists an optimal basis with structure $\{0 < i \le d, j, j+1\}$. Otherwise, we will have an optimal basis with structure $\{i, i+1, d \le j < n\}$.

We know that the best upper bounds can be obtained from solving the maximization problem.

– Case 1: As we know, if $2d \le \left\lfloor \frac{\mu_2}{\mu_1} \right\rfloor + 1$, the optimal basis is $\{0, j, j+1\}$ and $j = \left\lfloor \frac{\mu_2}{\mu_1} \right\rfloor$.

$$\begin{aligned} E[(X-d)_+] &= E[X] - E[\min(X, d)] \\ &= \mu_1 - d\left(p_j + p_{j+1}\right) \\ &= \mu_1 - \frac{d(\mu_1 + 2\mu_1 j - \mu_2)}{j^2 + j}. \end{aligned}$$

– Case 2: If $h(\lfloor x_2 \rfloor) < 2d - 1$, there exists an optimal basis with structure $\{0 < i \le d, j, j+1\}$. So, $E[(X-d)_+] = E[X] - E[\min(X, d)] = \mu_1 + (d-i)p_i - d$. If d stays the same, the objective function only depends on p_i. Based on the dual feasibility condition, we consider two cases: $\{2d - 1 - j, j, j+1\}$ and $\{2d - j, j, j+1\}$. Through simple algebra, we find that they have the same objective function value:

$$E[(X-d)_+] = \mu_1 + \frac{j + j^2 - \mu_1 - 2\mu_1 j + \mu_2}{2(2j + 1 - 2d)} - d.$$

Thus, in the optimal value, we can use $i + j = 2d$.
Since we are trying to minimize the dual objective, we consider

$$\begin{aligned} g(j) &= \frac{j + j^2 - \mu_1 - 2\mu_1 j + \mu_2}{2j + 1 - 2d} \\ &= \frac{1}{2}\left(j + \frac{1}{2} - 2\mu_1 + d\right) + \frac{d^2 - 2d\mu_1 + \mu_2 - \frac{1}{4}}{2j + 1 - 2d}, \end{aligned}$$

whose first derivative according to j is given by

$$g'(j) = \frac{1}{2} - \frac{2\left(d^2 - 2d\mu_1 + \mu_2 - \frac{1}{4}\right)}{(2j + 1 - 2d)^2}.$$

If $d^2 - 2d\mu_1 + \mu_2 - 0.25 \le 0$, $g'(j)$ is non-negative which means $g(j)$ is increasing in j. Therefore, $j = d + 1$ and $i = d - 1$ and the optimal value is

$$E[(X-d)_+] = \frac{1}{6}\left(3\mu_1 - 3d - 2d\mu_1 + \mu_2 + 2 + d^2\right).$$

On the other hand, if $d^2 - 2d\mu_1 + \mu_2 - 0.25 > 0$, this function is strongly unimodal in j because

$$g(j) = \frac{1}{4}(2j + 1 - 2d) + \frac{d^2 - 2d\mu_1 + \mu_2 - \frac{1}{4}}{2j + 1 - 2d} - \mu_1 - d$$

and its integer minimizer can indeed be obtained via rounding from its continuous solution. Let $L(j) = \frac{1}{4}(2j+1-2d) + \frac{d^2-2d\mu_1+\mu_2-\frac{1}{4}}{2j+1-2d}$. Since $L(j)$ is a unimodal function in j, an exact expression for the optimal j can be obtained by equating $\Delta L(j) = L(j+1) - L(j)$ to 0, in which case the result is a quadratic equation with solutions given by

$$\begin{aligned} j_1 &= d-1+\sqrt{d^2-2d\mu_1+\mu_2} \\ j_2 &= d-1-\sqrt{d^2-2d\mu_1+\mu_2}. \end{aligned}$$

We select j_1 since it is closer to $j_{\text{real}} = \frac{\sqrt{4\left(d^2-2d\mu_1+\mu_2\right)-1}+2d-1}{2}$. Then $j = \left\lfloor d-1+\sqrt{d^2-2d\mu_1+\mu_2}\right\rfloor + 1 = \left\lfloor d+\sqrt{d^2-2d\mu_1+\mu_2}\right\rfloor$ and $i = 2d - \left\lfloor d+\sqrt{d^2-2d\mu_1+\mu_2}\right\rfloor = \left\lceil d-\sqrt{d^2-2d\mu_1+\mu_2}\right\rceil$.
Similarly, for the case of $h(\lfloor x_2 \rfloor) \geq 2d-1$, we will have an optimal basis with structure $\{i, i+1, d \leq j < n\}$. Then, $E[(X-d)_+] = E[\max(X,d)] - E[d] = (j-d)\,p_j$ and the objective function only depends on p_j. We also consider two cases: $\{i, i+1, 2d-i\}$ and $\{i, i+1, 2d-i-1\}$ according to the dual feasibility. By substituting j with $2d-i$ or $2d-i-1$, we can find that they also have the same objective function:

$$E[(X-d)_+] = \frac{i+i^2+\mu_2-\mu_1-2i\mu_1}{2(2d-2i-1)}.$$

Thus, in the optimal value, we use $i+j = 2d$. Then, we can have

$$\begin{aligned} g(i) &= \frac{i+i^2+\mu_2-\mu_1-2i\mu_1}{2d-2i-1} \\ &= \frac{1}{4}(2d-2i-1) + \frac{d^2-2d\mu_1+\mu_2-\frac{1}{4}}{2d-2i-1} - d + \mu_1. \end{aligned}$$

So, the first derivative according to i is given by

$$g'(i) = -\frac{1}{2} + \frac{2\left(d^2-2d\mu_1+\mu_2-\frac{1}{4}\right)}{(2d-2i-1)^2}.$$

If $d^2-2d\mu_1+\mu_2-0.25 \leq 0$, $g'(i)$ is negative which means $g(i)$ is decreasing in i. Therefore, $i = d-2$ and $j = d+2$ and the optimal value is

$$E[(X-d)_+] = \frac{1}{6}\left(3\mu_1-3d-2d\mu_1+\mu_2+2+d^2\right).$$

On the other hand, if $d^2-2d\mu_1+\mu_2-0.25 > 0$, let $L(i) = \frac{1}{4}(2d-2i-1) + \frac{d^2-2d\mu_1+\mu_2-\frac{1}{4}}{2d-2i-1}$ and it is a unimodal function in i, so an exact expression for the optimal i can be obtained by equating $\Delta L(i) = L(i+1) - L(i)$ to 0 and we can get

$$\begin{aligned} i_1 &= d-1+\sqrt{d^2-2d\mu_1+\mu_2} \\ i_2 &= d-1-\sqrt{d^2-2d\mu_1+\mu_2}. \end{aligned}$$

We select i_2 since it is closer to $i_{\text{real}} = \dfrac{2d-1-\sqrt{4\left(d^2-2d\mu_1+\mu_2\right)-1}}{2}$. Then $i = \left\lfloor d-1-\sqrt{d^2-2d\mu_1+\mu_2}\right\rfloor + 1 = \left\lfloor d-\sqrt{d^2-2d\mu_1+\mu_2}\right\rfloor$ and $j = 2d - \left\lfloor d-\sqrt{d^2-2d\mu_1+\mu_2}\right\rfloor = \left\lceil d+\sqrt{d^2-2d\mu_1+\mu_2}\right\rceil$.

- Case 3: If $2d \geq n + \left\lfloor \frac{n\mu_1-\mu_2}{n-\mu_1}\right\rfloor$, the optimal basis is $\{i, i+1, n\}$ and $i = \left\lfloor \dfrac{n\mu_1-\mu_2}{n-\mu_1}\right\rfloor$.

$$
\begin{aligned}
E[(X-d)_+] &= E[X] - E\left[\min\left(X,d\right)\right] = (n-d)\,p_n \\
&= (n-d)\,\frac{i+i^2-\mu_1-2i\mu_1+\mu_2}{(i-n)+(i-n)^2}.
\end{aligned}
$$

It is straight-forward to derive the lower bounds, which is equivalent to solving the minimization problem.

- Case 1: If $d \leq \dfrac{\mu_2}{\mu_1} \leq n$, the optimal basis is $\{0, d, n\}$.

$$
\begin{aligned}
E[(X-d)_+] &= E[X] - E\left[\min\left(X,d\right)\right] \\
&= \mu_1 - d\left(1-p_0\right) \\
&= \frac{\mu_2-\mu_1 d}{n}.
\end{aligned}
$$

- Case 2: If $\left\lfloor \dfrac{d\mu_1-\mu_2}{d-\mu_1}\right\rfloor \geq d+1, d \leq \mu_1$, the optimal basis is $\{d, j, j+1\}$ and $j = \left\lfloor \dfrac{d\mu_1-\mu_2}{d-\mu_1}\right\rfloor$.

$$
\begin{aligned}
E[(X-d)_+] &= E[X] - E\left[\min\left(X,d\right)\right] \\
&= \mu_1 - d.
\end{aligned}
$$

- Case 3: If $\left\lfloor \dfrac{d\mu_1-\mu_2}{d-\mu_1}\right\rfloor \leq d-2, d \geq \mu_1$, the optimal basis is $\{i, i+1, d\}$ and $i = \left\lfloor \dfrac{d\mu_1-\mu_2}{d-\mu_1}\right\rfloor$.

$$
E[(X-d)_+] = E\left[\max\left(X,d\right)\right] - E\left[d\right] = 0.
$$

□

4 Numerical experiment

To gain insights into the optimal ordering decisions, we compare the solution of the discrete newsvendor problem to that of the continuous counterpart using the *maximin approach*. Let $\delta > 0$ be the finite increment between successive values in the support of the demand.

In our simulation study, we use the beta-binomial demand distribution with parameters $n = 20, a = 0.2, b = 0.25$ and the mixed binomial distribution with parameters $n_1 = 10, n_2 = 50, p_1 = 0.1, p_2 = 0.9, v_1 = v_2 = 0.5$. For instance, $\delta = 1$ means that the support of the demand distribution is $\{0, 1, \ldots, n\}$, while $\delta = 5$ suggests that the support set is $\{0, 5, \ldots, 5n\}$. This demand patterns can be observed for bundles of products for example.

Table 1 Comparison between the continuous solutions and the discrete solutions using maximin approach, when the demand follows a Beta-binomial distribution ($n = 20$, $a = 0.2$, $b = 0.25$)

β	δ	Continuous formulation		Discrete formulation	
		q	Profit	q	Profit
[0.01, 0.1]	1	27.61	7.55	20.00	7.80
	5	137.32	37.70	100.00	38.92
	10	274.75	75.43	200.00	77.87
	15	408.54	112.95	300.00	116.58
[0.1, 0.9]	1	6.49	1.47	8.09	1.53
	5	37.02	8.80	45.48	9.15
	10	67.38	14.83	84.11	15.41
	15	101.60	23.65	126.10	24.69
[0.9, 0.99]	1	0.00	0.00	0.00	0.00
	5	0.00	0.00	0.00	0.00
	10	0.00	0.00	0.00	0.00
	15	0.00	0.00	0.00	0.00

Table 2 Comparison between the continuous solutions and the discrete solutions using maximin approach, when the demand follows a mixed binomial distribution ($n_1 = 10$, $n_2 = 50$, $p_1 = 0.1$, $p_2 = 0.9$, $v_1 = v_2 = 0.5$)

β	δ	Continuous formulation		Discrete formulation	
		q	Profit	q	Profit
[0.01, 0.1]	1	72.39	19.49	50.00	20.28
	5	357.79	97.26	250.00	101.15
	10	724.77	194.89	500.00	202.67
	15	1086.98	292.28	750.00	303.93
[0.1, 0.9]	1	17.85	4.11	23.02	4.30
	5	84.05	18.52	107.24	19.36
	10	174.40	38.56	222.94	40.38
	15	254.02	55.58	328.44	58.18
[0.9, 0.99]	1	0.00	0.00	0.00	0.00
	5	0.00	0.00	0.00	0.00
	10	0.00	0.00	0.00	0.00
	15	0.00	0.00	0.00	0.00

Finally, we randomly choose 3000 values for the profit margin in the following three intervals: [0.01, 0.1], [0.1, 0.9] and [0.9, 0.99]. The average order quantities and respective profits, computed via the maximin approach, are reported in Tables 1, 2.

From the numerical results, we observe that the discrete newsvendor formulation yields a more profitable solution, compared to that from the continuous counterpart. The average profit improvement is about 3% for the Beta-binomial demand and 4% for the mixed binomial case. Our simulation also shows the conservative property of the maximin approach. In particular, when profit margins are randomly chosen between 0.9 and 0.99, the order quantities are always equal to zero. By using the assumption of discrete demand, the new formulation does improve on the order quantities for a wide range of profit margin.

For robustness, we conducted similar experiments for other common discrete distributions (e.g., Discrete uniform, Geometric, Beta-Pascal and Gamma-Poisson distributions) and analyze their respective outputs. If the underlying demand distribution has a single mode, or if it behaves similarly to the normal distribution, the proposed discrete formulation only achieves a minor improvement compared to using the continuous one (i.e., less than 1%). Overall, the performance of our approach significantly improves with multi-modal demand distributions (see Tables 1, 2). Specifically, the best profit improvement of 10% occurs when the two modes are furthest away from each other in the mixed binomial demand model.

5 Discussion

In this paper, we formulate and solve the distribution-free newsvendor problem under the assumption that the random demand is discrete. Our solution approach is based on the theory of discrete moment problems. In particular, Lemma 1 provides a characterization for the dual feasible basis structure for both the maximization and minimization bounding problems on a general stop-loss function. Thus, the reported results can also be of great interest for insurance and risk management applications. In what follows, we discuss several future research directions.

First, the discussion in Sect. 4 poses an immediate question of how to leverage shape information into the decision-making process. Linear shape constraints (e.g., unimodality constraints) in discrete moment problems have been well studied in the literature. We refer the readers to Prékopa et al. (2008), Subasi et al. (2009), Kumaran and Swarnalatha (2017), and Swarnalatha and Kumaran (2017). The common approach is to introduce an additional set of linear constraints and solves its relaxation using Prékopa's dual algorithm. However, so far, there exists no closed-form bounds on the stop-loss objective functions with unimodality constraints.

In addition, it may be worthwhile to study the impact of nonlinear shape constraints in DMPs. For instance, if the underlying random demand is known to be logconcave, i.e., its probability mass function satisfies the relation

$$x_i^2 \geq x_{i-1}x_{i+1}, \quad i = 1, \ldots, n-1, \tag{9}$$

which is a tighter condition than unimodality because logconcavity implies unimodality (see, e.g., Ninh and Pham 2018). Note that the non-convexity of (9) can pose a challenge for optimization. Thus, in the DMPs literature, unimodality constraints have been frequently used as a proxy instead (see, Subasi et al. 2009). Developing algorithms to solve DMP with logconcave constraint is a new area of research.

Lastly, one can also include the following log-convex constraints into a standard DMP

$$x_i^2 \leq x_{i-1}x_{i+1}, \quad i = 1, \ldots, n-1. \tag{10}$$

Though less popular than its log-concave counterpart, log-convexity is important in many application areas (Prékopa 1995). In insurance, compound distributions are frequently used to model losses (see, e.g., Ninh and Prékopa 2015) and there is a wide class of log-convex compound distributions. Note that the log-convexity constraints defined in (10) forms a convex set, and hence, the resulting mathematical program is a convex optimization problem and can be efficiently solved. In particular, these constraints can be rewritten into second-order cone constraints as follows

$$
\begin{aligned}
\min(\max) \quad & \sum_{i=0}^{n} f_i x_i \\
\text{subject to} \quad & \\
& \sum_{i=0}^{n} z_i^k x_i = \mu_k \qquad k = 0, \ldots, m, \\
& \left\| \begin{pmatrix} 2x_i \\ x_{i-1} - x_{i+1} \end{pmatrix} \right\| \le x_{i-1} + x_{i+1}, \; i = 1, \ldots, n-1, \\
& x_i \ge 0 \qquad i = 0, \ldots, n,
\end{aligned} \tag{11}
$$

and we can utilize primal–dual interior point methods to solve problems (11) to optimality (see, e.g., Alizadeh and Goldfarb 2003).

Acknowledgements The authors would like to thank the referees for the constructive comments and discussions that led to this improved version of the paper.

References

Alizadeh, F., & Goldfarb, D. (2003). Second-order cone programming. *Mathematical Programming*, *95*(1), 3–51.

Andersson, J., Jörnsten, K., Nonås, S. L., Sandal, L., & Ubøe, J. (2013). A maximum entropy approach to the newsvendor problem with partial information. *European Journal of Operational Research*, *228*(1), 190–200.

Axsäter, S. (2013). When is it feasible to model low discrete demand by a normal distribution? *OR spectrum*, *35*(1), 153–162.

Courtois, C., & Denuit, M. (2009). Moment bounds on discrete expected stop-loss transforms, with applications. *Methodology and Computing in Applied Probability*, *11*(3), 307–338.

Gallego, G., & Moon, I. (1993). The distribution free newsboy problem: Review and extensions. *Journal of the Operational Research Society*, *44*(8), 825–834.

Kumaran, V., & Swarnalatha, R. (2017). Bounds for the probability of union of events following monotonic distribution. *Discrete Applied Mathematics*, *223*, 98–119.

Lemke, C. E. (1954). The dual method of solving linear programming problem. *Naval Research Logistics Quarterly*, *1*(1), 36–47.

Mádi-Nagy, G., & Prékopa, A. (2004). On multivariate discrete moment problems and their applications to bounding expectations and probabilities. *Mathematics of Operations Research*, *29*(2), 229–258.

Mádi-Nagy, G. (2008). On multivariate discrete moment problems: Generalization of the bivariate min algorithm for higher dimensions. *SIAM Journal on Optimization*, *19*(4), 1781–1806.

Mádi-Nagy, G. (2012). Polynomial bases on the numerical solution of the multivariate discrete moment problem. *Annals of Operations Research*, *200*(1), 75–92.

Natarajan, K., Sim, M., & Uichanco, J. (2018). Asymmetry and ambiguity in newsvendor models. *Management Science*, *64*(7), 2973–3468.

Ninh, A., & Prékopa, A. (2013). The discrete moment problem with fractional moments. *Operations Research Letters*, *41*(6), 715–718.

Ninh, A., & Prékopa, A. (2015). Log-concavity of compound distributions with applications in stochastic optimization. *Discrete Applied Mathematics*, *161*(18), 3017–3027.

Ninh, A., & Pham, M. (2018). Logconcavity, twice-logconcavity and Turán-type inequalities. *Annals of Operations Research*,. https://doi.org/10.1007/s10479-018-2923-y.

Perakis, G., & Guillaume, R. (2008). Regret in the newsvendor model with partial information. *Operations Research*, *56*(1), 188–203.

Prékopa, A. (1988). Boole-Bonferroni inequalities and linear programming. *Operations Research*, *36*(1), 145–162.

Prékopa, A. (1990a). Sharp bounds on probabilities using linear programming. *Operations Research*, *38*(2), 227–239.

Prékopa, A. (1990b). The discrete moment problem and linear programming. *Discrete Applied Mathematics*, *27*(3), 235–254.

Prékopa, A. (1990c). Totally positive linear programming problems. In L. J. Leifmann (Ed.), *Functional analysis, optimization and mathematical economics. A collection of papers dedicated to the memory of L. V. Kantorovich* (pp. 197–207). New York: Oxford University Press.

Prékopa, A. (1992). Inequalities on expectations based on the knowledge of multivariate moments. Lecture Notes-Monograph Series, 309–331.

Prékopa, A. (1995). *Stochastic Programming*. Dordrecht: Kluwer Scientific.

Prékopa, A. (1998). Bounds on probabilities and expectations using multivariate moments of discrete distributions. *Studia Scientiarum Mathematicarum Hungarica*, *34*(1), 349–378.

Prékopa, A. (2001). Discrete higher order convex functions and their applications. In N. Hadjisavvas, J. E. Martínez-Legaz, J. P. Penot (Eds.), *Generalized Convexity and Generalized Monotonicity. Lecture Notes in Economics and Mathematical Systems* (Vol. 502). Berlin, Heidelberg: Springer.

Prékopa, A. (2009). Inequalities for discrete higher order convex functions. *Journal of Mathematical Inequalities*, *3*(4), 485–498.

Prékopa, A., Ninh, A., & Alexe, G. (2016). On the relationship between the discrete and continuous bounding moment problems and their numerical solutions. *Annals of Operations Research*, *238*(1–2), 521–75.

Prékopa, A., Subasi, M., & Subasi, E. (2008). Sharp bounds for the probability of the union of events under unimodality condition. *European Journal of Pure and Applied Mathematics*, *1*(1), 60–81.

Scarf, H. (1958). A min-max solution of an inventory problem. In *Studies in the Mathematical Theory of Inventory and Production*. Stanford University Press.

Subasi, E., Subasi, M., & Prékopa, A. (2009). Discrete moment problems with distributions known to be unimodal. *Mathematical Inequalities and Applications*, *12*(3), 587–610.

Swaminathan, J. M., & Shanthikumar, J. G. (1999). Supplier diversification: Effect of discrete demand. *Operations Research Letters*, *24*(5), 213–221.

Swarnalatha, R., & Kumaran, V. (2017). Bounds for the probability of the union of events with unimodality. *Annals of Operations Research*,. https://doi.org/10.1007/s10479-017-2629-6.