Robust Approach for Optimizing Uncertainty and its Application in Newsvendor

Hyunji Moon*

December 14, 2020

Abstract

Robust optimization aims to provide reasonable solutions under varying coefficients and constraints. This paper proposes support division (SD) model which addresses drawbacks of stochastic programming by transforming the form of uncertainty: the value of a random variable is mapped to its support range index. The model shows its best performance when a relatively accurate range forecast is available. Also, separate approaches are suggested depending on the initial data amount or the distribution shape. When data is small and uncertainty has multimodal form, mixture support division (MSD) model is recommended which first process data with Gaussian mixture model before applying SD. To validate the effect, newsvendor problem setting is introduced along with two existing approaches. Based on the experiment with normal, uniform, and three types of multimodal uncertainty, SD and MSD outperformed the existing two models. Along with the means to control the model's conservativeness, SD and MSD could be applied to a wide spectrum of domains in terms of their required robustness.

Keywords— robust optimization, support division, multimodal distribution, newsvendor

1 Introduction

Most systems have multiple objects which need to be balanced. Based on the value given upon each object by the system, it is modeler's role to find the optimal balance. The process includes representing the final objective as a weighted average of multiple objectives (Boser et al., 1997; Tibshirani, 1996) or solving a sequence of single-objective optimization problems (Arora, 2012). Robustness and efficiency are two popular objectives and this paper aims to suggest a model that balance them. Although robust optimization provides useful solutions in the face of uncertain situations, its conservative result has been criticized (Bertsimas and Sim, 2004). Limiting the degree of inefficiency by narrowing down the uncertainty set is the main motivation of this research. Fields including stochastic programming, scenario optimization, and distributionally robust optimization which are being studied actively along with robust optimization, all share their vein; they just differ in how and what form they model the uncertainty.

The suggested model, support division (SD) extends the idea of robust optimization by discretizing a random variable. This transformation makes model more intuitive which makes reflecting prior information on the uncertainty easy. For example, for retailers, it is easier to give rough estimation on the range of the demand than the accurate value. These prior information are often valuable to build a robust model, especially when there are not enough data. Another value of this research is its customized approach for two difficult conditions: lack of data and uncertainty with multimodal distribution. Many researches (Gelman

*Department of Industrial Engineering, Seoul National University (hjmoon0710@snu.ac.kr).

et al., 2017; Margossian et al., 2020) have discussed the hardness of model fitting and highlighted the effect of priors in the face of these conditions. In that context, mixture support division (MSD) model which combines Gaussian mixture model with SD is proposed. Its performance is tested on different forms of multimodal distributions with different parameters. Lastly, an extended version of SD, which gives the modeler a means to control the model's conservativeness is introduced. With this, SD and its variants could be applied to wide spectrum of domains in terms of the required robustness from risk-taking retailer selling imperishable goods to risk-averse military where stock-out could lead to detrimental results.

2 Literature review

Previous research that motivated SD and MSD model are introduced. Proposed models extend the spirit of robust optimization and directly address multimodal problem. Also, one domain example where models could be directly applied to is newsvendor which is popular in business field.

2.1 Robust optimization

Robust optimization attempts to design a solution which are immune to data uncertainty (Bertsimas and Sim, 2004). Depending on the existence of random variable in the model, it could be classified into non-probabilistic or probabilistic robust optimization. Wald's model (Wald, 1945) is a dominating paradigm for non-probabilistic model. As in equation 1, solutions are ranked on the basis of their worst-case costs and one with the least worst outcome is the optimal. Conservatism of its solutions has been noted (Bertsimas and Sim, 2004). For example, distribution free newsvendor model (Gallego and Moon, 1993) uses maximin approach to optimize order quantity based on limited information of uncertain demand: mean and standard deviation. Though this model could provide robust solutions for diverse distribution, its worst case is two point mass, which is unlikely to be realized.

$$\min_{x \in X} \max_{c \in C(x)} f(x, c) \tag{1}$$

Probabilistic robust optimization, or stochastic programming (SP), is a numerical optimization problem that arises from observing data from random data-generating process (Duchi, 2018). It includes random variables which lead to random objective functions or random constraints. Shapiro (2012) explains minimax SP as a compromise between probabilistic and non-probabilistic optimization. The former has been criticized for its difficulty to specify the distribution of unknown parameters whereas the latter for its conservative outcome.

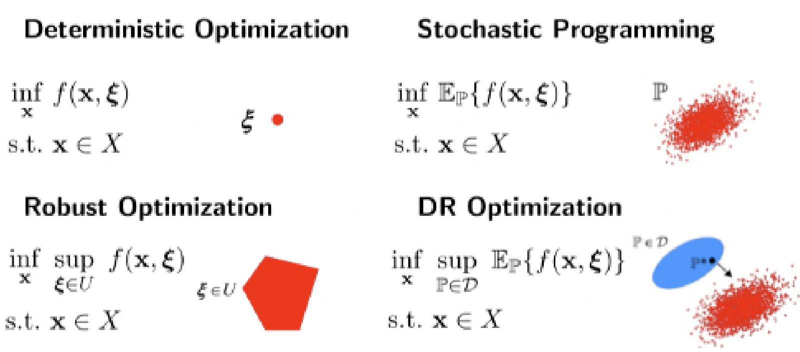

Figure 1: Uncertainty optimization concepts.

Distributionally Robust Optimization (DRO) addresses these issues. Based on the available empirical data, partial distribution information such as support set or moment statistics are extracted to construct uncertainty set, D. DRO optimize the worst-case expected performance on a set constituted by an infinite number of distributions contained in uncertainty set. DRO is less conservative than RO as RO tries to protect against parameter's worst-case realization whereas DRO prevents its worst-case distribution. On the other hand, DRO is easier to formulate than SP as only the support of uncertain parameters are requested in advance compared to their exact distribution. Other concepts such as scenario-based SP (Chen et al., 2002) exist that fill the gap of the four mentioned optimization. Figure 1 from Shang and You (2018) compares the four concepts. Uncertainty set in RO is expressed as a red polytope which includes different parameter configurations. Uncertainty set in DRO is expressed as a blue ellipsoid that contains an infinite number of distributions. Table 1, summarizes the uncertainty set and type of robust optimization, SP, and DRO. One example that could help understand the difference is robust approximation from Boyd et al. (2004). Stochastic and worst case approaches are compared with their minimizing objectives being, $E||Ax - b||$ and $sup_{A \in \mathcal{A}}||Ax - b||$. Furthermore, for DRO, $sup_{u \in \mathcal{U}} E_u||A(u)x - b||$ becomes the objective value, where u is a random variable that controls the randomness of A. Statistical measures are needed to set $\mathcal{U}$.

Table 1: Uncertainty set and type of uncertainty optimization concepts.

	Uncertainty set	Uncertainty type
Robust	polytope	deterministic
SP	-	stochastic
DRO	ellipsoid	both

2.2 Multimodality

Multimodal distribution holds an integral role for two reasons. First from methodology point of view, there are many known problems suffering from multimodal distribution. For example, the quality of approximation is bad for embedded Laplace approximation (Margossian et al., 2020) or mean field variational inference (Barber et al., 2011). Moreover, in DRO, uncertainty sets disregarding the multimodality will produce overly conservative results (Hanasusanto et al., 2015) and for simulation-based methods including Monte Carlo, its estimation error could be great when the random variable's density has multimodal distribution. Ekin et al. (2020) explains the existence of local optimum in multimodal distribution as the cause of these frequent suffering. Another hardship multimodal poses is its number of parameter. Compared to simple Poisson or normal distribution, multimodal has different types of parameter corresponding to weight, mean, and, deviation. This makes the model even difficult when there are not enough data. This is the motive behind GSD, an extension of SD for multimodal uncertainty in small data regime.

However, in application point of view, many multimodal distributions are observed owing to different latent variables that generates data such as different types of customer, topics, or trends (Hanasusanto et al., 2015). These difficulty and naturalness of multimodal uncertainty make research on optimizing bimodal uncertainty, as a simplified example of multimodal, fruitful.

2.3 Newsvendor

Newsvendor model suggests an optimal one-time ordering inventory amount that minimizes the expected cost. It is widely used in business settings (Cachon and Terwiesch, 2008). If the order is larger than the actual demand, overage cost occurs, and if it is smaller, underage cost due to opportunity cost of sales occurs. Therefore, it is important to have an adequate amount of inventory. In the past, there have been studies on inventory optimization algorithms assuming the distribution of demand. Cachon and Terwiesch (2008) argued that if demand follows a normal distribution, the optimal value can be obtained in a closed form. Though normal distribution are frequently observed, the optimal solution that assumes the demand distribution could be inaccurate and not robust. This is especially true if the amount of demand is low. For example, Daskin et al. (2002) argued that demand follows a Poisson process. A demand prediction model that lowers the dimension of the distribution was presented by using the Lagrangian relaxation optimization algorithm that divides the demand distribution into multiple distributions. The following equation shows the basic structure of newsvendor model where c, p is cost and price and q is order quantity. F is the cumulative distribution of the demand.

$$\text{E[opportunity cost + overage cost]} = cqF(q) - c\int_{x\leq q} xf(x)dx + (p-c)\left[\int_{x>q} xf(x)dx - q(1-F(q))\right]$$

If price and demand information are known, optimal stock level can be derived through newsvendor model. Since real demand includes uncertainty, it is not desirable to fit the demand distribution into a specific distribution. There have been studies to derive the optimum value without fitting the demand distribution to a specific distribution. Gallego and Moon (1993) proposed a distribution free model when information on demand distribution is limited to mean and variance only. This model assumes follows the minimax decision-making method that selects the highest value among the lowest profits. The free distribution model has robustness in the form of demand distribution because it derives an optimal value by considering all possible demands using limited information. Considering all possible demands means that even the number of cases with small probability is considered, and for this reason, there is a limitation that the free distribution model may present values different from the actual values when deriving the maximum profit. In this study, the distribution free model was selected as the comparative model together with the normal distribution model. This is to compare the robustness of the segmentation model when a distribution containing uncertainty occurs, contrary to the case of a stable normal distribution. Perakis and Roels (2008) hypothesized a situation in which various information such as the mean and variance of the distribution are given. The decision-making of the model follows the criterion of minimizing maximum regret. In addition, Aryanezhad et al. (2012) paid attention to the difference between the order volume and the actual delivery volume caused by uncertainty in demand. The difference generated by this difference flowing through the supply chain was derived using a nonlinear regression equation, and the optimal value was calculated by a systematic approach that considered the entire supply chain. Past news vendor models were fitted by assuming the distribution or did not assume the distribution. We tried to derive the optimal value using the information of the mean and variance. As the uncertainty of reality is included in the distribution, the actual distribution becomes complex. Therefore, the method that is suitable for assuming a specific distribution is limited in practical application. Models that do not assume a distribution can be said to be robust in that they derive an optimal value even if uncertainty is included, but there is a limit that satisfaction with the derived optimal value is low.

3 Support division model

3.1 Methodological viewpoint: robust optimization

When relatively accurate range forecast is possible, we could specify the uncertainty by transforming a continuous random variable d into discrete random variable m. m is an index that is set by learning the quantiles of historical data. This has two benefits. First, it is easier to average over m than d. Probability estimation becomes much more intuitive and easier when its object is range instead of exact value. Second, if we view the problem as two stage, where q is set in the first stage that minimizes the average of the second stage variable $C(q|d)$ from each d's scenario, the transformation has the effect of reducing the number of scenario. For simplicity, five quantiles are chosen but any numbers and forms of range is possible as long as we know $p(m=1), ..., p(m=M)$. By defining the cost for each given range as a maximum cost over possible demand in the range i.e. $C(q|m) := max_{d \in d(m)} C(q|d)$, the following are equivalent.

$$min_q E_d[C(q|d)] \tag{2a}$$

$$min_q E_m[C(q|m)] \tag{2b}$$

$$min_q E_m[max_{d \in d(m)} C(q|d)] \tag{2c}$$

$$min_q E_m[t_m] \; s.t. C(q|d_m) \le t_m \; \forall d_m \in d(m), \forall m = 1, ..., M \tag{2d}$$

From the equation, d is an uncertainty random variable which follows distribution D and m is uncertainty index random variable that denotes which range of support a certain d belongs to i.e. scenario. d(m) denotes m_{th} range of support division. Compare the original minimax objective in equation (3a) with equation (2c). Averaging over the specified uncertainty has an effect of curtailing conservatism cost.

$$min_q max_{d \in \cup_m d(m)} C(q|d) \tag{3a}$$

Another extension is to incorporate the parameter that represents robustness, Γ, as in Bertsimas and Sim (2004) and Han et al. (2014). The parameter controls the tradeoff between robustness and the level of conservativeness of the solution for each scenario by restricting the maximum number of the worst-case happening simultaneously. Note that L_m and H_m correspond to best case cost and worst case cost that can happen in each set d(m). Q is an arbitrary number with large value used to get best case cost.

$$min_q E_m[\frac{L_m + H_m}{2} * (1 - x_m) + H_m * x_m] \tag{4a}$$

$$C(q|d_m) \le H_m \tag{4b}$$

$$\sum_m x_m \le \Gamma \tag{4c}$$

$$C(q|d_m) \le L_m + Q * (1 - y_{m,k}) \tag{4d}$$

$$\Sigma_k y_{m,k} = 1 \tag{4e}$$

$$\forall d_m \in d(m), x \in B^M, y_{m,k} \in B^k \; \forall m = 1, ..., M, \tag{4f}$$

3.2 Application viewpoint: newsvendor

To validate this model, performance with two existing models are compared. Observing how performance changes depending on different characteristics regarding the true distribution (uniform, exponential, normal, normal mixture) and the amount of data is the main interest of this research. Normal assumption (NA) model and distribution free (DF) model are chosen as two comparison models. Figure 2 compares the difference between existing two approaches with SD. NA and DF are two extremes regarding how much we assume the true distribution. Based on the given mean μ and standard deviation σ of the true distribution, NA optimize over a normal distribution $N(\mu, \sigma)$, assuming a full distribution, while DF does not add any assumption. SD is closer to DF in that it does not presume further on the distribution, but the difference lies in the incorporation of random variable, m. Based on this variable which has equal probability, $1/M$ for each division, worst case conditional on m is searched. Our final objective function is the average of the worst case cost averaged over m. This leads to a more realistic worst case and therefore prevents the model from being overly inefficient in an attempt to be robust. Using a linear structure of the worst cost for each m, it can be shown that objective function, $E_m[C(q|m)]$ could be expressed as the following closed form. $C(q|m)$ is a worst case cost when true distribution belongs to m (m = 1, .., M) and q is a decision variable. Proof can be found in appendix section 6.

$\mathrm{E}_m[C(q|m)] = \frac{1}{M}[\sum_{i=1}^{k-1}(q-d_{i-1})*(p-c)+\sum_{i=k+1}^{M}(d_i-q)*(p-c)+\max[(d_k-q)*(p-c),(q-d_{k-1})*(c-s)]]$

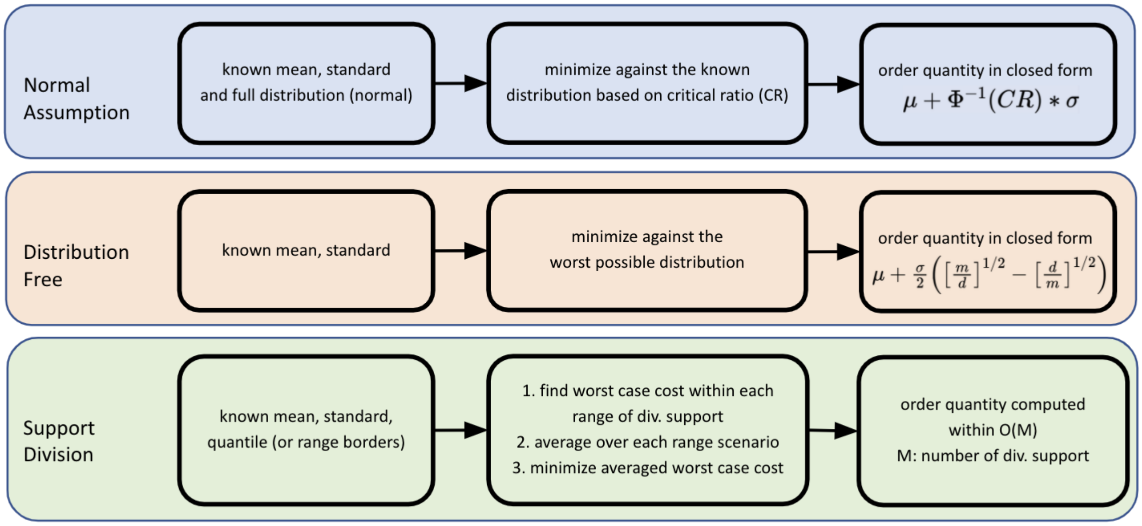

Figure 2: Comparison of profit ratio under normal, uniform, and multimodal distribution.

3.3 Methodological viewpoint: small data and multimodality

Two harsh conditions, small data and multimodality motivates the extended version of our original model, namely MSD. In most cases, inverse of cumulative distribution of historical data could be used to get the quantile needed for the setting its range. If the problem in interest has a structure where quantile is not easily obtained, small data regime for example, different approach is needed. Also, in SD, the number and form of range are flexible, and it is recommended that the modeler should carefully choose them. More accurate mapping between the value and its range makes the result more optimal. When data is multimodal, it is easier to demarcate ranges compared to uniform as can be seen from Figure 3. Considering the cause of the multimodal distribution mentioned in Section 2, it is natural that maximum worst cost should be calculated on separate sets before being aggregated.

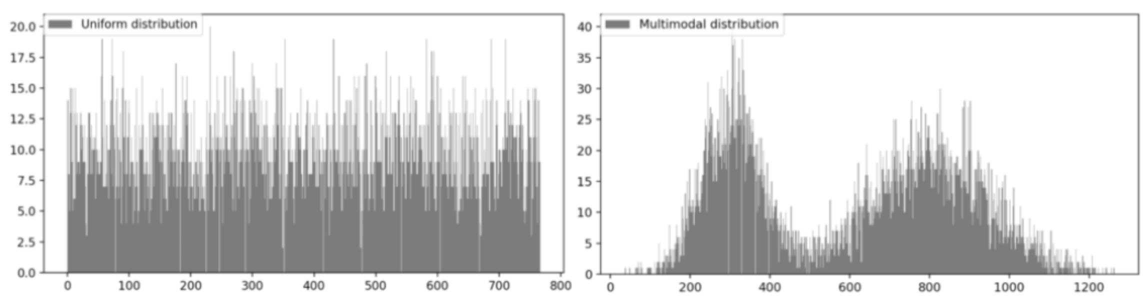

Figure 3: Multimodal uncertainty is easier to map value to range than uniform.

Based on this motivation, MSD which adds Gaussian mixture model with SD is proposed. The process is described in Figure 4. First, GMM is applied to learn the underlying structure of the data. It dissects the support based on the predicted classification probability of each value. For example, in Figure 4, the total of four ranges are set. Based on this discretized range, SD is applied as described in 2.3.

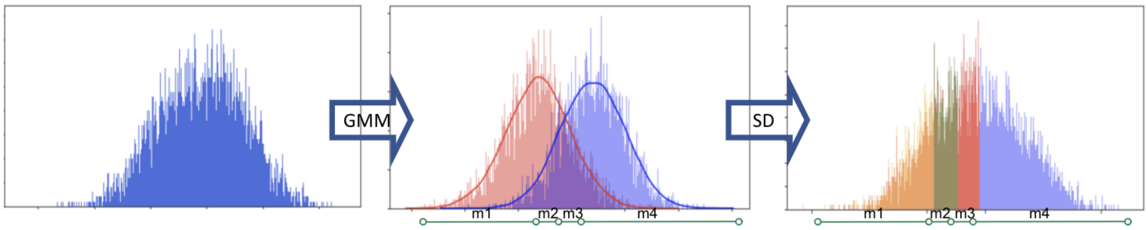

Figure 4: Optimization process of MSD.

4 Experiment and Result

Data were simulated from normal, uniform, and multimodal distribution. As can be seen from Table 2 and Figure 4, SD outperforms NA and DF in most cases. As expected, NA model shows best performance for normal distribution, but the scale of difference is relatively small compared to other differences and could be highly subject to randomness. Note that β which is the margin of the good is used as the axis of Figure 4. In other word, based on the result, we could conclude that SD model give robust optimal results under different distribution and situations.

Table 2: Relative cost of NA, DF, SD for different distribution.

	Normal	Uniform	Multimodal
NA	1.001	0.9959	0.9955
DF	1.0006	1	0.998
SD	0.9984	1.0041	1.0065

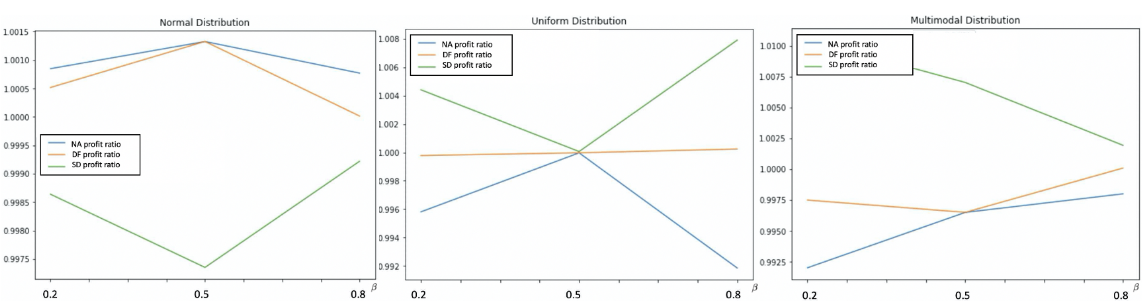

Figure 5: Comparison of profit ratio under normal, uniform, and multimodal distribution.

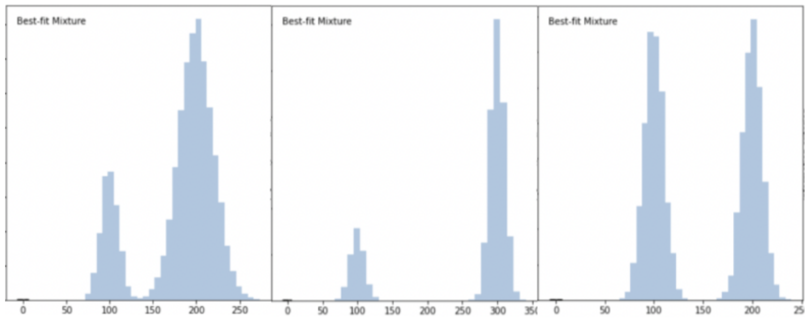

Figure 6: Different bimodal distributions used for the experiment.

Further experiments are performed to investigate the performance of MSD. Different bimodal distributions seen in Figure 5 are tested. As can be seen from Table 3, the SD shines in this setting. It is notable that the more distinct the difference between two modes are, the better result SD gives.

Table 3: Relative cost of NA, DF, SD for bimodal distribution.

$\mu_1, \mu_2, \sigma_1, \sigma_2, w_1$	(100, 200, 10, 20, .2)	(100, 300, 10, 20, .2)	(100, 200, 10, 20, .5)
NA	0.94	0.93	0.95
DF	0.93	0.93	0.95
SD	1.12	1.13	1.09

5 Conclusion

The contribution of this paper is to suggest support division approach which addresses drawbacks of stochastic problem through transforming random variable from value to its range index. Methodologies for range setting in small and large data regime are suggested. Moreover, discussion on addressing multimodal distribution especially in the context of small data have been included along with a extended model, MSD. The model's effectiveness has been tested on different shape and type of distributions.

The limitation and the direction for our research is as follows. First, prior design especially for multimodal distribution is needed. Truncated normal prior which are frequently used for order-constrained settings could be applied. Prior could greatly affect the performance of the model if properly understood in the context of the entire Bayesian analysis, from inference to prediction to model evaluation. Second, is application of hierarchical (or multilevel) model. It is more flexible than discrete or continuous mixture model when addressing heterogeneity and over-dispersion (Gelman et al., 2017). It would be interesting to design a hierachical prior and a likelihood where shared parameter could partial pool the information between subgroups.

The result could be directly applied in business settings to recommend optimal inventory level or order quantity that maximize the profit. Another application domain is military equipment. Efficient conservatism is needed in military; however as budgets are limited efficient management is needed. Determining the amount of spare engines for each battle ship or plane is an integral issue where underestimation could lead to mission failure whereas overestimation could lead to wasted budget. With the means to set the degree of conservatism, SD model could be used in diverse settings to provide both robust and efficient solution.

6 Appendix

Proof. q is the order quantity and d is the demand.

Case 1. $q < d_{m-1}$
Only overage cost exists. For each divided support $[d_{m-1}, d_m]$, cost incurred for each demand is (d-q) * (p-c) which is a piece-wise linear function of d. Therefore, the worst case happens when demand is d_i and its cost is d_i is $(d_i - q) * (p - c)$.

Case 2. $q > d_m$
Only underage cost exists. For each divided support $[d_{m-1}, d_m]$, cost incurred for each demand is (q-d) * (c-s) which is a piece-wise linear function of d. Therefore, the worst case happens when demand is d_{i-1} and its cost is $(q - d_i) * (c - s)$.

Case 3. $q \in [d_{m-1}, d_m]$
From case3-1 and 3-2, the overall worst case cost for case 3 is $\max[(d_k - q) * (p - c), (q - d_{k-1}) * (c - s)]]$.

Case 3-1. $d \in [d_{m-1}, q]$
The worst case happens when demand is d_{i-1} with its cost $(d_{i-1} - q) * (p - c)$.

Case 3-2. $d \in [q, d_m]$
The worst case happens when demand is d_i with its cost $(d_{i-1} - q) * (p - c)$.

To combine the above results, our final objective is

$$\mathrm{E}_m[C(q|m)] = \frac{1}{M}[\sum_{i=1}^{k-1}(q-d_{i-1})*(p-c)+\sum_{i=k+1}^{M}(d_i-q)*(p-c)+\max[(d_k-q)*(p-c),(q-d_{k-1})*(c-s)]].$$

References

J. S. Arora. Introduction to optimum design. 2012.

M.-B. Aryanezhad, S. G. J. Naini, and A. Jabbarzadeh. An integrated model for designing supply chain network under demand and supply uncertainty. *African journal of Business management*, 6(7):2678–2696, 2012.

D. Barber, A. T. Cemgil, and S. Chiappa. *Bayesian time series models*. Cambridge University Press, 2011.

D. Bertsimas and M. Sim. The price of robustness. *Operations research*, 52(1):35–53, 2004.

B. Boser, I. Guyon, and V. Vapnik. Pattern recognition system using support vectors, July 15 1997. US Patent 5,649,068.

S. Boyd, S. P. Boyd, and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

G. Cachon and C. Terwiesch. *Matching supply with demand*. McGraw-Hill Publishing, 2008.

Z.-L. Chen, S. Li, and D. Tirupati. A scenario-based stochastic programming approach for technology and capacity planning. *Computers & Operations Research*, 29(7):781–806, 2002.

M. S. Daskin, C. R. Coullard, and Z.-J. M. Shen. An inventory-location model: Formulation, solution algorithm and computational results. *Annals of operations research*, 110(1-4):83–106, 2002.

J. C. Duchi. Introductory lectures on stochastic optimization. *The mathematics of data*, 25:99, 2018.

T. Ekin, S. Walker, and P. Damien. Augmented simulation methods for discrete stochastic optimization with recourse. *Annals of Operations Research*, pages 1–23, 2020.

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

A. Gelman, D. Simpson, and M. Betancourt. The prior can often only be understood in the context of the likelihood. *Entropy*, 19(10):555, 2017.

J. Han, C. Lee, and S. Park. A robust scenario approach for the vehicle routing problem with uncertain travel times. *Transportation science*, 48(3):373–390, 2014.

G. A. Hanasusanto, D. Kuhn, S. W. Wallace, and S. Zymler. Distributionally robust multi-item newsvendor problems with multimodal demand distributions. *Mathematical Programming*, 152(1-2):1–32, 2015.

C. Margossian, A. Vehtari, D. Simpson, and R. Agrawal. Hamiltonian monte carlo using an adjoint-differentiated laplace approximation: Bayesian inference for latent gaussian models and beyond. *Advances in Neural Information Processing Systems*, 33, 2020.

G. Perakis and G. Roels. Regret in the newsvendor model with partial information. *Operations Research*, 56 (1):188–203, 2008.

C. Shang and F. You. Distributionally robust optimization for planning and scheduling under uncertainty. *Computers & Chemical Engineering*, 110:53–68, 2018.

A. Shapiro. Minimax and risk averse multistage stochastic programming. *European Journal of Operational Research*, 219(3):719–726, 2012.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

A. Wald. Statistical decision functions which minimize the maximum risk. *Annals of Mathematics*, pages 265–280, 1945.