On the Heavy Tail Behavior of the Distributionally Robust Newsvendor Model with Moment Constraints

March 2018

Anulekha Dhara, University of Michigan
Bikramjit Das, Singapore University of Technology and Design
Karthik Natarajan, Singapore University of Technology and Design

Distributionally Robust Newsvendor (DRN)

- A newsvendor needs to decide on the number of units of an item to order before the actual demand is observed.
- The unit purchase cost is c and the unit revenue is $p > c > 0$.
- Any unsold items at the end have zero salvage value.
- Demand $\tilde{d}$ for the item is random.
- The probability distribution of the demand denoted by $F(\cdot)$ is ambiguous and only assumed to lie a set of possible distributions denoted by $\mathcal{F}$.

$$\text{(DRN)} \quad \max_{q \in \mathbb{R}_+} \inf_{F \in \mathcal{F}} \left(p\mathbb{E}_F[\min(q, \tilde{d})] - cq \right)$$

- Let $\alpha = 1 - c/p \in (0, 1)$ denote the critical ratio and the mean of demand be specified in $\mathcal{F}$.

$$\text{(DRN)} \quad \min_{q \in \mathbb{R}_+} \sup_{F \in \mathcal{F}} \left(\mathbb{E}_F[\tilde{d} - q]^+ + (1 - \alpha)q \right)$$

Scarf's Model

- Set of demand distributions in Scarf's model (1958) is:

$$\mathcal{F}_{1,2} = \left\{ F \in \mathbb{M}(\Re_+) : \int_0^\infty dF(w) = 1, \int_0^\infty w dF(w) = m_1, \int_0^\infty w^2 dF(w) = m_2 \right\}$$

- Given an order quantity $q > m_2/2m_1$, the worst-case demand distribution for $\sup_{F \in \mathcal{F}_{1,2}} \mathbb{E}_F[\tilde{d} - q]_+$ is two-point:

$$\tilde{d}_q^* = \begin{cases} q - \sqrt{q^2 - 2m_1 q + m_2}, & \text{w.p. } \frac{1}{2}\left(1 + \frac{q - m_1}{\sqrt{q^2 - 2m_1 q + m_2}}\right) \\ q + \sqrt{q^2 - 2m_1 q + m_2}, & \text{w.p. } \frac{1}{2}\left(1 - \frac{q - m_1}{\sqrt{q^2 - 2m_1 q + m_2}}\right) \end{cases}$$

- The support points and the probabilities are dependent on q (the power of the adversary).
- Given an order quantity q in the range $0 \leq q \leq m_2/2m_1$, the worst-case distribution is two-point but fixed and given by $\tilde{d}^*_{m_2/2m_1}$.

Scarf's Model

- The worst-case bound is given as:

$$\sup_{F \in \mathcal{F}_{1,2}} \mathbb{E}_f[\tilde{d} - q]_+ = \begin{cases} \frac{1}{2}\left(\sqrt{q^2 - 2m_1 q + m_2} - (q - m_1)\right), & \text{if } q > \frac{m_2}{2m_1} \\ m_1 - \frac{q m_1^2}{m_2}, & \text{if } 0 \leq q \leq \frac{m_2}{2m_1} \end{cases}$$

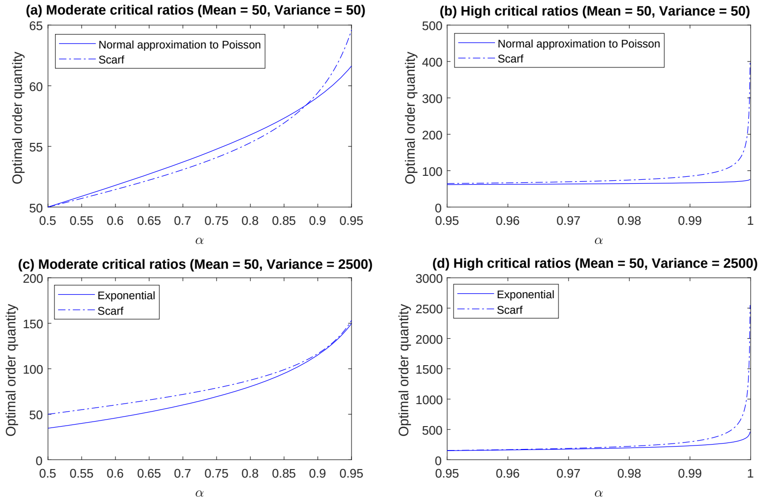
- A closed form solution for the optimal order quantity is as follows:

$$q^* = \begin{cases} m_1 + \frac{\sqrt{m_2 - m_1^2}}{2} \frac{2\alpha - 1}{\sqrt{\alpha(1-\alpha)}}, & \text{if } \frac{m_2 - m_1^2}{m_2} < \alpha < 1 \\ 0, & \text{if } 0 \leq \alpha \leq \frac{m_2 - m_1^2}{m_2} \end{cases}$$

Scarf's Model

- Scarf (1958) observed that for a large range of critical ratios, the optimal order quantity for the two moment model is very close to that of a Poisson distribution while for very high critical ratios, the model prescribes higher order quantities than the Poisson distribution.
- Gallego and Moon (1993) compared the order quantity from Scarf's model with the optimal order quantity for normally distributed demands and concluded numerically that for a large range of critical ratios, the loss is profit is not significant.
- Wang, Glynn and Ye (2015): "In the distributionally robust optimization approach, the worst-case distribution for a decision is often unrealistic. Scarf (1958) shows that the worst-case distribution in the newsvendor context is a two-point distribution. This raises the concern that the decision chosen by this approach is guarding under some overly conservative scenarios, while performing poorly in more likely scenarios. Unfortunately, these drawbacks seem to be inherent in the model choice and cannot be remedied easily."

Scarf's Model



Distributionally Robust Newsvendor (DRN)

- Shapiro and Kleywegt (2002) and Shapiro and Ahmed (2004) show that DRN can be equivalently formulated as a classical newsvendor problem through the construction of a new probability demand distribution.
- Given a set $\mathcal{F}$, there exists a nonnegative random variable $\tilde{d}^*$ with probability distribution F^* such that:

$$\sup_{F \in \mathcal{F}} \mathbb{E}_F[\tilde{d} - q]_+ = \mathbb{E}_{F^*}[\tilde{d}^* - q]_+, \ \forall q \in \Re_+$$

- The right hand side of this equation corresponds to a random variable $\tilde{d}^*$ with a distribution F^* that dominates all the random variables $\tilde{d}$ in the set $\mathcal{F}$ in an increasing convex order sense (see Müller and Stoyan (2002), Shaked and Shanthikumar (2007)).
- However distribution F^* in most cases does not have an explicit characterization and might not even lie in the original set of distributions $\mathcal{F}$.

Revisiting Scarf's Model

- The distribution F^* from Scarf's model is obtained by differentiating the worst-case bound and is given as follows:

$$F^*(q) = \mathbb{P}(\tilde{d}^* \leq q) = \begin{cases} \frac{1}{2}\left(1 + \frac{q - m_1}{\sqrt{q^2 - 2m_1 q + m_2}}\right), & \text{if } q > \frac{m_2}{2m_1} \\ 1 - \frac{m_1^2}{m_2}, & \text{if } 0 \leq q \leq \frac{m_2}{2m_1} \end{cases}$$

- F^* defines a censored student-t random variable with d.o.f 2:

$$\tilde{d}^* = \begin{cases} \tilde{t}_2\left(m_1, (m_2 - m_1^2)/2\right), & \text{if } \tilde{t}_2\left(m_1, (m_2 - m_1^2)/2\right) > \frac{m_2}{2m_1} \\ 0, & \text{otherwise} \end{cases}$$

- The demand distribution in the standard newsvendor model has to possess infinite variance to recreate the solution from Scarf's worst-case model with for all values of α - Müller and Stoyan (2002), Gallego (1998).
- This characterizes a "heavy tail optimality" property of the Scarf's model and explains why the order quantities are higher than that for light-tailed distributions such as normal and exponential for high critical ratios.

Related DRN Models

- Our focus is on sets where demand might take any value in $[0, \infty)$ and the high service level regime. Anderson, Fitzsimons and Simester (2006) provide evidence that a stockout for retailer has significant short and long-term effects and needs to be minimized.
- Closed form solution: Ben-Tal and Hochman (1972, 1976) - mean, mean absolute deviation, Natarajan, Sim and Uichanco (2017) - mean, variance and semivariance.
- Bertsimas and Popescu (2002, 2005) and Lasserre (2002) - SDP techniques to compute the worst-case bounds given:

$$\mathcal{F}_{1,2,\dots,n} = \left\{ F \in \mathbb{M}(\Re_+) : \int_0^\infty dF(w) = 1, \int_0^\infty w^i dF(w) = m_i, \ i = 1, 2, \dots, n \right\}$$

- While some attempts has been made to solve this problem analytically for $n = 3$ and $n = 4$, the tight worst-case bounds have complicated expressions involving roots of cubic and quartic equations (see Jansen, Haezendonck, and Goovaerts (1986) Zuluaga, Peña and Du (2009)).

Related DRN Models

- Additional structural properties such as symmetry and unimodality have been added to the ambiguity sets with moments - SDP and SOCP methods (Popescu (2005), Van Parys, Goulart and Kuhn (2016), Li, Jiang and Mathieu (2017)).
- Lam and Mottet (2017) analyzed the worst-case behavior with information on the tail probability of the random variable for a given threshold, the density function and its left-derivative along with convexity of the tail density function - Worst-case distribution is extremely light tailed or extremely heavy tailed.
- Blanchet and Murthy (2016) computed the worst-case tail probabilities where the ambiguity set is defined as the set of all distributions within a ball of distance δ around a reference distribution using Kullback-Leibler and Renyi divergence measures.
- Ambiguity sets around a reference measure defined with the Wasserstein metric has also been analyzed in the DRN model (see Esfahani and Kuhn (2017) and Gao and Kleywegt (2017)).

Related DRN Models

- A related moment model that has been studied in option pricing but remains unexplored in the newsvendor model (see Grundy (1991)):

$$\mathcal{F}_n = \left\{ F \in \mathbb{M}(\Re_+) : \int_0^\infty dF(w) = 1, \int_0^\infty w^n dF(w) = m_n \right\}$$

- For $n \geq 1$, given a value $q > (n-1)m_n^{1/n}/n$, the worst-case distribution is two-point:

$$\tilde{d}_q^* = \begin{cases} \dfrac{(n-1)q}{n}, & \text{w.p. } \dfrac{(n-1)^n m_n}{n^n q^n} \\ 0, & \text{otherwise} \end{cases}$$

- For $0 \leq q \leq (n-1)m_n^{1/n}/n$, the worst-case demand distribution is degenerate with the mass at the point $m_n^{1/n}$.
- The distribution F^* in this model defines a Pareto random variable:

$$\tilde{d}^* = \text{Pareto}\left(\frac{(n-1)m_n^{1/n}}{n}, n \right)$$

Evidence of Heavy Tailed Demand

Reference	Data	Demand
Clauset, Shalizi & Newman (2009)	24 datasets Two sets - Number of calls, books sold	Power law is a good fit in 17 sets Hard to rule out heavy-tail distributions such as lognormal
Gaffeo, Scorcu & Vici (2008)	Italian books	Power law tail Exponent < 2
Chevalier & Goolsbee (2003)	Sales of books on Amazon.com	Pareto Parameter 1.2
Bimpkis & Markakis (2016)	626 products North American retailer	Pareto Parameter 1.38 Benefits from pooling - low
Natarajan, Sim & Uichanco (2017)	Spare parts	Best fit is often heavy-tailed distributions

Model

- Consider an ambiguity set defined by a known value of the first and the nth moment:

$$\mathcal{F}_{1,n} = \left\{ F \in \mathbb{M}(\Re_+) : \int_0^\infty dF(w) = 1, \int_0^\infty w dF(w) = m_1, \int_0^\infty w^n dF(w) = m_n \right\}$$

- We allow for any real number $n > 1$ (not necessarily an integer).
- Scarf's model: $n = 2$.
- To the best of our knowledge, no closed form solution for $\sup_{F \in \mathcal{F}_{1,n}} \mathbb{E}_F[\tilde{d} - q]_+$ is currently known for general values of n and seems to be hard to find.
- For rational values of n, the DRN problem under this ambiguity set can be solved as a SDP.
- We develop lower and upper bounds that are approximately optimal for large values of q, which helps us characterize F^*.

Main Results

Proposition

Given an ambiguity set $\mathcal{F}_{1,n}$, there exists a positive quantity $\underline{q}(m_1, m_n, n)$ which depends on the parameters m_1, m_n and n, such that:

$$\sup_{F \in \mathcal{F}_{1,n}} \mathbb{E}_F[\tilde{d} - q]_+ \geq \frac{(m_n - m_1^n)}{n^n q^{n-1}}(n-1)^{n-1}, \quad \forall q \geq \underline{q}(m_1, m_n, n).$$

Main Results

Proposition

Consider the ambiguity set $\mathcal{F}_{1,n}$.

(a) *Given $n \in (2, \infty)$, there exists a positive quantity $\overline{q}(m_1, m_n, n)$ which depends on the parameters m_1, m_n and n, such that:*

$$\sup_{F \in \mathcal{F}_{1,n}} \mathbb{E}_F[\tilde{d} - q]_+ \quad \leq \quad \frac{(m_n - m_1^n)}{n^n q^{n-1} - n^2 m_1^{n-1}(n-1)^{n-1}}(n-1)^{n-1} \quad \forall q \geq \overline{q}(m_1, m_n, n).$$

(b) *Given $n \in (1, 2)$, for any $\epsilon > 0$, there exists a positive quantity $\overline{q}(m_1, m_n, n, \epsilon)$ which depends on the parameters m_1, m_n, n and ϵ such that:*

$$\sup_{F \in \mathcal{F}_{1,n}} \mathbb{E}_F[\tilde{d} - q]_+ \quad \leq \quad \frac{(m_n - m_1^n)}{n^n q^{n-1} - n^2 m_1^{n-1}(n-1)^{n-1} - \epsilon}(n-1)^{n-1} \quad \forall q \geq \overline{q}(m_1, m_n, n, \epsilon).$$

Outline of Proof: Lower Bound

- Primal feasible - Consider a three point random variable $\tilde{d}$ with a distribution defined as follows:

$$\tilde{d} = \begin{cases} \dfrac{qn}{n-1}, & \text{w.p. } \dfrac{(m_n - m_1^n)}{n^n q^n}(n-1)^n \\ d, & \text{w.p. } p \\ 0, & \text{w.p. } 1 - p - \dfrac{(m_n - m_1^n)}{n^n q^n}(n-1)^n \end{cases}$$

- Find the values d and p as a function of q such that moment constraints for m_1 and m_n are met.
- Show that for large values of q, the value d is less than q and the corresponding p defines a valid probability measure, using the generalized binomial theorem.

Outline of Proof: Upper Bound

- Dual feasible - The dual formulation is:

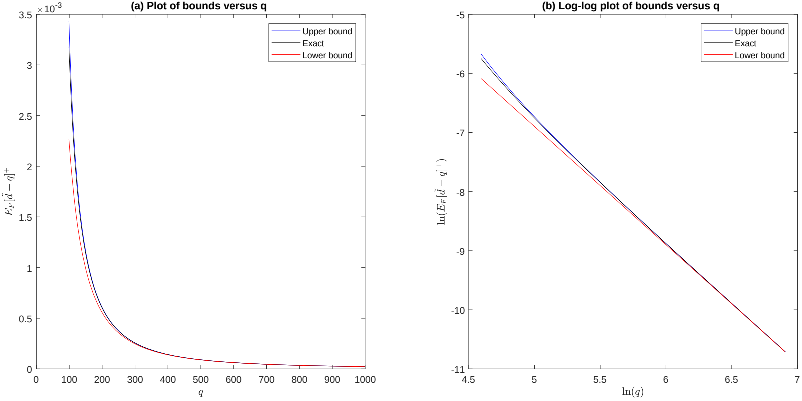
$$\begin{array}{lll} \inf & y_0 + y_1 m_1 + y_n m_n & \\ \text{s.t.} & y_0 + y_1 d + y_n d^n \geq 0, & \forall d \geq 0, \\ & y_0 + y_1 d + y_n d^n \geq d - q, & \forall d \geq 0, \end{array}$$

- Set:

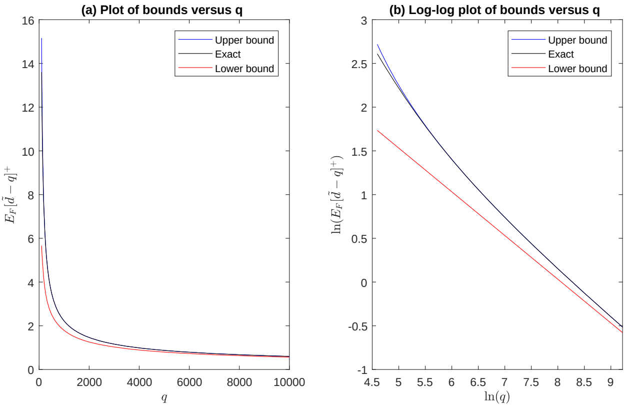
$$\begin{aligned} y_0 &= \frac{(n-1)m_1^n(n-1)^{n-1}}{n^n(q^{n-1}-K)}, \\ y_1 &= \frac{-nm_1^{n-1}(n-1)^{n-1}}{n^n(q^{n-1}-K)}, \\ y_n &= \frac{(n-1)^{n-1}}{n^n(q^{n-1}-K)}, \end{aligned}$$

- Choose K carefully for each of the cases with $n \in (2, \infty)$ and $n \in (1, 2)$ which ensures dual feasibility for large values of q.

Numerical ($n = 3,\ m_1 = 50,\ m_3 = 125150$)



Numerical ($n = 3/2,\ m_1 = 50,\ m_{3/2} = 500, \epsilon = 0.2$)



Main Results

- A function $f : R_+ \to R_+$ is regularly varying at infinity if $\lim_{t\to\infty} f(tx)/f(t) = t^{\beta}$ for some $\beta \in \Re$. We denote this function as $f \in RV_{\beta}$ where β is the index of regular variation.
- A nonnegative random variable with distribution function F is regularly varying if $\bar{F} := 1 - F \in RV_{-\beta}$ for some $\beta \geq 0$.

Theorem

Given an ambiguity set $\mathcal{F}_{1,n}$ and the fixed parameters m_1, m_n and n, as q tends to infinity, the function:

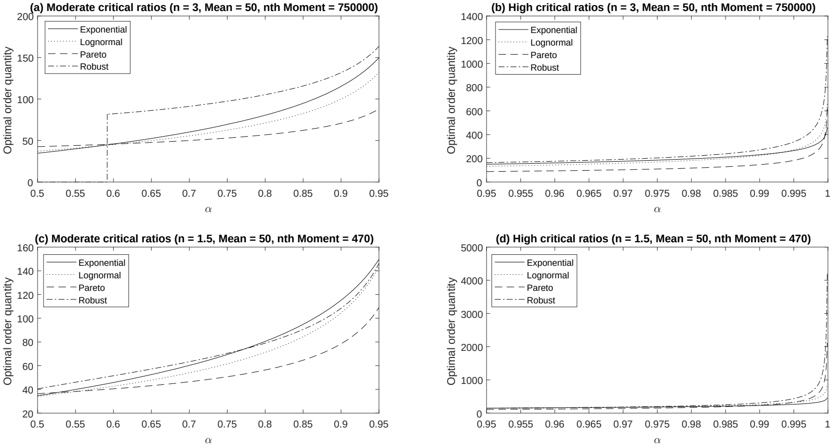
$$\mathbb{E}_{F^*}[\tilde{d} - q]_+ = \sup_{F \in \mathcal{F}_{1,n}} \mathbb{E}_F[\tilde{d} - q]_+ \quad \in \quad RV_{-(n-1)}.$$

An application of the reverse implication of Karamata's theorem implies that:

$$\bar{F}^*(q) \quad \in \quad RV_{-n}.$$

with $\mathbb{E}_{F^*}[\tilde{d}^{*k}] < \infty$ for all $0 \leq k < n$ and $\mathbb{E}_{F^*}[\tilde{d}^{*k}] = \infty$ if $k \geq n$.

Numerical: Optimal Order Quantity



Final Thoughts

- Distributionally robust optimization finds the best decision for the worst distribution. A natural question that researchers have been concerned with is - What are the types of "average" scenarios under which such an optimal decision would also do well?
- For the newsvendor problem, Scarf's solution with first two moment information is optimal for a heavy tailed censored t-distribution with parameter 2 for all critical ratios.
- More generally, with the first and the nth moment given, for high critical ratios, the optimal order quantity from a distributionally robust newsvendor model is optimal for a regularly varying distribution with index $-n$ (power-law tail behavior).