On the Heavy Tail Behavior of the Distributionally Robust Newsvendor

March 2021

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

Distributionally Robust Newsvendor

- 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 assumed to lie a set of possible distributions denoted by $\mathcal{F}$. Solve:

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

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

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

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 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:

$$\sup_{F\in\mathcal{F}_{1,2}} \mathbb{E}_f[\tilde{d}-q]_+ = \begin{cases} \frac{1}{2}\left(\sqrt{q^2-2m_1q+m_2}-(q-m_1)\right), & \text{if } q > \frac{m_2}{2m_1} \\ m_1 - \frac{qm_1^2}{m_2}, & \text{if } 0 \le q \le \frac{m_2}{2m_1} \end{cases}$$

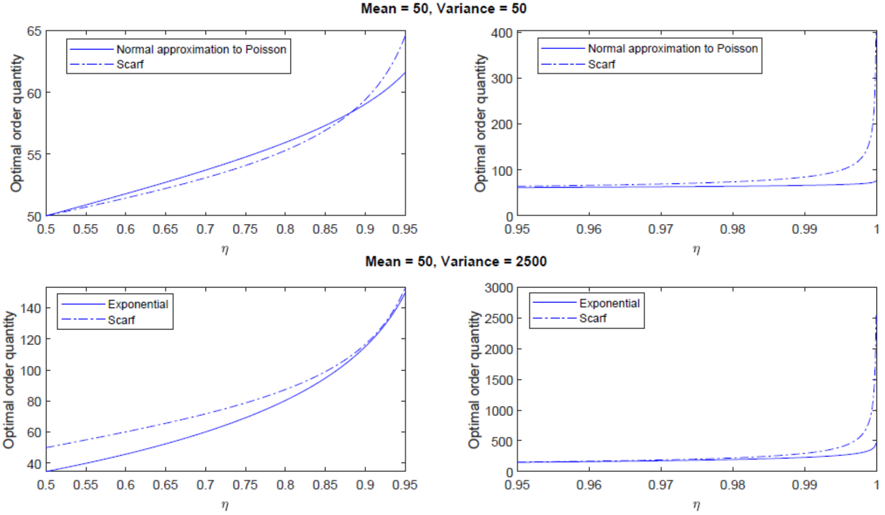
- A closed form solution for the optimal order quantity is:

$$q_\eta^{\text{scarf}} = \begin{cases} m_1 + \frac{\sqrt{m_2-m_1^2}}{2}\frac{2\eta-1}{\sqrt{\eta(1-\eta)}}, & \text{if } \frac{m_2-m_1^2}{m_2} < \eta < 1 \\ 0, & \text{if } 0 \le \eta < \frac{m_2-m_1^2}{m_2} \end{cases}$$

Scarf's Model: Observations

- Scarf (1958) observed that for a large range of critical ratios, the optimal order quantity for the two moment model is close to that of a Poisson distribution while for high critical ratios, the model prescribes higher order quantities than the Poisson distribution. Gallego and Moon (1993) discuss similar findings for normally distributed demands.
- Wang, Glynn and Ye (2015) compared Scarf's model with the optimal order quantity for exponential demand distribution and normally distributed demand with the same mean and variance and found Scarf's solution differed more in comparison to the exponential demand distribution for moderate critical ratios.

Scarf's Model: Observations



Distributionally Robust Newsvendor

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

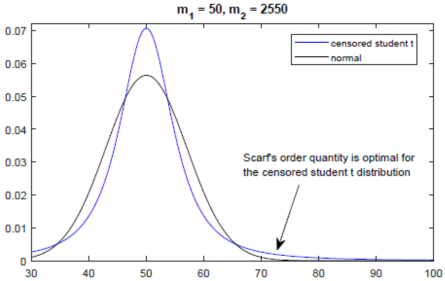
$$\sup_{F \in \mathcal{F}} \mathbb{E}_F[\tilde{d} - q]_+ = \mathbb{E}_{F^*}[\tilde{d}^* - q]_+, \ \forall q.$$

- 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 the distribution F^* in many cases does not have an explicit characterization and might not even lie in the original set of distributions $\mathcal{F}$.

Revisiting Scarf's Model

- Scarf's model: 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 in Scarf's model for every η (see Müller and Stoyan (2002), Gallego (1998)) - "heavy tail optimality" property of Scarf's order quantity.

Related Models

- Our focus is on cases where demand might take any value in $[0, \infty)$ and the high service level regime.
- Closed form solution: Ben-Tal and Hochman (1972, 1976) - mean, mean absolute deviation, Chen, He and Zhang (2011), Natarajan, Sim and Uichanco (2017) - mean, variance and semivariance, Bertsimas, Gupta and Kallus (2018) - robust sample average approximation
- Partial attempts to solve it analytically with third and fourth moments - Jansen, Haezendonck, and Goovaerts (1986) Zuluaga, Peña and Du (2009).
- Bertsimas and Popescu (2002, 2005) and Lasserre (2002) - SDP techniques for ambiguity sets of the form where n is a positive integer:

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

Related 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)).
- Wang, Glynn and Ye (2015) developed a likelihood robust optimization model for this problem that uses convex optimization.
- Lam and Mottet (2017) incorporated 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.
- 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.
- Wasserstein metric based ambiguity set has also been analyzed in this model (see Esfahani and Kuhn (2017) and Gao and Kleywegt (2017)).

Related Models

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

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

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

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

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

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

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)	Automotive demand	Best fit is often heavy-tailed distributions

Model

- Consider an ambiguity set defined by a known value of the first and the αth moment denoted by $\mathcal{F}_{1,\alpha}$.
- Non-parametric statistical tests are available to verify the existence of finite moments (Fedotenkov 2013).
- Primal formulation for worst-case bound:

$$
\begin{aligned}
\sup \quad & \int_0^\infty [w-q]_+ dF(w) \\
\text{s.t.} \quad & \int_0^\infty \mathrm{d}F(w) = 1 \\
& \int_0^\infty w\,\mathrm{d}F(w) = m_1 \\
& \int_0^\infty w^\alpha \mathrm{d}F(w) = m_\alpha \\
& F \in \mathbb{M}(\Re_+)
\end{aligned}
$$

- We allow for any real number $\alpha > 1$ (not necessarily an integer); Scarf's model $\alpha = 2$.

Model

- Dual:

$$\begin{array}{lll} \inf & y_0 + y_1 m_1 + y_\alpha m_\alpha & \\ \text{s.t.} & y_0 + y_1 w + y_\alpha w^\alpha \geq 0 & \forall w \geq 0 \\ & y_0 + y_1 w + y_\alpha w^\alpha \geq w - q & \forall w \geq 0 \end{array}$$

- For rational values of α, the problem can be solved as a SDP.
- For small values of q, we can solve this problem in closed form.
- Given an arbitrary $\alpha > 1$ and $q > 0$, a closed form solution for $\sup_{F \in \mathcal{F}_{1,\alpha}} \mathbb{E}_F[\tilde{d} - q]_+$ appears to be hard to find; under some conditions, you have to solve:

$$\min_{y_1} q \left(\frac{(-y_1)^{\frac{\alpha}{\alpha-1}}}{(1-y_1)^{\frac{\alpha}{\alpha-1}} - (-y_1)^{\frac{\alpha}{\alpha-1}}} \right) + m_1 y_1 + \frac{m_\alpha (\alpha-1)^{\alpha-1}}{\alpha^\alpha q^{\alpha-1}} \left((1-y_1)^{\frac{\alpha}{\alpha-1}} - (-y_1)^{\frac{\alpha}{\alpha-1}} \right)^{\alpha-1}$$

- We develop lower and upper bounds that are approximately optimal for large values of q, which helps us characterize F^*.

Main Results (Lower Bound)

Proposition

Given an ambiguity set $\mathcal{F}_{1,\alpha}$, the following lower bound is valid:

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

where:

$$\underline{q} \quad = \quad \left(\left(\frac{m_\alpha - m_1^\alpha}{m_1} \right) \left(\frac{\alpha - 1}{\alpha} \right)^{\alpha-1} + m_1^{\alpha-1} \right)^{1/(\alpha-1)} .$$

Outline of Proof: Lower Bound

- Unlike trying to find analytical expressions for primal and dual feasible solutions that are optimal, we find "approximately optimal" primal and dual feasible solutions.
- For the lower bound, we extend Grundy's two point distribution to a three point distribution as follows:

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

  where we choose specific values of w and p to ensure this demand satisfies the moment conditions.
- For large values of q, only the first atom can shown to be strictly above q.

Main Results (Upper Bound)

Proposition

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

(a) *When $\alpha \in (2,\infty)$, the following upper bound is valid:*

$$\sup_{F\in\mathcal{F}_{1,\alpha}} \mathbb{E}_F[\tilde{d}-q]_+ \leq \frac{(m_\alpha - m_1^\alpha)}{\alpha^\alpha q^{\alpha-1} - \alpha^2 m_1^{\alpha-1}(\alpha-1)^{\alpha-1}}(\alpha-1)^{\alpha-1}, \forall q > \overline{q},$$

where:

$$\overline{q} \quad = \quad m_1(\alpha-1)\alpha^{(2-\alpha)/(\alpha-1)}.$$

(b) *When $\alpha \in (1,2)$, for all $\epsilon \in (0,(\alpha/(\alpha-1))^{\alpha-1}-\alpha)$, the following upper bound is valid:*

$$\sup_{F\in\mathcal{F}_{1,\alpha}} \mathbb{E}_F[\tilde{d}-q]_+ \leq \frac{(m_\alpha - m_1^\alpha)}{\alpha^\alpha q^{\alpha-1} - (\alpha+\epsilon)\alpha m_1^{\alpha-1}(\alpha-1)^{\alpha-1}}(\alpha-1)^{\alpha-1}, \forall q > \overline{q},$$

where $\overline{q} = m_1(\alpha-1)x/\alpha$ and x is the unique root in the interval $((\alpha+\epsilon)^{1/(\alpha-1)},\infty)$ to the equation:

$$x^\alpha - (\alpha+\epsilon)x + 1 - \left(x^{\alpha-1} - \alpha - \epsilon + 1\right)^{\alpha/(\alpha-1)} \quad = \quad 0.$$

Outline of Proof: Upper Bound

- For the upper bound, we consider the dual formulation:

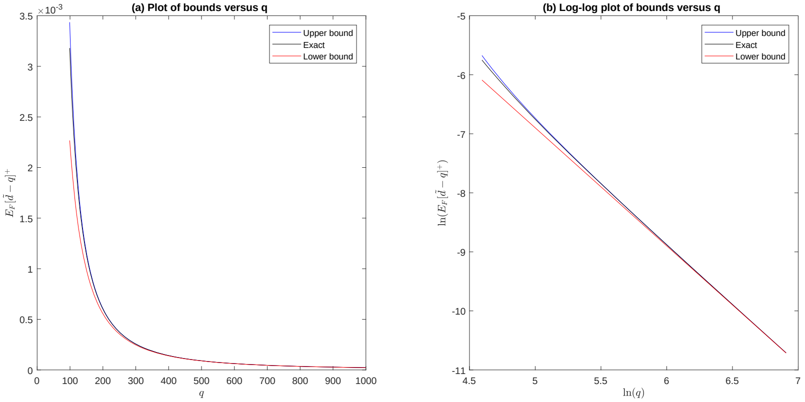
$$\begin{array}{rll} \inf & y_0 + y_1 m_1 + y_\alpha m_\alpha & \\ \text{s.t.} & y_0 + y_1 w + y_\alpha w^\alpha \geq 0, & \forall w \geq 0, \\ & y_0 + y_1 w + y_\alpha w^\alpha \geq w - q, & \forall w \geq 0, \end{array}$$

- Set:

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

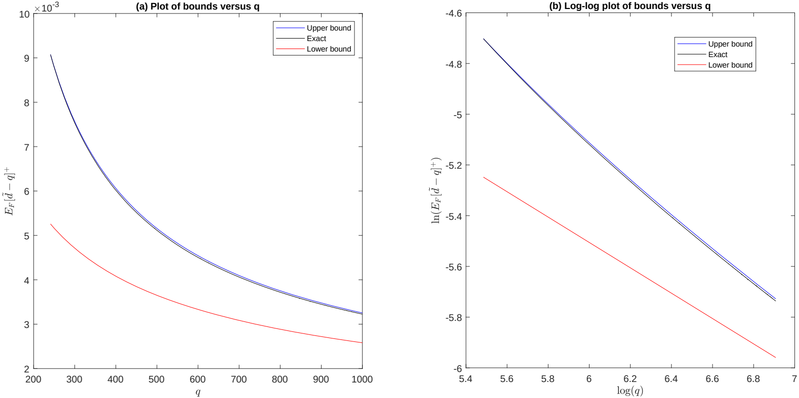
where we choose K carefully for each of the cases with $\alpha \in (2, \infty)$ and $\alpha \in (1, 2)$ which ensures dual feasibility for large values of q.

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



The lower bound is valid for $q > 50.013$ and the upper bound is valid for $q > 57.735$.

Numerical ($\alpha = 3/2$, $m_1 = 50$, $m_{3/2} = \sqrt{125150}$, $\epsilon = 0.1$)



The lower bound is valid for $q > 50.034$ and the upper bound is valid for $q > 240.67$ (as ϵ decreases, $\overline{q}$ increases).

Regularly Varying Functions

- A function $u : \Re_+ \to \Re_+$ is said to be regularly varying at infinity (see Bingham, Goldie and Teugels (1998), Resnick (2008)) if for all $t > 0$, we have:

$$\lim_{x \to \infty} \frac{u(tx)}{u(x)} = t^{\alpha},$$

  for some $\alpha \in \Re$. We denote this function as $f \in RV_{\alpha}$.
- Examples: $u(x) = x^{\alpha}, x^{\alpha} \ln(1 + x)$
- A nonnegative random variable with distribution function F is regularly varying if $\bar{F} := 1 - F \in RV_{-\alpha}$ for some $\alpha \geq 0$.
- Examples: Pareto, Cauchy, Burr, Log-gamma, Student-t with finite d.o.f.

Main Results (Bounds and regularly varying functions)

Theorem

Given the worst-case expected value $\Pi_{1,\alpha}(q) = \sup_{F \in \mathcal{F}_{1,\alpha}} \mathbb{E}_F[\tilde{d} - q]_+$ *where* $\Pi_{1,\alpha} : \Re_+ \to \Re_+$*, the following holds:*

(a) $\Pi_{1,\alpha} \in RV_{-\alpha+1}$.

(b) There exists a random variable $\tilde{d}^*$ *with distribution function* $F^* \notin \mathcal{F}_{1,\alpha}$ *such that*

$$\Pi_{1,\alpha}(q) = \mathbb{E}_{F^*}[\tilde{d}^* - q]_+, \quad \forall q \geq 0.$$

For the distribution we have $\overline{F}^* \in RV_{-\alpha}$ *with* $\mathbb{E}_{F^*}[(\tilde{d}^*)^{\alpha_1}] < \infty$ *for all* $0 \leq \alpha_1 < \alpha$ *and* $\mathbb{E}_{F^*}[(\tilde{d}^*)^{\alpha_1}] = \infty$ *if* $\alpha_1 \geq \alpha$*. Here* F^* *is independent of the choice of* q*.*

Outline of Proof

- For an appropriate constant C_1 with $q > \underline{q}$, we have

$$\Pi_{1,\alpha}(q) \geq C_1 \frac{1}{q^{\alpha-1}}$$

- For an appropriate constant C_2 with $q > \overline{q}$, we have

$$\Pi_{1,\alpha}(q) \leq C_1 \frac{1}{q^{\alpha-1}} \left(1 - \frac{C_2}{q^{\alpha-1}}\right)^{-1}.$$

- For $q > \max(\underline{q}, \overline{q})$,

$$t^{-\alpha+1} \left(1 - \frac{C_2}{q^{\alpha-1}}\right) \leq \frac{\Pi_{1,\alpha}(tq)}{\Pi_{1,\alpha}(q)} \leq t^{-\alpha+1} \left(1 - \frac{C_2}{(tq)^{\alpha-1}}\right)^{-1}.$$

- Apply converse of Karamata's theorem which guarantees preservation of regular variation under differentiation.

Main Results (Bounds and regularly varying functions)

Proposition

Consider the ambiguity set $\mathcal{F}_{1,\alpha}$ with $\alpha > 1$. For $\eta \in [0, 1)$, let q^*_η be an optimal order quantity of the distributionally robust newsvendor and C^*_η be the optimal cost. Then q^*_η is also optimal to a standard newsvendor problem with the underlying demand distribution F^* described in the previous theorem and satisfies the property:

$$q^*_\eta \sim \frac{\alpha - 1}{\alpha} \frac{1}{1 - \eta} C^*_\eta, \quad \text{as} \quad \eta \to 1.$$

A similar characterization between the Value-at-Risk and Conditional Value-at-Risk in risk management for distributions with regularly varying tails has been made in prior work (see Hua and Joe (2011)).

Numerical Results: Solution Method

- For $\alpha = 3$, you need to solve a SDP:

$$
\begin{aligned}
\min_{q,y_0,y_1,y_3,a_1,b_1,c_1,a_2,b_2,c_2} \quad & (1-\eta)q + y_0 + y_1 m_1 + y_3 m_3 \\
\text{s.t.} \quad & \begin{bmatrix} y_0 & 0 & a_1 & b_1 \\ 0 & y_1 - 2a_1 & -b_1 & c_1 \\ a_1 & -b_1 & -2c_1 & 0 \\ b_1 & c_1 & 0 & y_3 \end{bmatrix} \succeq 0, \\
& \begin{bmatrix} y_0 + q & 0 & a_2 & b_2 \\ 0 & y_1 - 1 - 2a_2 & -b_2 & c_2 \\ a_2 & -b_2 & -2c_2 & 0 \\ b_2 & c_2 & 0 & y_3 \end{bmatrix} \succeq 0, \\
& q \geq 0,
\end{aligned}
$$

- In working with larger integers α (or $\alpha = p/q$ with high values of p or q), directly applying the SDP will need the incorporation of many extra variables.

Numerical Results: Solution Method

- Nice trick: Change of variable - $w = e^z$. Divide by e^z to rewrite the problem as a relative entropy optimization problem using a technique from Chandrasekaran and Shah (2016) for dealing with nonnegative signomials with one negative coefficient:

$$
\begin{array}{ll}
\min & (1-\eta)q + y_0 + y_1 m_1 + y_\alpha m_\alpha \\
\text{s.t.} & y_0 e^{-z} + y_\alpha e^{(\alpha-1)z} \geq -y_1, \qquad\qquad \forall z, \\
& (y_0 + q)e^{-z} + y_\alpha e^{(\alpha-1)z} \geq 1 - y_1, \quad \forall z. \\
& q \geq 0.
\end{array}
$$

$$
\begin{array}{ll}
\min_{q,y_0,y_1,y_\alpha,v_1,v_2,v_3,v_4} & (1-\eta)q + y_0 + y_1 m_1 + y_\alpha m_\alpha \\
\text{s.t.} & v_1 \log\left(\frac{v_1}{e y_0}\right) + v_2 \log\left(\frac{v_2}{e y_\alpha}\right) \leq y_1, \\
& v_3 \log\left(\frac{v_3}{e(y_0+q)}\right) + v_4 \log\left(\frac{v_4}{e y_\alpha}\right) \leq y_1 - 1, \\
& (\alpha-1)v_2 = v_1, \\
& (\alpha-1)v_4 = v_3, \\
& q, v_1, v_2, v_3, v_4 \geq 0.
\end{array}
$$

Value of Incorporating Moments Beyond Scarf's Model

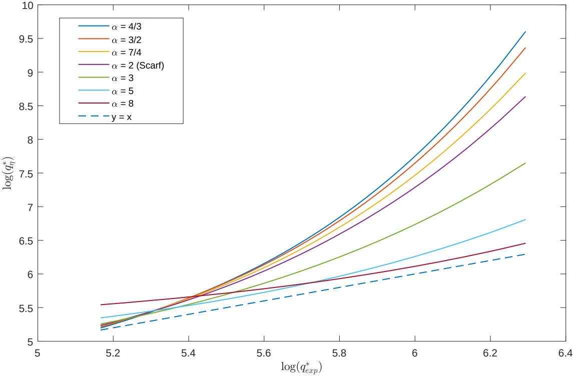
- Three demand distributions that possess different kinds of tail behavior:

  (a) Exponential random variable with mean 50

  (b) Lognormal random variable with parameters $m = \log(50/\sqrt{2})$ and $s = \sqrt{\log(2)}$

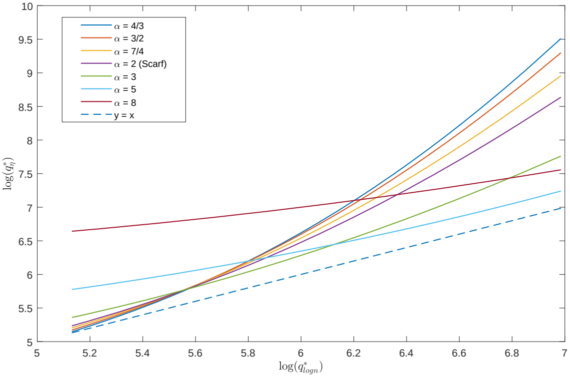
  (c) Pareto random variable with shape parameter $\beta = 1 + \sqrt{2}$ and scale parameter $x_m = 50\sqrt{2}/(1 + \sqrt{2})$.
- Scarf's model prescribes the same optimal order quantity in all the three cases.
- Using the moments of the distributions for other values of α, need not prescribe the same solution.

Value of Incorporating Moments Beyond Scarf's Model



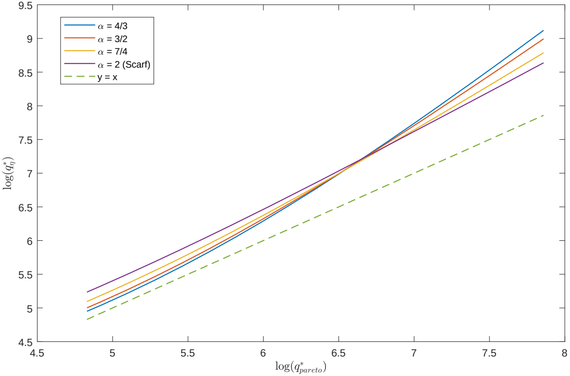
The log-log plot compares the optimal order quantities for the distributionally robust newsvendor with $\alpha = 4/3, 3/2, 7/4, 2, 3, 5, 8$ with the optimal order quantity for the exponential distribution for $\eta \in [0.97, 0.99998]$. As the figure illustrates for larger critical ratios, the knowledge of higher moment information makes the robust model less conservative and closer to the $y = x$ line.

Value of Incorporating Moments Beyond Scarf's Model



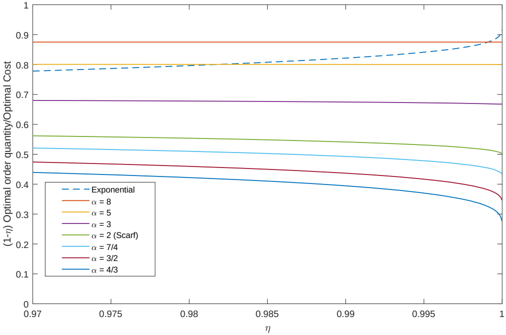
The log-log plot compares the optimal order quantities for the distributionally robust newsvendor with $\alpha = 4/3, 3/2, 7/4, 2, 3, 5, 8$ with the optimal order quantity for the lognormal distribution for $\eta \in [0.97, 0.99998]$. As the figure illustrates for larger critical ratios, the knowledge of higher moment information makes the robust model less conservative and closer to the $y = x$ line. Only for $\alpha = 8$, the line is above the $\alpha = 5$ line for the chosen critical ratios, but the slope indicates that for even higher critical ratios, the robust order quantities for $\alpha = 8$ will get closer to the $y = x$ in comparison to the $\alpha = 5$.

Value of Incorporating Moments Beyond Scarf's Model



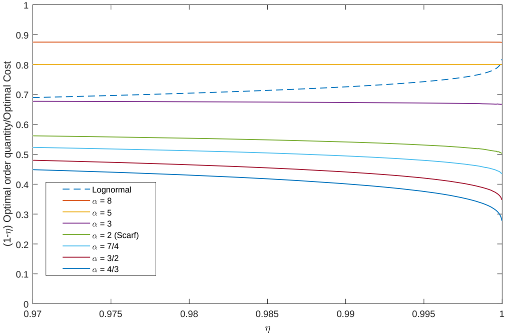
The log-log plot compares the optimal order quantities for the distributionally robust newsvendor with $\alpha = 4/3, 3/2, 7/4, 2$ with the optimal order quantity for the Pareto distribution for $\eta \in [0.97, 0.99998]$. As the figure illustrates for larger critical ratios, the knowledge of higher moment information makes the robust model less conservative and closer to the $y = x$ line.

Scaling of Optimal Order Quantities and Optimal Costs



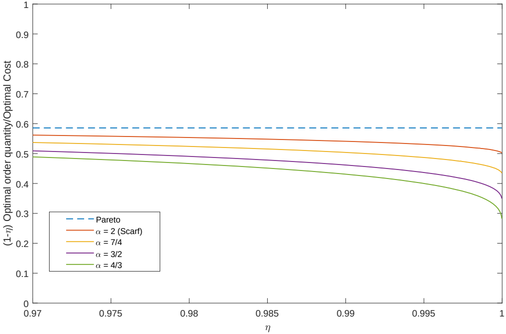
The plot provides the ratios $(1-\eta)q^*_\eta / C^*_\eta$ for the distributionally robust newsvendor model (with different values of α) which converges to $(\alpha - 1)/\alpha$ and the corresponding values for the exponential distribution (which is the dashed line, which tends to 1 as η tends to 1).

Scaling of Optimal Order Quantities and Optimal Costs



The plot provides the ratios $(1-\eta)q^*_\eta/C^*_\eta$ for the distributionally robust newsvendor model (with different values of α) which converges to $(\alpha-1)/\alpha$ and the corresponding values for the lognormal distribution (which is the dashed line which tends to 1 as η tends to 1).

Scaling of Optimal Order Quantities and Optimal Costs



The plot provides the ratios $(1-\eta)q_\eta^*/C_\eta^*$ for the distributionally robust newsvendor model (with different values of α) which converges to $(\alpha-1)/\alpha$ where the moments are obtained from a Pareto distribution. The Pareto distribution is regularly varying and in this case, the ratio is exactly $\sqrt{2}/(1+\sqrt{2}) \approx 0.5858$.

Robustness to Contamination

- Use a mixture of two distributions, F_0 and F_1, where F_0 is the original distribution and F_1 is a heavy-tailed (contamination) distribution parameterized by $\lambda \in [0, 1]$:

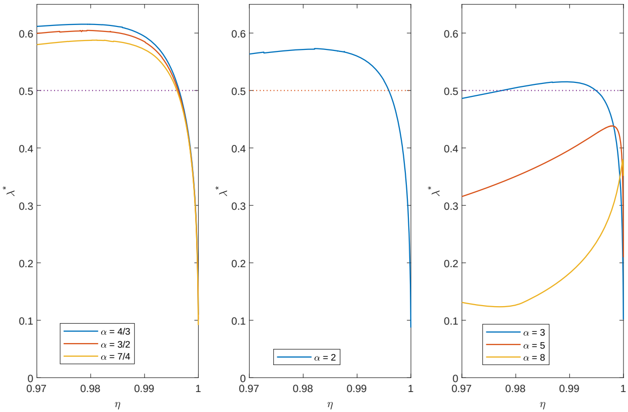
$$F_\lambda \equiv (1 - \lambda)F_0 + \lambda F_1, \quad 0 \leq \lambda \leq 1.$$

- Denote the corresponding optimal order quantities by q_0^* and q_1^* respectively. The order quantities satisfy:

$$C_0(q_0^*) \leq C_0(q_1^*) \quad \text{and} \quad C_1(q_0^*) \geq C_1(q_1^*).$$

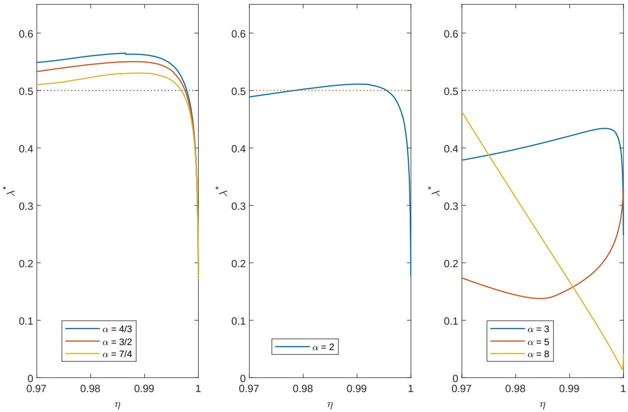
- λ^* - Value beyond which the robust order quantity outperforms the classical solution under contamination
- $\lambda^* < 0.5$, less than 50% contamination, the robust order quantity outperforms the standard newsvendor order quantity
- $\lambda^* > 0.5$, more than 50% contamination is needed for the robust order quantity to outperform the standard newsvendor order quantity.

Robustness to Contamination



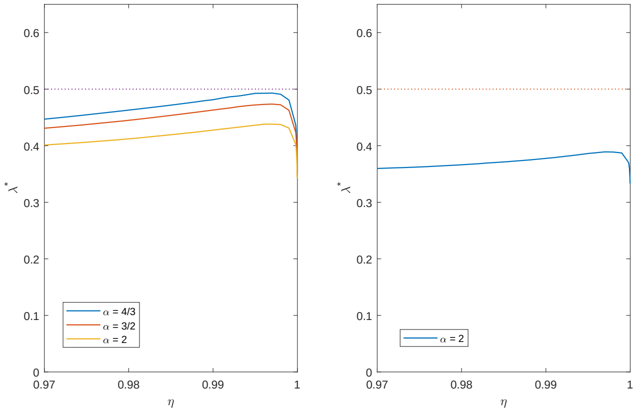
The plot provide the λ^* values for the case where the original distribution is exponential. In each of the cases, we see that as the critical ratio approaches 1, the value of λ^* rapidly drops to 0. This indicates that a small amount of contamination is sufficient for the robust solution to outperform the classical solution for high service levels.

Robustness to Contamination



The plot provide the λ^* values for the case where the original distribution is lognormal. In each of the cases, we see that as the critical ratio approaches 1, the value of λ^* rapidly drops to 0. This indicates that a small amount of contamination is sufficient for the robust solution to outperform the classical solution for high service levels.

Robustness to Contamination



The plot provide the λ^* values for the case where the original distribution is Pareto. In each of the cases, we see that as the critical ratio approaches 1, the value of λ^* rapidly drops to 0. This indicates that a small amount of contamination is sufficient for the robust solution to outperform the classical solution for high service levels.

Final Thoughts

- How do you generalize results/approach to other ambiguity sets? for multivariate contexts? for other objective functions? characterizing the body of the distribution F^*?