On The Relationship Between Entropy, Demand Uncertainty, and Expected Loss

Adam J. Fleischhacker
Department of Business Administration, University of Delaware, Newark, DE 19716, 302.831.6966, ajf@udel.edu

Pak-Wing Fok
Department of Mathematics, University of Delaware, Newark, DE 19716, pakwing@udel.edu

We analyze the effect of demand uncertainty, as measured by entropy, on expected costs in a stochastic inventory model. Existing models studying demand variability's impact use either stochastic ordering techniques or use variance as a measure of uncertainty. Due to both axiomatic appeal and recent use of entropy in the operations management literature, this paper develops entropy's use as a demand uncertainty measure. Our key contribution is an insightful proof quantifying how costs are non-increasing when entropy is reduced.

Key words: Entropy, Measuring Demand Uncertainty, Stochastic Inventory Models, Demand Variability

1. Background

Entropy is a measure of uncertainty strongly advocated by Jaynes (2003) and originally popularized by Shannon (1948). Let D be a discrete random variable with probability mass function $\mathbf{p} = (p_1, ..., p_N)$, then entropy $H(\mathbf{p})$ is defined as

$$H(p_1, ..., p_n) = -\sum_{i=1}^{n} p_i \log(p_i). \tag{1}$$

A good introduction to entropy as a measure of uncertainty can be found in Abbas (2006). Maximizing Eq. (1) subject to constraints based on existing knowledge (e.g. the mean, support, moments, etc.) is known as the maximum entropy principle (Jaynes 1957), and is considered a uniquely correct method for inductive inference (Shore and Johnson 1980, Johnson and Shore 1983).

Our work in this paper solidifies the theoretical connection between entropy as a measure of demand uncertainty and expected loss. Despite the concept of entropy being around for over 60 years, the operations management literature has been slow to adopt this uncertainty measure. Eren and Maglaras (2009) comment that "to the best of our knowledge, the operations management and revenue management literatures have not explored the use of ME (maximum entropy) techniques to approximate unknown demand or willingness-to-pay distributions." Recently, however, entropy is being explored. References to entropy within both operations management (see for example Shuiabi et al. 2005, Perakis and Roels 2008, Eren and Maglaras 2009, Andersson et al. 2011) and related contexts are increasing. This includes the contexts of pricing models (Lim and Shanthikumar

2007), portfolio optimization (Glasserman and Xu 2013), and discrete optimization (Nakagawa et al. 2013). Of particular interest, Andersson et al. (2011) numerically demonstrate promising performance characteristics of using entropy-based demand distributions for ordering decisions; our work complements this insight by theoretically connecting entropy and loss in a similar setting.

2. Entropy Versus Alternative Demand Uncertainty Measures

Existing theoretical connections between uncertainty and supply/demand mismatch costs are often based on a measure of spread or on stochastic ordering techniques. In this work we use entropy to connect uncertainty and expected loss in a way that is intuitively satisfying, facilitates numerical evaluation of uncertainty's effects on mismatch costs, and is also theoretically justified. In addition, it enables the exploration of uncertainty reduction without restricting expected demand to remain constant as uncertainty is increased/decreased. For example, a retailer reducing demand uncertainty by transitioning from hi-lo pricing to everyday low pricing will most certainly see a change in expected demand (see discussion in Lee et al. 1997).

Variance is arguably the most preferred measure of demand uncertainty found in the inventory management literature (see for example Ridder et al. 1998, Taylor and Xiao 2010, Kwak and Gavirneni 2011). While intuitively, variance provides meaning regarding the spread of a distribution and hence, uncertainty; problems arise when one tries to operationalize variance as a measure of uncertainty (Jaynes 2003, p.345). In an often cited example, imagine a six-sided die with a known bias such that the expected value of a roll is 4.5 instead of 3.5. How can we assign probabilities to the six outcomes in this clearly under-specified problem? Following Jaynes' logic, we should pick the probability assignment implying maximal uncertainty subject to the problem's constraints. In other words, if every feasible probability assignment were to be evaluated by a numerical measure of uncertainty, then one chooses the assignment that has the largest measurement; to choose otherwise would imply additional information beyond that which is available. Thus, applying variance as an uncertainty measure to the six-sided die problem, assign probability, p_i, by maximizing $\mathrm{Var}(\{p_i\}) = \sum_{i=1}^{6} p_i(i-9/2)^2$ subject to $\sum_{i=1}^{6} p_i = 1$ and $\sum_{i=1}^{6} ip_i = 9/2$. Solving this maximization problem, all probability is placed on the outcomes of one and six with $p_1 = 0.3$ and $p_6 = 0.7$. Unfortunately, this extreme placement of probabilities leads to dissatisfaction with the implied results. Shouldn't some probability remain on outcomes of 2, 3, 4, and 5? Variance as an uncertainty measure leads to these counter-intuitive results and thus, maximal uncertainty and maximal variance are not equal.

More intuitive results are found when entropy is used in place of variance. Mathematically, we maximize $-\sum_{i=1}^{6} p_i \log p_i$ subject to $\sum_{i=1}^{6} p_i = 1$ and $\sum_{i=1}^{6} ip_i = 9/2$. The Lagrange function is

$$\mathcal{L} \equiv -\sum_{i=1}^{6} p_i \log p_i - \lambda_1 \left(\sum_{i=1}^{6} p_i - 1 \right) - \lambda_2 \left(\sum_{i=1}^{6} ip_i - 9/2 \right), \tag{2}$$

where λ_1 and λ_2 are lagrange multipliers and log is always the natural logarithm in this paper. Maximizing $\mathcal{L}$, we find maximum uncertainty corresponds to $p_1 \approx 5.4\%$, $p_2 \approx 7.9\%$, $p_3 \approx 11.4\%$, $p_4 \approx 16.5\%$, $p_5 \approx 24.0\%$, and $p_6 \approx 34.7\%$. Outcomes of 2 through 5 are no longer an impossibility and much more in line with what common sense might dictate.

Besides entropy and variance (see Ebrahimi et al. 1999, for further comparison of entropy and variance), other alternative uncertainty measures exist. A popular alternative to study the effects of demand variability is to use stochastic ordering criteria as done in Song (1994), Jemaï and Karaesmen (2005), Song et al. (2010), and Xu et al. (2010). In all of these works, the authors confirm the intuition that uncertainty generally leads to increased costs. Though, specific cases where larger uncertainty is associated with reduced costs can also be found (Ridder et al. 1998). The previously mentioned works use of stochastic ordering lead to qualitative insight, but do not facilitate numerical connection of uncertainty and mismatch costs. Gerchak and Mossman (1992) enable more numerically driven computations involving uncertainty through use of the mean preserving transformation. Unfortunately, the mean preserving transformation, like the previously mentioned stochastic ordering techniques, assumes uncertainty reduction never leads to changes in expected demand. Common sense requires expected demand changes may indeed be a possible outcome of a demand uncertainty reduction effort; otherwise, why do it? Hence, the extant demand variability literature is largely void of a method of measuring uncertainty that enables us to numerically relate uncertainty and expected mismatch costs, is consistent with common sense, and provides quantifiable evaluation of uncertainty's effect on expected loss.

Risk measures such as VaR (Value-at-Risk) and ES (Expected Shortfall) are common concepts in asset management and portfolio theory. The VaR_q is calculated so that with probability q (sometimes called a confidence level), the portfolio's value will not exceed the VaR. For example, the $q = 95\%$ VaR is only exceeded by a portfolio of assets 5% of the time. The expected shortfall is an average of the VaR_q, restricted to values of q close to 1. To calculate these risk measures, one starts with a probabilistic model for the assets and computes the likelihood of the portfolio having a certain value, thereby quantifying the risk. By treating the stochastic inventory problem with portfolio theory, we can highlight the differences between risk measures and entropy. Given $\{p_i\}$ that specify the stochasticity in demand, underage/overage costs and an ordering strategy, the VaR and ES could be computed. However, the demand entropy is simply $-\sum_i p_i \log p_i$. It is independent of underage/overage costs and ordering strategy. Rather than giving a dollar-amount estimate for the inventory's value (with a certain probability), entropy only quantifies uncertainty in the demand for stock. It is more similar to the confidence level 'q' in VaR_q, rather than the VaR itself: entropy reflects the newsvendor's uncertainty in the same way that the q in VaR_q reflects the portfolio manager's confidence level.

Our use of entropy and its comparison to variance connotes mathematical risk. However, entropy is more a measure of randomness than risk; its value is independent of preferences among potential outcomes of demand (Szegö 2005). This independence property proves desirable as a decision maker achieves less uncertainty through information acquisition efforts (e.g. focus groups, expert opinion, etc.) and not better demand outcomes. A reduction in a risk measure, on the other hand, would quantify a reduction in the "likelihood of loss or less than expected returns" (see McNeil et al. 2010, Ch. 1). Our work here connects an uncertainty measure to expected loss and hence, its main purpose is to formalize the intuitive relationship that uncertainty tends to lead to less desirable outcomes. Our goal is to facilitate information valuation, whereas a risk-centric focus in this setting is more geared towards optimizing ordering decisions (see related discussion in Choi et al. 2011, and refences therein).

3. Entropy and Expected Loss

In this paper, we quantify the relationship between expected supply/demand mismatch costs and entropy in the context of a general stochastic inventory problem. We commence our analysis defining maximal uncertainty and subsequently examine the effects of reducing uncertainty. Consider the loss matrix $\mathbf{A} \in \mathbb{R}^{+N\times N}$ whose elements represent the expected loss associated with N possible ordering decisions and N possible outcomes (e.g. demand realizations). An ordering decision $j \in \{1,2,\ldots,N\}$ is chosen such that the loss, $\mathcal{L}_j(\mathbf{p})$, is minimized

$$\mathcal{L}_{\min}(\mathbf{p}) = \min_{1\le j\le N} \mathbf{q}_j^T \mathbf{A}\mathbf{p}, \tag{3}$$

where demand distribution $\mathbf{p} = (p_1, p_2, \ldots, p_N)$ is unknown, and $\mathbf{q}_j$ is the jth unit (column) vector representing an order quantity of j units.

The determination of the optimal order quantity j^* and associated loss $\mathcal{L}_{j^*}(\mathbf{p})$ requires specification of the probability distribution $\mathbf{p}$. Since, we do not know $\mathbf{p}$ and any of an infinite number of N-tuple probabilities $(p_1, p_2, \ldots, p_N)$ may be valid representations of $\mathbf{p}$, we find the $\mathbf{p}$ consistent with maximal uncertainty via the principle of maximum entropy: maximize $\sum_{i=1}^N p_i \log p_i$, subject to $p_1 + p_2 + ... + p_N = 1$, with the unique solution $p_1 = p_2 = ... = p_N = 1/N$. The corresponding expected loss is $\mathcal{L}_{\min}(\mathbf{e}/N) = \mathbf{q}_{j^*}^T \mathbf{A}\mathbf{e}/N$ where $\mathbf{e}$ is the vector consisting of all 1s.

With maximal uncertainty (i.e. entropy) defined, we can now value expected loss at reduced levels of entropy. To relate uncertainty to loss, we define

$$S(h) \equiv \left\{ (p_1, \ldots, p_N) : H(\mathbf{p}) = h \ , \ \sum_{i=1}^N p_i = 1 \ , \ 0 < p_j < 1, j = 1, \ldots, N \right\}, \tag{4}$$

where $H(\mathbf{p}) = H(p_1, \ldots, p_N) = -\sum_{i=1}^N p_i \log p_i$. For a given entropy level h, Eq. (4) yields all N-tuples $(p_1, p_2, \ldots, p_N)$ that are consistent with this level of uncertainty.

Besides axiomatic derivation as a measure of uncertainty, entropy may also be derived via combinatorial arguments (Niven 2007). Hence, all N-tuple's of equal entropy are considered equiprobable. The expected loss subject to the constant entropy constraint $H(p_1, p_2, \ldots, p_N) = h$ is therefore an average over all possible N-tuples that lie on S:

$$\mathcal{E}[h] \equiv E[\mathcal{L}_{\min}(\mathbf{p}); S(h)] = \frac{\int_{S(h)} \mathcal{L}_{\min}(\mathbf{p}) \mathrm{d}s}{\int_{S(h)} \mathrm{d}s}. \tag{5}$$

Note that (5) is actually an expectation of expected losses.

Some isoentropy surfaces for the $N = 3$ case are shown in Figure 1. For the general N-decision case, the constant h isoentropy surfaces are $N-2$ dimensional hypershells lying in the $N-1$ dimensional simplex $\sum_{i=1}^{N} p_i = 1$, $0 < p_i < 1$. In Figure 1(a) we see that for values of h near the maximal value $\log 3$, the isoentropy contour is approximately circular, but for smaller values, they become more triangular. The contours are also geometrically symmetric in p_1, p_2 and p_3, as expected.

In Figure 1(b), we show that for different values of h, the isoentropy contour can pass through different "loss regions" (separated by the thin red lines) with different optimal order quantities (OQs). Loss region D_2 consists of all 3-tuples (p_1, p_2, p_3) where an OQ of 2 gives the smallest loss. In loss regions D_1 and D_3, OQ of 1 and 3 give the smallest loss. When $h = \log 3$ – corresponding to maximal uncertainty – the optimal decision would be to order 2 units. However, after uncertainty reduction efforts occur so that $h = 1.07$ (say), the potential probability distribution consistent with that given level of uncertainty may be any point on the dotted iso-entropy contour. Since, the entire iso-entropy curve lies completely in region D_2 where ordering 2 units is optimal, then any money or effort spent on reducing uncertainty from $h = \log 3$ to $h = 1.07$ is essentially wasted as the optimal OQ remains unchanged. However, when h is further decreased to $h = 1$, there are points on the contour corresponding to a change in decision. As the contour further expands (as h decreases) and passes through different loss regions, *on average*, the expected loss decreases. We formalize this crucial relationship between uncertainty (characterized by h) and expected loss in Theorem 1, §4. However, before we prove the main theorem, we prove some results concerning the geometry of isoentropy surfaces and set up some relevant notation pertinent to the theorem. We prove

LEMMA 1. *Let $S(h)$ be defined in (4). Then S has the following properties:*

1. *When $h = \log N$, $S = \{\mathbf{e}/N\}$.*
2. *For a point $\mathbf{p} = (p_1, \ldots, p_N) \in S$, the normal vector of S at $\mathbf{p}$ is proportional to*

$$\mathbf{r} \equiv \frac{1}{N} \left(\sum_{i=1}^{N} \log p_i \right) \mathbf{e} - (\log p_1, \log p_2, \ldots, \log p_N)^T. \tag{6}$$

Proof Proofs for all lemmas are provided in the appendix.

We now define a <u>loss region</u>. The aim is to partition the "belief space" $[0,1]^N$ into N loss regions where the index of the region corresponds to the optimal order quantity.

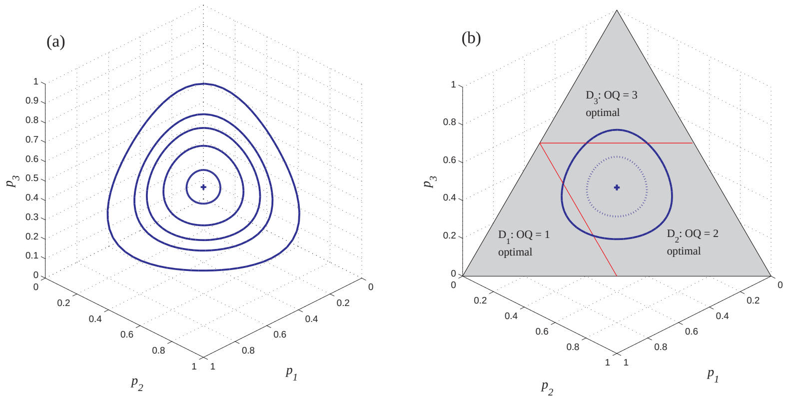

Figure 1 **(a) Isoentropy contours for** $h = \log 3, 1.09, 1.05, 1.0, 0.95$ **and** 0.8**, as defined by Eq. (4). (b) Loss regions (for a symmetric absolute linear loss function) and isoentropy contours for** $h = \log 3$ **(cross),** $h = 1.07$ **(dotted curve) and** $h = 1$ **(thick solid curve). OQ = Order Quantity.**

DEFINITION 1. For a given loss matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$, let $\mathbf{a}_i^T$ be the ith row. Then the loss regions are defined recursively as

$$D_1 = \left\{\mathbf{p} \mid \mathbf{a}_1^T\mathbf{p} \leq \mathbf{a}_m^T\mathbf{p}, \quad 1 < m \leq N\right\}, \tag{7}$$

$$D_j = \left\{\mathbf{p} \notin D_1 \cup \ldots \cup D_{j-1} \text{ and } \mathbf{a}_j^T\mathbf{p} \leq \mathbf{a}_m^T\mathbf{p}, \quad 1 \leq m \leq N, m \neq j\right\}, \qquad j = 2, \ldots, N. \tag{8}$$

LEMMA 2. *The "belief space"* $\mathbf{p} = (p_1, p_2, \ldots, p_N) \in [0,1]^N$ *is partioned into* N *loss regions* $D_1, D_2, \ldots, D_N$ *where the* D_j *are defined in (7) and (8).*

COROLLARY 1. *Within region* D_j*, the decision to order* j *units gives the smallest loss (though not necessarily uniquely):*

$$\mathcal{L}_{min}(\mathbf{p}) = \min_{1\leq k\leq N} \mathcal{L}_k(\mathbf{p}) = \mathbf{a}_j^T\mathbf{p} \qquad \textit{when} \qquad \mathbf{p} \in D_j, \tag{9}$$

4. N-DECISION CASE: CONNECTING ENTROPY AND EXPECTED LOSS

The following theorem formalizes how entropy affects expected loss and confirms that expected loss, over all possible demand distributions representing a given level of uncertainty, is non-increasing as uncertainty is reduced.

THEOREM 1. *Consider the* N*-decision problem where the rows of the loss matrix* $\mathbf{A} \in \mathbb{R}^{N\times N}$ *are* $\mathbf{a}_i^T$*,* $i = 1, \ldots, N$*. Then the expected loss*

$$\mathcal{E}[h] \equiv E[\mathcal{L}_{min}(\mathbf{p}); S(h)] = \frac{\int_{S(h)} \mathcal{L}_{min}(\mathbf{p})\, ds}{\int_{S(h)} ds}, \quad \mathcal{L}_{min}(\mathbf{p}) = \min_{1\leq j\leq N} \mathbf{a}_j^T\mathbf{p}, \tag{10}$$

satisfies

$$\mathcal{E}'[h] \geq 0, \ \textit{when} \qquad 0 < h < \log N, \tag{11}$$

where $S(h)$ *is defined in (4).*

Proof of Theorem 1 From lemma 1 part 1, when $h \to \log N^-$, $S \to \{\mathbf{e}/N\}$. Let J be the (unique) loss region containing $\mathbf{e}/N$. For this belief, the optimal decision is to order J units:

$$\mathbf{a}_J^T\mathbf{e} \le \mathbf{a}_j^T\mathbf{e}, \qquad j = 1, \ldots, N. \tag{12}$$

Furthermore, for $\mathbf{p} \in D_i$, $i \neq J$, an order quantity of i units gives a loss smaller than or equal to an order of J units:

$$\mathbf{p} \in D_i \Leftrightarrow \mathbf{w}_i^T\mathbf{p} \le 0, \qquad \mathbf{w}_i \equiv \mathbf{a}_i - \mathbf{a}_J. \tag{13}$$

Now we divide our analysis into two cases. Either $S(h)$ lies entirely within D_J or it lies in ≥ 2 different loss regions.

In the first case, $\mathcal{E}[h] = \frac{\mathbf{a}_J^T \int_{S(h)} \mathbf{p}\mathrm{d}s}{\int_{S(h)} \mathrm{d}s}$. Geometrically, $\int_{S(h)} \mathbf{p}\mathrm{d}s / \int_{S(h)} \mathrm{d}s$ is the center of mass of an $(N-2)$-dimensional shell of points in $S(h)$. Because the set $S(h)$ is symmetric in the probabilities $p_1, \ldots, p_N$ and $\sum_{i=1}^N p_i = 1$, we must have $\int_{S(h)} \mathbf{p}\mathrm{d}s / \int_{S(h)} \mathrm{d}s = \mathbf{e}/N$ for all h such that $S(h)$ lies entirely within D_J. Therefore $\mathcal{E}[h] = \mathbf{a}_J^T\mathbf{e}/N \Rightarrow E'[h] = 0$ and the theorem is proved.

In the second case, $S(h)$ is not contained only in D_J. Instead, different parts of S lie in different loss regions. The indices of these loss regions form a subset of the integers $I \subseteq \{1, 2, \ldots, N\}$. Defining $S_j(h) = \{\mathbf{p} : \mathbf{p} \in S \cap D_j\}$, $j \in I$, we can partition $S(h)$ as $S(h) = \bigcup_{j\in I} S_j(h)$. Therefore, we have

$$\begin{aligned}
\mathcal{E}[h] &= \frac{\sum_{j\in I} \int_{S_j} \mathbf{a}_j^T \mathbf{p}\mathrm{d}s}{\int_S \mathrm{d}s}, \\
&= \frac{\int_{S_J} \mathbf{a}_J^T\mathbf{p}\mathrm{d}s}{\int_S \mathrm{d}s} + \frac{\sum_{j\in I\setminus\{J\}} \int_{S_j} \mathbf{a}_J^T\mathbf{p}\mathrm{d}s}{\int_S \mathrm{d}s} - \frac{\sum_{j\in I\setminus\{J\}} \int_{S_j} \mathbf{a}_J^T\mathbf{p}\mathrm{d}s}{\int_S \mathrm{d}s} + \frac{\sum_{j\in I\setminus\{J\}} \int_{S_j} \mathbf{a}_j^T\mathbf{p}\mathrm{d}s}{\int_S \mathrm{d}s}, \\
&= \frac{\int_S \mathbf{a}_J^T\mathbf{p}\mathrm{d}s}{\int_S \mathrm{d}s} + \frac{\sum_{j\in I\setminus\{J\}} \int_{S_j} (\mathbf{a}_j - \mathbf{a}_J)^T\mathbf{p}\mathrm{d}s}{\int_S \mathrm{d}s}, \\
&= \mathbf{a}_J^T\mathbf{e}/N + \frac{\sum_{j\in I\setminus\{J\}} \int_{S_j(h)} \mathbf{w}_j^T\mathbf{p}\mathrm{d}s}{\int_{S(h)} \mathrm{d}s},
\end{aligned} \tag{14}$$

using the definition of $\mathbf{w}_j$ in (13). We can see from Eq. (14) that any deviation in $\mathcal{E}[h]$ from $\mathbf{a}_J^T\mathbf{e}/N$ arises from parts of the isosurface lying outside of D_J. Upon taking the derivative,

$$\mathcal{E}'[h] = \frac{\sum_{j\in I\setminus\{J\}} \int_{S_j(h)} \beta\hat{\mathbf{n}} \cdot \nabla(\mathbf{w}_j^T\mathbf{p})\mathrm{d}s}{\int_{S(h)} \mathrm{d}s}, \tag{15}$$

where $\hat{\mathbf{n}} = \mathbf{r}/|\mathbf{r}|$ is the unit normal in the direction of increasing h on $S(h)$ and $\mathbf{r}$ is defined in Eq. (6). In the appendix, we give an explicit derivation of (15) and the definition of β. We also show that $\beta > 0$. Since $S(h)$ only includes points $0 < p_j < 1$, $j = 1, \ldots, N$, $\mathbf{r}$ is always well-defined for all points on S.

Consider any of the integrands in (15) and note that $\hat{\mathbf{n}} \cdot \nabla(\mathbf{w}_j^T\mathbf{p})$ and $\mathbf{r} \cdot \nabla(\mathbf{w}_j^T\mathbf{p})$ have the same sign. We now show that $\mathbf{r} \cdot \nabla(\mathbf{w}_j^T\mathbf{p})$ is positive for any j. Since the denominator in (15) is always positive and $\beta > 0$, this is equivalent to showing that $\mathcal{E}'[h] > 0$.

Let w_{jk} be the kth component of $\mathbf{w}_j$ where $\mathbf{w}_j$ is defined in (13) and let $\mathbf{p} \in S_j$. Then

$$\begin{aligned}\mathbf{r}\cdot\nabla(\mathbf{w}_j^T\mathbf{p}) &= -\sum_{i=1}^{N}\log(p_i/\mu)w_{ji}, \qquad \mu=\left(\prod_{j=1}^{N}p_j\right)^{1/N},\\ &\geq \sum_{k=1}^{N}(1-p_i/\mu)w_{ji},\\ &= \sum_{i=1}^{N}w_{ji}-\frac{1}{\mu}\sum_{i=1}^{N}w_{ji}p_i,\\ &\geq 0,\end{aligned}$$

using $\sum_{i=1}^{N} w_{ji} = (\mathbf{a}_j - \mathbf{a}_J)^T\mathbf{e} \geq 0$ from (12), $-\sum_{i=1}^{N} w_{ji}p_i \geq 0$ from (13) and the fact that the geometric mean of the probabilities $\mu > 0$. Hence $E'[h] \geq 0$ and the theorem is proved. Q.E.D.

The intuitive notion that uncertainty is the driver of expected loss in this context is now mathematically demonstrated. The proof makes few restrictions on the loss matrix. As such, it validates the appeal of entropy as a measure of uncertainty beyond the context of the demand uncertainty reduction studied here. In addition, there is a subtle implication contained within the proof that is worth pointing out. When starting at maximal uncertainty, the expected loss does not decrease immediately as one becomes more certain. One must sufficiently reduce uncertainty to a point where the loss-minimizing decision may potentially change before any benefits of uncertainty reduction can be expected. In common sense terms, do not pursue uncertainty reduction if it is not going to change your decision.

Theorem 1 establishes a link between a given level of uncertainty (entropy h) and the expectation of expected loss over a belief space (the surface $S(h)$). This is in contrast to linking uncertainty and expected loss for a particular N-tuple of probabilities. As is the case with variance or coefficient of variation as uncertainty measurements (see examples in Ridder et al. 1998, Jemaï and Karaesmen 2005), entropy can be reduced and lead to an increased expected loss. For example, assume an $N = 3$ problem with the following loss matrix:

$$A_{ij} = \begin{cases} c_o(i-j), & \text{if } j < i, \\ c_u(j-i), & \text{if } j \geq i, \end{cases} \tag{16}$$

where c_o and c_u represent the per unit overage and underage costs, respectively, for the standard newsvendor problem (Porteus 2002). A reduction in entropy occurs if $p_1 = 0.3, p_2 = 0.3, p_3 = 0.4$ are new beliefs compared to the maximum entropy belief of $p_1 = p_2 = p_3 = 1/3$. However, the expected loss increases from \$33.33 to \$35.00 **Adam: what are the overages and underages for these dollar amounts?**. This underscores the point that Theorem 1 refers to the expectation of loss over *all* N-tuples corresponding to a given level of uncertainty.

5. CONCLUDING REMARKS

This paper explores the use of entropy as a measurement of demand uncertainty and places the relationship between demand uncertainty and expected loss on more firm theoretical footing. In developing the main theorem of this paper, intuition about uncertainty is developed and some advantages of entropy as compared to variance as an uncertainty measure are revealed. The intuitive notion that supply/demand mismatch costs are driven down by decreasing uncertainty is now confirmed in this setting. In addition, one notices that at any given level of uncertainty expected demand is not artificially forced to be constant for comparison purposes and multiple possible distributions are now consistent with any specified level of uncertainty. Through both theoretical and intuitive insight, entropy as an uncertainty measure provides results that contrast with those of (Ridder et al. 1998, Lemma 3) where larger demand variability, under the assumption of constant mean demand, is sometimes associated with higher costs.

Appendix A: Mathematical Proofs

A.1. Proof of Lemma 1

1. $p_1 = p_2 = \ldots = p_N = 1/N$ certainly satisfies $-\sum_{i=1}^N p_i \log p_i = \log N$ and $\sum_{i=1}^N p_i = 1$ so $\mathbf{e}/N \in S(\log N)$. Now we show this is the only point in $S(\log N)$. Consider maximizing $-\sum_{i=1}^N p_i \log p_i$ subject to the constraint $\sum_{i=1}^N p_i = 1$. Let

$$\mathcal{L}(p_1, \ldots, p_N) = -\sum_{i=1}^N p_i \log p_i - \lambda \left(\sum_{i=1}^N p_i - 1 \right). \tag{17}$$

Taking $\frac{\partial \mathcal{L}}{\partial p_i} = 0$, we find that $p_i = 1/N$. Furthermore, $\frac{\partial^2 \mathcal{L}}{\partial {p_i}^2} = -1/p_i < 0$ so $\mathcal{L}$ is strictly concave for all p_i and $p_1 = p_2 = \ldots = p_N = 1/N$ is the *unique* global maximizer. Therefore $\mathbf{e}/N$ is the only point in $S(\log N)$.

2. For the unconstrained isoentropy surface $H(p_1, p_2, \ldots, p_{N-1}, p_N) = -\sum_{i=1}^N p_i \log p_i$ the normal ∇H is given by

$$(\nabla H)_j = -(\log p_j + 1), \quad j = 1, \ldots, N. \tag{18}$$

To find $\mathbf{r}$, ∇H must be projected onto the hyperplane $\mathbf{p} \cdot \mathbf{e} = 1$. Recall that projection of a vector $\mathbf{v}$ onto a plane with unit normal $\hat{\mathbf{m}}$ is given by $(I - \hat{\mathbf{m}}\hat{\mathbf{m}}^T)\mathbf{v}$ and $P \equiv (I - \hat{\mathbf{m}}\hat{\mathbf{m}}^T)$ is the projection matrix. For $\hat{\mathbf{m}} = \mathbf{e}/\sqrt{N}$, the projection matrix is given by

$$P_{ij} = \begin{cases} 1 - 1/N & \text{if } i = j, \\ -1/N & \text{if } i \neq j. \end{cases} \tag{19}$$

After some calculation, we find that

$$\mathbf{r} \equiv P \nabla H = \left\{ \frac{1}{N} \left(\sum_{i=1}^N \log p_i \right) \mathbf{e} - (\log p_1, \log p_2, \ldots, \log p_N)^T \right\}. \quad Q.E.D. \tag{20}$$

A.2. Proof of Lemma 2

To show that $[0,1]^N$ is partitioned by $D_1,\ldots,D_N$, we must show that every $\mathbf{p}\in[0,1]^N$ belongs to exactly one loss region. For any point $\mathbf{p}\in[0,1]^N$ in the belief space, the loss corresponding to an order quantity of k units is given by $\mathbf{a}_k^T\mathbf{p}$. In particular, there will always be an order quantity that gives the smallest loss, though not necessarily uniquely. Assume first that this order quantity is unique; let this quantity be i. Then we have

$$\mathbf{a}_i^T\mathbf{p} < \mathbf{a}_j^T\mathbf{p}, \qquad i\neq j. \tag{21}$$

If $i=1$, then $\mathbf{p}\in D_1$ by (7). If $i>1$, then clearly $\mathbf{p}\notin D_1,\ldots,D_{i-1}$. So by (8), $\mathbf{p}\in D_i$. Now assume that the optimal order quantity is not unique and that the $M>1$ order quantities $i_1<i_2<\ldots<i_M$ all yield the smallest loss. Then

$$\mathbf{a}_{i_1}^T\mathbf{p}=\mathbf{a}_{i_2}^T\mathbf{p}=\ldots=\mathbf{a}_{i_M}^T\mathbf{p},\ \text{and}\ \mathbf{a}_i^T\mathbf{p}<\mathbf{a}_j^T\mathbf{p}, \qquad i\in\{i_1,i_2,\ldots,i_M\},\quad j\notin\{i_1,i_2,\ldots,i_M\}. \tag{22}$$

From (7) and (8), loss regions with smaller indices are defined before regions with larger indices. So $\mathbf{p}\in D_{i_1}$ and $\mathbf{p}$ is uniquely assigned to the i_1^{th} loss region. Q.E.D.

A.3. Derivation of Eq. (15)

LEMMA 3. *For a differentiable function $f(\mathbf{p})$ and the contour $S(h)$ defined in (4), we have*

1.
$$\frac{d}{dh}\int_{S(h)} f(\mathbf{p})\,ds = \int_{S(h)} \beta\hat{\mathbf{n}}\cdot\nabla f\,ds, \tag{23}$$

where $\beta=-(\hat{\mathbf{n}}\cdot\mathbf{b})^{-1}>0$, $b_j=1+\log p_j$ and $\hat{\mathbf{n}}=\mathbf{r}/||\mathbf{r}||$ with $\mathbf{r}$ defined in (6).

2.
$$\frac{d}{dh}\left(\frac{\int_{S(h)} f(\mathbf{p})\,ds}{\int_{S(h)} ds}\right) = \frac{\int_{S(h)} \beta\hat{\mathbf{n}}\cdot\nabla f\,ds}{\int_{S(h)} ds}. \tag{24}$$

Proofs 1. Using the definition of the derivative, we have

$$\frac{\mathrm{d}}{\mathrm{d}h}\int_{S(h)} f(\mathbf{p})\mathrm{d}s = \lim_{\delta h\to 0}\frac{1}{\delta h}\left[\int_{S(h+\delta h)} f(\mathbf{p})\mathrm{d}s - \int_{S(h)} f(\mathbf{p})\mathrm{d}s\right], \tag{25}$$

$$= \lim_{\delta h\to 0}\frac{1}{\delta h}\int_{S(h)}[f(\mathbf{p}+\delta\mathbf{p})-f(\mathbf{p})]\mathrm{d}s. \tag{26}$$

Since the contour $S(h+\delta h)$ consists of points $\mathbf{p}\in S(h)$ advected along the normal direction of $S(h)$, $\mathbf{p}+\delta\mathbf{p}\in S(h+\delta h)$ for some $\delta\mathbf{p}\propto\hat{\mathbf{n}}$. We now calculate $\delta\mathbf{p}$ explicitly. Without loss of generality, we can take

$$\delta\mathbf{p}=\beta(\delta h)\times\hat{\mathbf{n}}, \tag{27}$$

for some scalar β. Since $\mathbf{p}+\delta\mathbf{p}\in S(h+\delta h)$,

$$-\sum_{i=1}^{N}(p_i+\delta p_i)\ln(p_i+\delta p_i) \;=\; h+\delta h, \tag{28}$$

$$\Rightarrow -\sum_{i=1}^{N}(\ln p_i+\delta p_i/p_i+\ldots)(p_i+\delta p_i) \;=\; h+\delta h. \tag{29}$$

But since $\mathbf{p}\in S(h)$, to leading order, (29) reduces to

$$-\sum_{i=1}^{N}\delta p_i[1+\ln p_i]=\delta h \Longrightarrow -\delta\mathbf{p}\cdot\mathbf{b}=\delta h, \tag{30}$$

where $b_j = 1 + \ln p_j$. Substituting (27) into (30), we deduce that

$$\beta = -\frac{1}{\hat{\mathbf{n}} \cdot \mathbf{b}}, \tag{31}$$

and (26) gives

$$\frac{\mathrm{d}}{\mathrm{d}h} \int_{S(h)} f(\mathbf{p}) \mathrm{d}s = \lim_{\delta h \to 0} \frac{1}{\delta h} \int_{S(h)} \delta \mathbf{p} \cdot \nabla f \mathrm{d}s = \int_{S(h)} \beta \hat{\mathbf{n}} \cdot \nabla f \mathrm{d}s. \tag{32}$$

The scalar $\beta > 0$ since

$$\beta = -(\hat{\mathbf{n}} \cdot \mathbf{b})^{-1} = -\frac{||\mathbf{r}||}{\mathbf{r} \cdot \mathbf{b}}, \tag{33}$$

and with $x_i \equiv \ln p_i$, one can easily show that

$$-\mathbf{r} \cdot \mathbf{b} = \sum_{i=1}^{N} \left[x_i - \left(\frac{\sum_{j=1}^{N} x_j}{N} \right) \right]^2 > 0, \tag{34}$$

using the definition of $\mathbf{r}$ in (6).

2. Using the result from part 1. of this lemma,

$$\frac{\mathrm{d}}{\mathrm{d}h} \left(\frac{\int_{S(h)} f(\mathbf{p}) \mathrm{d}s}{\int_{S(h)} \mathrm{d}s} \right) = \frac{\int_{S(h)} \beta \hat{\mathbf{n}} \cdot \nabla f \mathrm{d}s \int_{S(h)} \mathrm{d}s - \int_{S(h)} f(\mathbf{p}) \mathrm{d}s \int_{S(h)} \beta \hat{\mathbf{n}} \cdot \nabla (1) \mathrm{d}s}{\left(\int_{S(h)} \mathrm{d}s \right)^2}, \tag{35}$$

$$= \frac{\int_{S(h)} \beta \hat{\mathbf{n}} \cdot \nabla f \mathrm{d}s}{\int_{S(h)} \mathrm{d}s}. \tag{36}$$

Q.E.D.

References

Abbas, A.E. 2006. Entropy. *John Wiley Encyclopedia of Biomedical Engineering*. John Wiley and Sons, Inc.

Andersson, Jonas, Kurt Jörnsten, Sigrid Lise Nonås, Leif K. Sandal, Jan Ubøe. 2011. A maximum entropy approach to the newsvendor problem with partial information. *Forthcoming in the European Journal of Operational Research* .

Choi, Sungyong, Andrzej Ruszczynski, Yao Zhao. 2011. A multiproduct risk-averse newsvendor with law-invariant coherent measures of risk. *Operations Research* **59**(2) 346–364.

Ebrahimi, Nader, Esfandiar Maasoumi, Ehsan S Soofi. 1999. Ordering univariate distributions by entropy and variance. *Journal of Econometrics* **90**(2) 317–336.

Eren, Serkan, Costis Maglaras. 2009. A Maximum Entropy Joint Demand Estimation and Capacity Control Policy. *SSRN eLibrary* .

Gerchak, Yigal, David Mossman. 1992. On the effect of demand randomness on inventories and costs. *Operations Research* **40**(4) pp. 804–807.

Glasserman, Paul, Xingbo Xu. 2013. Robust portfolio control with stochastic factor dynamics. *Operations Research* **61**(4) 874–893.

Jaynes, E.T. 1957. Information theory and statistical mechanics. *Phys. Rev.* **106**.

Jaynes, E.T. 2003. *Probability Theory: The Logic of Science*. Cambridge University Press.

Jemaï, Zied, Fikri Karaesmen. 2005. The influence of demand variability on the performance of a make-to-stock queue. *European Journal of Operational Research* **164**(1) 195–205.

Johnson, R., J. Shore. 1983. Comments on and correction to "axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy" (jan 80 26-37) [corresp.]. *Information Theory, IEEE Transactions on* **29**(6) 942 –943.

Kwak, Jin Kyung, Srinagesh Gavirneni. 2011. Retailer policy, uncertainty reduction, and supply chain performance. *International Journal of Production Economics* **132**(2) 271 – 278.

Lee, Hau L, V Padmanabhan, Seungjin Whang. 1997. The bullwhip effect in supply chains1. *Sloan management review* **38**(3) 93–102.

Lim, Andrew EB, J George Shanthikumar. 2007. Relative entropy, exponential utility, and robust dynamic pricing. *Operations Research* **55**(2) 198–214.

McNeil, Alexander J, Rüdiger Frey, Paul Embrechts. 2010. *Quantitative risk management: concepts, techniques, and tools*. Princeton university press.

Nakagawa, Yuji, R.J.W. James, César Rego, Chanaka Edirisinghe. 2013. Entropy-based optimization of nonlinear separable discrete decision models. *Management Science* .

Niven, Robert K. 2007. Origins of the combinatorial basis of entropy. *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, vol. 954. 133–142.

Perakis, Georgia, Guillaume Roels. 2008. Regret in the newsvendor model with partial information. *Oper. Res.* **56** 188–203.

Porteus, E.L. 2002. *Foundations of stochastic inventory theory*. Stanford Business Books.

Ridder, Ad, Erwin van der Laan, Marc Salomon. 1998. How larger demand variability may lead to lower costs in the newsvendor problem. *Operations Research* **46**(6) pp. 934–936.

Shannon, C. E. 1948. A mathematical theory of communication. *Bell Sys. Tech. J.* **27** 379–423,623–656.

Shore, J., R. Johnson. 1980. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *Information Theory, IEEE Transactions on* **26**(1) 26 – 37.

Shuiabi, Eyas, Vince Thomson, Nadia Bhuiyan. 2005. Entropy as a measure of operational flexibility. *European Journal of Operational Research* **165**(3) 696–707.

Song, Jing-Sheng. 1994. The effect of leadtime uncertainty in a simple stochastic inventory model. *Management Science* **40**(5) pp. 603–613.

Song, Jing-Sheng, Hanqin Zhang, Yumei Hou, Mingzheng Wang. 2010. The effect of lead time and demand uncertainties in (r, q) inventory systems. *Operations Research* **58**(1) 68–80.

Szegö, Giorgio. 2005. Measures of risk. *European Journal of Operational Research* **163**(1) 5–19.

Taylor, Terry A., Wenqiang Xiao. 2010. Does a manufacturer benefit from selling to a better-forecasting retailer? *Management Science* **56**(9) pp. 1584–1598.

Xu, Minghui, Youhua Frank Chen, Xiaolin Xu. 2010. The effect of demand uncertainty in a price-setting newsvendor model. *European Journal of Operational Research* **207**(2) 946–957.