An Entropy Based Methodology for Valuation of Demand Uncertainty Reduction

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

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

We propose a distribution-free entropy-based methodology to calculate the expected value of an uncertainty reduction effort and present our results within the context of reducing demand uncertainty. In contrast to existing techniques, the methodology does not require a priori assumptions regarding the underlying demand distribution, does not require sampled observations to be the mechanism by which uncertainty is reduced, and provides an expectation of information value as opposed to an upper bound. In our methodology, a decision maker uses his existing knowledge combined with the maximum entropy principle to model both his present and potential future states of uncertainty as probability densities over all possible demand distributions. Modeling uncertainty in this way provides for a theoretically justified and intuitively satisfying method of valuing an uncertainty reduction effort without knowing the information to be revealed. We demonstrate the methodology's use in three different settings: 1) a newsvendor valuing knowledge of expected demand, 2) a short-lifecycle product supply manager considering the adoption of a quick response strategy, and 3) a revenue manager making a pricing decision with limited knowledge of the market potential for his product.

Key words: Maximum Entropy Principle, Expected Value of Information, Distribution Free Models, Demand and Inventory Management

---

1. INTRODUCTION

For decision makers facing uncertainty, a natural response is to collect more information. "Better information, better decisions" is an often heard adage. At the same time, many managers lament the loss of time spent in meetings upon meetings discussing information gathering efforts that seem to have minimal impact on changing the decision at hand. This conflict between time and information forces managers to

> ...navigate between two deadly extremes: on the one hand, ill-conceived and arbitrary decisions made without systematic study and reflection ("extinction by instinct") and on the other, a retreat into abstraction and conservatism that relies obsessively on numbers, analyses, and reports ("paralysis by analysis") (**?**).

When a decision ultimately gets made, the implied belief is that expected costs of further uncertainty reduction exceed the expected benefit, but justified and rigorous methods of calculating that benefit have proved elusive. To remedy this, we present a methodology for valuing the pursuit of information and do so within the context of *demand uncertainty* reduction. Increased demand volatility and increased global sourcing (i.e. long leadtimes) make information valuation particularly relevant in this context. For example, to forecast demand for a new fashion apparel item, limited demand information is available and a manager is naturally uncomfortable making a stocking decision for the entire selling season. In response, management might seek input from a consumer focus group to reduce pre-season uncertainty or try to postpone production decisions to a time where uncertainty is reduced. Valuing the associated improvement in expected outcome is required, but how to value that effort remains an under-explored question in the literature.

Our valuation approach is to consider and compare the expected outcomes of two decision makers; the *ignorant manager* who has not pursued uncertainty reduction and the *informed manager* who has. As opposed to using a single demand distribution, akin to work by **???**, we assign a probability distribution over all possible demand distributions. The principle of maximum entropy (**?**) is adopted for this probability assignment purpose. We refer to any specific demand distribution as a *belief* and refer to a probability distribution over all possible beliefs as a *belief distribution*. The constructed belief distribution is deemed most consistent with currently available information and provides the foundation for creating probability distributions for potential future information. Given a distribution over all potential future information, the informational advantage of the informed manager over the ignorant manager can be calculated. To our knowledge, this is the first paper to demonstrate application of the maximum entropy principle for valuation of uncertainty reduction.

The largest advantage of this approach is that it enables numeric calculation of the *expected value* of information without knowing the information that will be revealed. Existing valuation techniques only provide bounds on the value of information and value-estimates driven by upper bounds tend to overstate the information's impact. Other numeric information valuation techniques in the literature rely on a priori knowledge of the information that will be revealed; through comparison of decisions made with and without the known information, information value can be investigated. While this comparison provides insight, it lacks prescriptive guidance for a manager. If the manager knew the information to be revealed, then there would be no need for computing its value in the first place.

2. LITERATURE REVIEW

Three key aspects of the existing literature serve as the foundation of our analysis. First, we examine distribution free approaches to modeling demand uncertainty and uncertainty reduction. These approaches facilitate robust and tractable decision making and share our perspective that distributional assumptions should not be restricted to specific distribution parameters and families. Second, our work's objective is to value information and we examine existing techniques used for this purpose. Lastly, we review entropy's role as an uncertainty measure and motivate its usefulness in assigning a probability density over all possible demand distributions.

2.1. Distribution Free Approaches

When the assumption of a specific model or model family may be incorrect, the assumed model is often outperformed by other techniques that make less restrictive assumptions about demand (see for example **????**). One active research area that facilitates these less restrictive assumptions is robust optimization. The robust optimization framework finds "worst-case" demand distributions subject to constraints imposed by any existing information (see pioneering work by **??**). Successful works leveraging the robust framework for information valuation are plentiful. Both **?** and **?** show how to calculate the maximum expected value of distribution information given a specified base-state of knowledge (such as a given mean and standard deviation). **?** show the tractability of optimal inventory decisions which minimize maximum regret (as opposed to maximizing minimum profit as in **?**). The authors study information valuation in this context, but the value of information is accomplished via comparing the profit of an uninformed decision maker to that of an oracle.

Recent robust optimization techniques (for example **????**) continue to facilitate decision making where distributional uncertainty exists. Of particular note, **?** explores the value of full stochastic modeling and asks a related question to our own, "How can we find out if we would achieve more with a stochastic model without developing the stochastic model?" In contrast to all of these "robust" approaches, our modeling of demand distribution uncertainty seeks an expectation of information value whereas the robust alternatives are restricted to providing an upper bound or a maximum expected value of information. This is a direct consequence of robust approaches relying on a "very unreasonable type of [demand] distribution" (**?**). Mathematically, a two-point demand distribution is often assumed.

Data-driven approaches are another form of distribution free response to managing demand uncertainty. Data driven approaches include operational statistics (**?**), sample average approximation (**??**), Kaplan-Meier estimators (**?**), and other non-parametric algortihms (**??**). For our intent, data is not assumed.

2.2. Valuing Uncertainty Reduction

The inventory control literature often uses variance (or equivalently, standard deviation) to quantify uncertainty surrounding possible values of demand (see for example **???**). Hence, the literature often adopts using variance reduction as a proxy for modeling uncertainty reduction. The seminal work by **?** models uncertainty in a particular item's forecast using an estimate of the standard deviation in forecast error. Assuming a specific distribution, the bivariate normal, differentiates their work from ours which uses a distribution-free approach (see related discussion in **?**, for comparison of entropy and variance).

Academic authors have successfully used other uncertainty-reduction valuation methods besides variance reduction. **?** study the effects of leadtime uncertainty on long-run average inventory costs using stochastic ordering criteria. While stochastic ordering defines conditions for which one random variable is more variable than another, generated insights tend to be more qualitative than quantitative. **?** enable more numerically driven studies of uncertainty's effects by introducing a quantitative methodology borrowed from the micro-economics literature, namely mean preserving transformation. Unfortunately, there is a downside of using this transformation. All potential levels of uncertainty, from high to low, have the same expectation of demand. As such, the technique fails to accurately model potential future states of uncertainty that may exist after one pursues uncertainty reduction; the probability of a change in ordering decision due to full or partial uncertainty cannot be modeled properly.

Partial uncertainty reduction is often modeled through a Bayesian approach using observations of demand to reduce uncertainty. The seminal work of **?** is often thought of as pioneering the Bayesian updating approach and has proven successful in the literature over the last half-century. **?** use a Bayesian updating procedure to classify future demand as being consistent with a "hot" selling product or a "cold" product. Other papers applying this Bayesian approach include **?**, **?**, and **?**. Extensions to accommodate unobservable lost sales are noteworthy and include **?**, **?**, **?**, **?**, and **?**. In contrast to the aforementioned work, we can accommodate, but do not require distributional assumptions or demand observations to model uncertainty reduction valuation.

2.3. Using the Principle of Maximum Entropy for Assigning Probabilities to Outcomes

The principle of maximum entropy, as originally developed in **?** as an extension of **?**, has seen tremendous success in assigning probability distributions to under-specified problems where multiple distributions are consistent with the provided information. Maximizing entropy, subject to constraints based on existing knowledge (e.g. the mean, support, moments, etc.), is the "means of determining quantitatively the full extent of uncertainty already present" (**?**). Axiomatic derivation of the principle of maximum entropy shows that under certain conditions it is a uniquely correct method for inductive inference (**????**). Recent successes from wildlife research (**?**), linguistics (**?**), biology (**?**), software engineering (**?**), and notably operations management (**???**) are examples of the continuing success of the maximum entropy principle in applied settings. Despite an abundance of successes, debate surrounding the maximum entropy principle's rationale does exist (see references within introduction of **?**). We do not seek to resolve this debate, but rather demonstrate the promise of using entropy and the maximum entropy principle in our context.

For probability assignment, the maximum entropy principle for the differential (or continuous) form of entropy is used in this paper (see **?**, for introductory material). Over the domain, $[1, N]$ differential entropy is defined as: $H(x) = -\int_1^N f(x)\log(f(x))\,dx$ where $f(\cdot)$ is a probability density function. By finding the density $f(\cdot)$, maximizing $H(x)$ subject to any constraints, one is able to pick a distribution that is considered most consistent with the constraints of the problem. In the univariate case, many well-known distributions are maximum entropy distributions when particular constraints are imposed. For example, when only the support of the distribution is known, the uniform distribution is entropy maximizing; if only the mean and variance are prescribed, then the normal distribution is entropy maximizing; given just a mean and support of $[0, \infty]$, then the exponential distribution is entropy maximizing. For more examples, readers are referred to the list of distributions in **?** and the more in depth derivations of **?**.

Our use of differential entropy implies a definition of ignorance that should be stated explicitly. Namely, we assume ignorance represents an inability to prefer any one demand distribution to any other. Without information, all possible demand distributions are considered equally likely. Other assumptions of ignorance (see **?**) can be accommodated through the use of relative (or cross-) entropy instead of the differential entropy used in this paper. **?** and **?** discuss the required mathematical machinery for these extensions and our method is equivalent to the use of relative entropy when the objective is to minimize disparity between the desired density and the uniform density (**?**). Our use of differential entropy is similar to the techniques of **?** with Crook's metaprobability being analogous to a belief distribution.

3. THEORETICAL PRELIMINARIES

In this section, we cover three critical elements enabling the valuation of uncertainty reduction. The first element of our methodology is to leverage the maximum entropy principle to form *belief distributions*. These distributions represent a decision maker's state of uncertainty. The second element is to create a distribution for potential future information in a way that is consistent with a given belief distribution. We call this an *information distribution* and it enables uncertainty reduction modeling without a priori knowledge of the information to be received. The third critical element is the *expected regret function* which computes the expected value of an uncertainty reduction effort. Combining all of these elements, one can value an uncertainty reduction effort without making any assumptions, distributional or otherwise, beyond the decision maker's current knowledge. For expositional ease, we summarize the important notation to be introduced in this section:

Sets, Random Variables, and Realizations

d_i: Possible demand realizations indexed by $i \in \{1, 2, \ldots, N\}$.
D: A random variable representing demand with realizations d_i where $i \in \{1, 2, \ldots, N\}$.
p_i: Probability that demand, D, is equal to d_i.
u_j: Information in the form of a statistic(s), indexed $j \in \{1, 2, \ldots, m\}$, about the unknown demand distribution.
$\mathcal{P}$: Set of all possible demand distributions.
$\mathcal{Q}$: Set of all possible demand distributions after information is realized.
$p_1, \ldots, p_N$: A **belief** - single (generic) instance of a demand distribution where $p_1, \ldots, p_N \in \mathcal{P}$.
$\mathbf{p} = p_1, \ldots, p_N$: A random variable, in the Bayesian sense, representing the true demand distribution.
$\mathbf{x} = x_1, \ldots, x_N$: A demand distribution realization of $\mathbf{p} = p_1, \ldots, p_N$.
$\mathbf{q} = q_1, \ldots, q_N$: A random variable, in the Bayesian sense, representing the true demand distribution after information is realized.
$\mathbf{u} = u_1, \ldots, u_m$: A random variable, in the Bayesian sense, representing a vector of information.
$\mathbf{y} = y_1, \ldots, y_m$: An information vector realization of $\mathbf{u} = u_1, \ldots, u_m$.

Important Functions

$\hat{\mathbf{p}} = E[\mathbf{p}]$: An **operational belief** representing the demand distribution (belief) used for decision making.
$\hat{\mathbf{q}} = E[\mathbf{q}]$: An **operational belief** representing the demand distribution (belief) used for decision making after information is realized.
$f(\mathbf{x})$: A **belief distribution** - a probability distribution over all possible beliefs $p_1, \ldots, p_N \in \mathcal{P}$ occasionally subscripted, $f_{\mathbf{p}}(\cdot)$ and $f_{\mathbf{q}}(\cdot)$, referring to the ignorant and informed managers' belief distributions, respectively.
$\mathcal{M}(\mathbf{x})$: A mapping function whose input is a belief (i.e. demand distribution) and whose output is a statistic (e.g. mean, median, mode). Multiple maps, $\mathcal{M}_j(\cdot)$, are indexed by $j \in \{1, 2, \ldots, m\}$ where m separate statistics are calculated for comparison against information u_j.

3.1. Belief Distributions

For this paper, we impose two requirements on a decision maker's knowledge of a demand distribution. First, demand is discrete and second, the support of demand is known. With just these two constraints, a decision maker faces ambiguity over which of the infinitely-many feasible demand distributions, denoted by $\mathcal{P}$, to employ for decision making. In contrast to existing techniques, we do not seek a single probability distribution for demand $p_1, \ldots, p_N$; rather our fundamental quantity of interest is a *belief distribution* $f(x_1, \ldots, x_N)$: a probability distribution, over all possible beliefs $(p_1, \ldots, p_N) \in \mathcal{P}$, that is consistent with our knowledge. To form a belief distribution, one assigns probability to each n-tuple as if the n-tuples themselves are drawn from some distribution (analogous to modeling metaprobability as described in **?**):

$$\mathbf{p} \equiv (p_1, p_2, \ldots, p_N) \sim f(x_1, \ldots, x_N).$$

where larger values of $f(\cdot)$ correspond to more plausible beliefs.

We define two types of belief distributions: 1) the ignorant manager's belief distribution, $f_{\mathbf{p}}(\cdot)$, and 2) the informed manager's belief distribution, $f_{\mathbf{q}}(\cdot)$. The ignorant manager's belief distribution represents the current state of uncertainty while an informed belief distribution represents

	Information u
Mode of Demand	$d_i^* : p_{i^*} = \max(\{p_1, p_2, \ldots, p_N\})$
Median of Demand	$d_i^* : \sum_{i=1}^{i^*-1} p_i < 1/2, \quad \sum_{i=1}^{i^*} p_i > 1/2$
Mean Demand	$\sum_{i=1}^{N} d_i p_i$

Table 1 Possible information that can be priced in our methodology. The demand distribution takes the form $P(\mathbf{Demand} = d_i) = p_i$, $i = 1, \ldots, N$. In this paper, our focus is on valuing the mean: $u = \sum_{i=1}^{N} d_i p_i$.

uncertainty given additional information. In this paper, we assume that the additional information takes the form of constraints on the admissible beliefs in $\mathcal{P}$ and our goal is to value this information. Generally, we can assume that the m pieces of information take the form of mappings $\mathcal{M}_j(p_1, \ldots, p_N) = u_j$, $1 \le j \le m$. When $m = 1$, we simply write $\mathcal{M}(p_1, \ldots, p_N) = u$. Then in light of the constraints, the set of all demand distributions is restricted to a subset $\mathcal{Q}$ with elements $\mathbf{q}$ such that

$$\mathcal{Q} = \mathcal{P} \mid \mathcal{M}(\mathbf{p}) = \mathbf{u} \quad \Longleftrightarrow \quad \mathbf{q} = \mathbf{p} \mid \mathcal{M}(\mathbf{p}) = \mathbf{u}$$

and the corresponding conditional density is $f_{\mathbf{q}} = f_{\mathbf{p}|\mathcal{M}=\mathbf{u}}$. An obvious corollary is $\mathcal{M}(\mathbf{q}) = \mathbf{u}$.

Some examples of $\mathcal{M}$ are shown in Table 1. The examples shown are statistics, but our approach can also be used to value information that does not relate to "well-known" statistics. For example, when $N = 3$, for some $\delta \ll 1$ we could define

$$u = \begin{cases} 1, \; |p_1 - p_3| < \delta, \\ 0, \; \text{otherwise} \end{cases} \tag{1}$$

as measuring how symmetric the demand distribution is with respect to d_1 and d_3. Knowledge of u as defined in (1) can also be valued within our theoretical framework.

If u is known, the constraint restricts the set of possible beliefs. For example, when $d_i = i$ and $N = 3$, if the mean demand is known to be 2.5, then only beliefs (p_1, p_2, p_3) that satisfy $p_1 + 2p_2 + 3p_3 = 2.5$ are admissible. "Milder" constraints can also be imposed on belief distributions that do not restrict the support of the belief distribution. For example, moments of f could be specified. These *shaping* constraints do not eliminate potential beliefs; they simply elevate certain beliefs to be more plausible than others. In §5.2, we present a problem where more plausible beliefs have expected variance near a certain value, but all beliefs remain feasible representations of demand. In this paper, we assume that the information represented by the shaping constraints is available to both ignorant and informed managers. The valuation of such information is not treated in this paper, but is the subject of future work.

What is the form of f? Naturally, since it is a joint continuous probability distribution, its integral must be 1. There may also be other constraints on f that reflect the ignorant manager's state of knowledge, for example certain moments of f may be known. This knowledge on f can be incorporated into a *differential entropy* functional S that incorporates m constraints:

$$S[f(x_1,\ldots,x_N)] = -\int_{V_{N-1}} f\log f \mathrm{d}V + \sum_{j=1}^{m}\lambda_j\left[\int_{V_{N-1}} B_j(x_1,\ldots,x_N)f(x_1,\ldots,x_N)\mathrm{d}V - C_j\right], \quad (2)$$

where

$$V_{N-1} = \left\{(x_1,\ldots,x_N) \mid 0\le x_j\le 1, 1\le j\le N, \text{ and } \sum_{j=1}^{N} x_j = 1\right\},$$

the B_j, $1\le j\le N$ are known functions, C_j, $1\le j\le N$ are given constants and λ_j, $1\le j\le m$ are Lagrange multipliers. Equivalently, the constraint $\sum_{j=1}^{N} x_j = 1$ can be used directly to reduce the dimension of the domain of integration leading to the functional

$$\begin{aligned} S^*[\phi(x_2,\ldots,x_N)] = &-\int_{V^*_{N-1}} \phi\log\phi \mathrm{d}x_2\ldots\mathrm{d}x_N + \quad (3)\\ &\sum_{j=1}^{m}\lambda_j\left[\int_{V^*_{N-1}} B_j\left(1-x_2-\ldots-x_N, x_2,\ldots,x_N\right)\phi(x_2,\ldots,x_N)\mathrm{d}x_2\ldots\mathrm{d}x_N - C_j\right], \end{aligned}$$

where

$$V^*_{N-1} = \left\{(x_2,\ldots,x_N) \mid 0\le x_j\le 1, 2\le j\le N, \text{ and } \sum_{j=2}^{N} x_j \le 1\right\}.$$

Maximization of the functionals (2) and (3) yield the maximum entropy densities $f(x_1,x_2,\ldots,x_N)$ and $\phi(x_2,\ldots,x_N)$. The two approaches are equivalent since f and ϕ are simply related through a multiplicative Jacobian of transformation. The differential entropy method is common in the statistical mechanics literature where its maximization is used to find canonical equilibrium distributions of many-particle systems, see **?** for example. **?** gives many examples of maximum differential entropy distributions under different constraints. In the simplest case, $m=1$, $B_1=C_1=1$, yielding a uniform Dirichlet distribution.

For notational purposes, we derive the informed belief distribution assuming a single piece of information so that $\mathbf{u}=u$. The generalization to multiple pieces of information is straightforward. The manager adopts an informed belief distribution $f_{\mathbf{q}}$ defined as the probability density over $\mathcal{P}$ conditional on constraint $\mathcal{M}(p_1,\ldots,p_N)=u$. Let $\Omega(u)=\{\mathbf{p}\in V_{N-1} \mid \mathcal{M}(p_1,\ldots,p_N)=u\}$ denote the restricted set of beliefs given u and let $I_\Omega(\mathbf{x})$ denote the usual indicator function: $I_\Omega(\mathbf{x})=1$ if $\mathbf{x}\in\Omega$ and $I_\Omega(\mathbf{x})=0$ if $\mathbf{x}\notin\Omega$. Suppose the random variable u is discrete so that $u\in\{y_1,y_2,\ldots,y_L\}$ with corresponding probabilities $\mathrm{Prob}(u=y_i)=g_{y_i}$. Then

$$\begin{aligned} P(\mathbf{x}\le\mathbf{p}\le\mathbf{x}+\mathrm{d}\mathbf{x} \mid u=y_i) &= \frac{P(\mathbf{x}\le\mathbf{p}\le\mathbf{x}+\mathrm{d}\mathbf{x} \,\cap\, u=y_i)}{P(u=y_i)},\\ &= \frac{P(\mathbf{x}\le\mathbf{p}\le\mathbf{x}+\mathrm{d}\mathbf{x} \,\cap\, M(\mathbf{q})=y_i)}{P(u=y_i)}. \end{aligned}$$

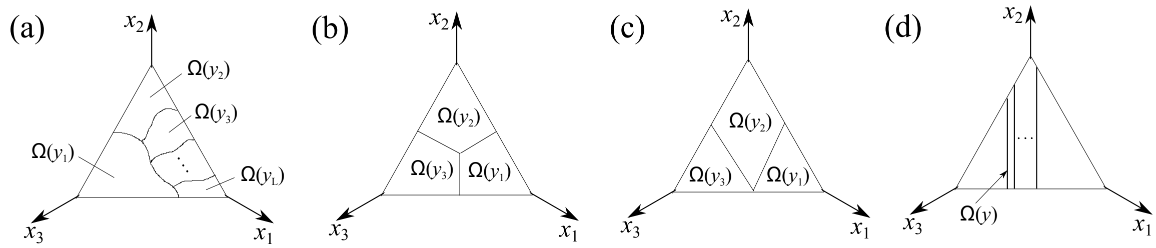

Figure 1 **(a) The constraint $\mathcal{M}(p_1,...,p_N) = u$ partitions V into L subsets (L can be infinite). Specific cases where u is the mode, median and mean are shown in (b), (c) and (d).**

The numerator is equal to zero if $\mathbf{x} \notin \Omega(y_i)$ and is just $P(\mathbf{x} \le \mathbf{p} \le \mathbf{x} + \mathrm{d}\mathbf{x})$ if $\mathbf{x} \in \Omega(y_i)$. Therefore

$$f_{\mathbf{q}}(\mathbf{x}; y_i) = \frac{f_{\mathbf{p}}(\mathbf{x}) I_{\Omega(y_i)}(\mathbf{x})}{g_{y_i}}. \tag{4}$$

The constraints $\mathcal{M}(p_1,\dots,p_N) = y_i,\ i = 1,\dots,L$ partition the simplex V_{N-1} into L subsets $\Omega(y_1),\dots,\Omega(y_L)$ and $f_q(\mathbf{x}; y_i) \propto f_p(\mathbf{x})$ on $\Omega(y_i)$; see Fig. 1(a). Specific cases where u is the mode and median are shown in (b) and (c). Now suppose that u is a continuous random variable so that $\mathrm{Prob}(y \le u \le y + \mathrm{d}y) = g_u(y)\mathrm{d}y$. Then the continuous generalization of (4) is

$$f_{\mathbf{q}}(\mathbf{x}; y) = \frac{f_{\mathbf{p}}(\mathbf{x}) I_{\Omega(y)}(\mathbf{x})}{g_u(y)}. \tag{5}$$

The constraints $\mathcal{M}(p_1,\dots,p_N) = y,\ y_{\min} \le y \le y_{\max}$ partition the simplex V_{N-1} into an infinite number of subsets, each indexed by y and $f_{\mathbf{q}}(\mathbf{x}; y) \propto f_{\mathbf{p}}(\mathbf{x})$ on $\Omega(y)$. As an example, when u is the mean, the appropriate partitions are shown in Fig. 1(d).

3.2. Information Distributions

The ignorant belief distribution $f_{\mathbf{p}}$ assigns probability to all possible beliefs of demand. The constraint

$$\mathcal{M}(p_1,\dots,p_N) = \mathbf{u}, \tag{6}$$

can be viewed in two ways. If $\mathbf{u}$ is known, it can be interpreted as a constraint on the possible beliefs that the ignorant manager can take. On the other hand, if $\mathbf{u}$ is unknown and $\mathbf{p} \sim f_p$, the distribution of $\mathbf{u}$ can be computed. We call this distribution the *information distribution*, $g_{\mathbf{u}}$. We have already seen the role of the information distribution for computing the informed belief distribution in eqs. (4) and (5).

Mathematically, $g_{\mathbf{u}}$ is always well-defined so long as we can associate every belief $p_1,\dots,p_N$ with a single finite value of $\mathbf{u}$ – mathematically, this means that $\mathcal{M}(p_1,\dots,p_N)$ is single-valued and defined for every $\mathbf{p} \in V_{N-1}$. The information distribution can always be found numerically if one can sample from $f_{\mathbf{p}}$. In some special cases, it may even be possible to find an analytic form for $g_{\mathbf{u}}$. In §3.2.1 we will explore such a case.

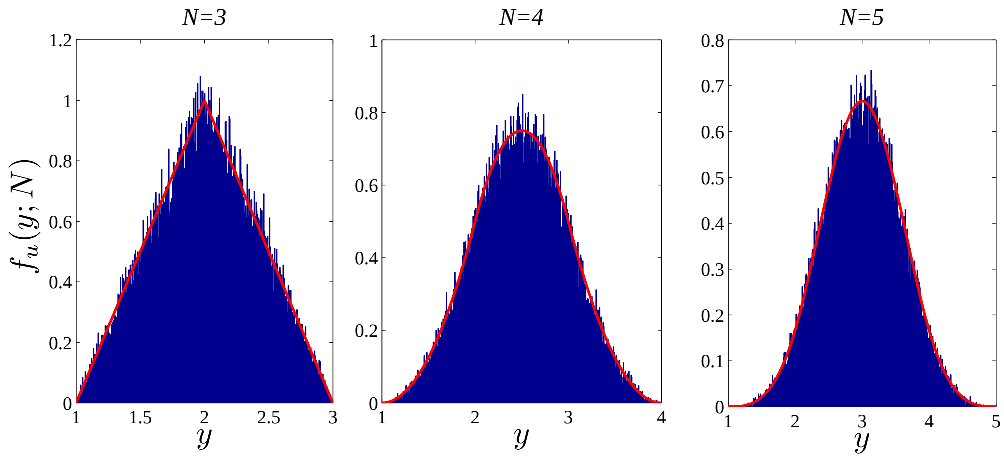

Figure 2 **Monte Carlo simulation results for the distribution of $\sum_{j=1}^{N} jp_j$ where p_j are uniformly sampled from the simplex V_{N-1}. Red curve shows analytic solutions using Lemma 1 for the cases $N=3,4,5$.**

3.2.1. Probability Distributions for Future Information When Only The Support of Demand is Known When the ignorant manager only knows the support of demand, the maximum (differential) entropy principle prescribes that all N-tuples $(p_1, p_2, \ldots, p_N)$ are considered equally likely (**?**, see pp. 123-127). Equivalently, the potential demand distribution (N-tuple) that accurately describes demand is uniformly drawn from the simplex V_{N-1}. The p_i are treated as random variables and hence, any function of $\mathbf{p}$ is also a random variable. In this section, we leverage the work of **?** to derive closed-form expressions for the distribution of the mean demand.

Suppose that the information gained from the uncertainty reduction effort is the mean u. Analytic forms for the distribution of $u \equiv p_1 + 2p_2 + \ldots + Np_N$ can be found through Lemma 1.

LEMMA 1. *Let $(p_1, p_2, \ldots, p_N)$ be a vector random variable, uniformly distributed over the simplex defined by $\sum_{i=1}^{N} p_i = 1$, $p_i \geq 0$. The probability distribution function for $u \equiv p_1 + 2p_2 + \ldots + Np_N$ is*

$$g_u(y;N) = \frac{(y-1)^{N-2}}{(N-2)!} - (N-1) \sum_{k=1}^{\lceil y \rceil -2} \frac{(y-1-k)^{N-2}}{k \prod_{i \neq N-k}(N-k-i)} \tag{7}$$

for $1 \leq y \leq N$ and $\lceil y \rceil$ is the smallest integer greater than or equal to y. The product in (7) is taken over all integers i from 1 to $N-1$ except for $N-k$.

Proof of Lemma 1 All proofs are provided in the Appendix.

For the special cases $N = 3, 4, 5$, Fig. 2 shows the analytic form of $g_u(y;N)$ and associated confirmation of the form through Monte Carlo simulations.

REMARK 1. The densities in (7) have an interesting property when N is large. Specifically, numerical investigations suggest the scaling

$$g_u(y;N) \sim \frac{1}{\sqrt{N}} \times H\left(\frac{y-(N+1)/2}{\sqrt{N}}\right), \qquad N \to \infty. \tag{8}$$

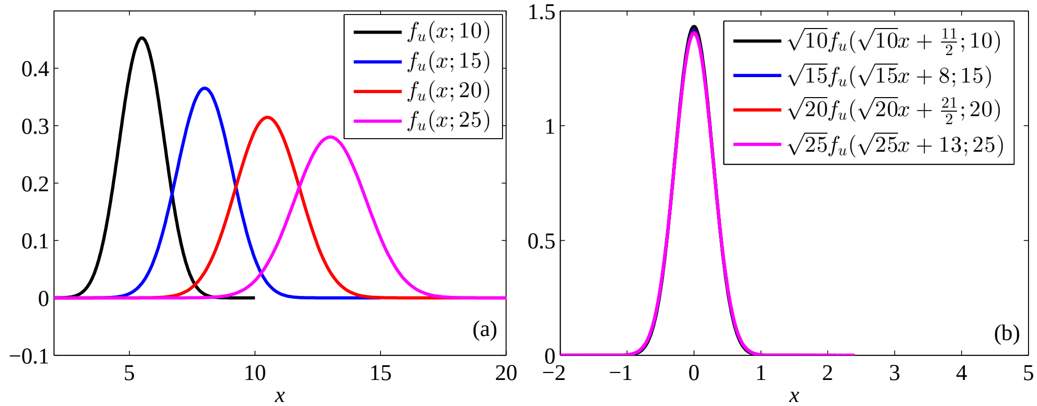

Figure 3 **(a) Probability density for the mean $f_u(x;N)$ as given by eq. (7) when $N=10,15,20,25$. (b) A rescaling of the x and y axes collapses all the densities onto approximately the same curve $H(\cdot)$ (see eq. (8)) providing N is sufficiently large.**

for some function $H(y)$. This function is found empirically in Fig. 3(b).

In other words, when N is large, $g_u(x;N)$ can be obtained from a single function $H(\cdot)$ which is approximated in Fig. 3(b). We use this property of f_u in the revenue management problem of §5.3.

3.3. Operational Beliefs and The Notion Of Regret

Assume loss matrix element L_{ki} represents the loss of making decision k when outcome d_i is realized. A classical example for this decision setting is that of a newsvendor who faces a stochastic demand and orders a certain number of newspapers (**?**). Thus, the expected loss of decision k, where we know only that $P(D=d_i)=p_i$, can be given by looking at the kth row of loss vector $\mathcal{L}$ such that:

$$\mathcal{L}_k = \sum_{i=1}^{N} L_{ki} \times P(D=d_i) = \mathbf{e}_k^T \mathbf{L}\mathbf{p}, \tag{9}$$

where $\mathbf{e}_k$ is the kth unit column vector (zeros everywhere except for a 1 in the kth row), $\mathbf{L}$ is the loss matrix with entries L_{ki}, and $\mathbf{p}$ is the true demand distribution.

Since $\mathbf{p}\sim f_{\mathbf{p}}$ is a stochastic quantity, the loss vector $\mathcal{L}$ is also stochastic. For decision making purposes, a manager with belief distribution $f_{\mathbf{p}}$ calculates the *expected* value of loss vector $\mathcal{L}$: $E[\mathcal{L}]=E[\mathbf{L}\mathbf{p}]=\mathbf{L}E[\mathbf{p}]$. Therefore, he chooses a single N-tuple, $\hat{\mathbf{p}}=(\hat{p}_1,\hat{p}_2,\ldots,\hat{p}_N)$, which we call an *operational belief*, such that each $\hat{p}_i$ represents the marginal probability of demand d_i:

$$\hat{\mathbf{p}} = E[\mathbf{p}]. \tag{10}$$

Notationally, we use $\hat{\mathbf{p}}$ to represent either a generic operational belief or the ignorant manager's operational belief, depending on context. For the informed manager, that is to say the perspective of a manager with information $\mathbf{u}$, eqns. (4) - (5) are used to generate the informed manager's belief distribution $\mathbf{q}(\mathbf{u}) \sim f_q$. The associated operational belief is

$$\hat{\mathbf{q}}(\mathbf{u}) = E[\mathbf{q}(\mathbf{u})]. \tag{11}$$

It is important to note that each operational belief, both ignorant and informed, is adopted strictly for decision making purposes; it is used only as input into loss function (9) and not as a statement about which demand distribution is in some sense more plausible. The plausibility is already captured in a belief distribution. When analytic forms of the operational beliefs are not available, the high dimensional integrals implied by $E[\cdot]$ in (10) and (11) can be numerically approximated using the Metropolis-Hastings algorithm (**?**).

The value of information is measured using the notion of regret. In this paper, regret measures the reduction in expected loss (e.g. expected supply/demand mismatch costs) that could be obtained when making a decision from a more informed perspective as opposed to sticking with the decision made from the more ignorant perspective. Assuming that managers act rationally in order to minimize loss, the regret of the ignorant manager who fails to use or acquire information $\mathbf{u}$ is

$$R(\mathbf{u}) = \mathbf{e}_{K^*}^T \mathbf{L}\hat{\mathbf{q}}(\mathbf{u}) - \mathbf{e}_{k^*}^T \mathbf{L}\hat{\mathbf{q}}(\mathbf{u}), \tag{12}$$

where the integers K^* and k^* satisfy

$$K^* = k \mid \mathbf{e}_k^T \mathbf{L}\hat{\mathbf{p}} = \min, \tag{13}$$

$$k^* = k \mid \mathbf{e}_k^T \mathbf{L}\hat{\mathbf{q}}\,(\mathbf{u}) = \min, \tag{14}$$

and represent the ignorant and informed managers' decisions, respectively. Note that the informed decision k^* depends on the information $\mathbf{u}$ while the ignorant decision K^* does not. Hence, the first term of (12) represents the expected loss when using *ignorant* decision K^*. The second term represents the expected loss when using the more *informed* decision k^*. Each term is calculated using the informed manager's operational belief. Thus, the difference of the two terms is the informed manager's expectation of the value of information $\mathbf{u}$.

The regret function (12) values $\mathbf{u}$ – if it is known – by comparing the decisions of the ignorant and informed managers. However, for our information-valuation purposes, $\mathbf{u}$ is not known, and the goal is to value the potential information prior to knowing what information is to be revealed. Enabled by the existence of information distribution $g_{\mathbf{u}}$, an ignorant manager can now value an uncertainty reduction effort by calculating $E[R(\mathbf{u})]$. When $\mathbf{u}$ are continuous, the expected dollar value of pursuing an uncertainty reduction is

$$E[R(\mathbf{u})] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} R(u_1', \ldots, u_m') g_{\mathbf{u}}(u_1', \ldots, u_m') \mathrm{d}u_1' \ldots \mathrm{d}u_m'. \tag{15}$$

If $g_{\mathbf{u}}$ cannot be found analytically (which is often the case), one can still approximate (15) through

$$E[R(\mathbf{u})] \approx \frac{1}{M}\sum_{i=1}^{M} R(u_1^{(i)},\ldots,u_m^{(i)}), \tag{16}$$

where the m-tuples $(u_1^{(i)},\ldots,u_m^{(i)})$, $1 \leq i \leq M$ are realized by sampling $\mathbf{p} \sim f_{\mathbf{p}}$ and using (6).

Eq. (15) values information by considering all possible realizations of information $\mathbf{u}$ with each realization leading to a different value for regret. In addition, each realization leads to a different loss associated with the ignorant decision. The next lemma establishes a consistency between the distribution of the informed manager's expectation of loss associated with the ignorant decision and the ignorant manager's expectation of his own loss.

Lemma 2. *The expected loss associated with the ignorant manager's decision, as measured using the informed manager's perspective over all possible realizations of information* $\mathbf{u}$*, is the same as the ignorant manager's calculation of his own expected loss, i.e.*

$$E_{\mathbf{u}}[\mathbf{e}_{K^*}^T \mathbf{L}\hat{\mathbf{q}}(\mathbf{u})] = \mathbf{e}_{K^*}^T \mathbf{L}\hat{\mathbf{p}}. \tag{17}$$

Proof of Lemma 2 All proofs are provided in the Appendix.

Thus, our valuation of information $E[R(\mathbf{u})]$ can also be thought of as the difference between the ignorant manager's expected loss $\mathbf{e}_{K^*}^T \mathbf{L}\hat{\mathbf{p}}$ and the expectation of the informed manager's loss $E[\mathbf{e}_{k^*}^T \mathbf{L}\hat{\mathbf{q}}(\mathbf{u})]$.[1] This representation of regret is more common (see **?**, p.376).

4. Methodology for Pricing Information

With the theoretical preliminaries in place, we provide the steps to value information using only one's current information:

(1) **Form the ignorant belief distribution:** Apply the maximum entropy principle to create an ignorant belief distribution, $f_{\mathbf{p}}$, that is most consistent with current information.

(2) **Find the ignorant operational belief and associated decision:** From the ignorant belief distribution, form an operational belief, $\hat{\mathbf{p}}$; see eq. (10) and note the associated decision K^*.

(3) **Derive the information distribution:** Characterize how the information $\mathbf{u}$ is related to the demand distribution by determining the relation $\mathcal{M}(p_1,\ldots,p_N) = \mathbf{u}$ in (6). Since the distribution of $\mathbf{p}$ is known, the distribution $g_{\mathbf{u}}$ is also known. The information distribution may have a closed form; if not, it can be realized numerically.

(4) **Characterize the informed belief distributions:** For each possible $\mathbf{u}$, determine how the belief space is restricted and determine how the belief space is partitioned by different $\mathbf{u}$ values; see Fig. 1. Mathematically, the informed belief $\mathbf{q}$ follows a distribution $f_{\mathbf{q}}$ given by (5).

[1] For notational brevity the dependence of informed decision k^* on information $\mathbf{u}$ is not explicitly shown.

(5) **Find all informed operational beliefs and decisions:** For each informed belief distribution (of which there may be an infinite number) make an informed operational belief $\hat{\mathbf{q}}$; see eq. (11) and note the associated decision k^*.

(6) **Form the regret function:** Calculate the regret for each specific $\mathbf{u}$; see eq. (12). Typically, this is calculated as the difference in expected loss (under the informed operational belief) between decision K^* and k^*.

(7) **Compute expected regret:** Compute expected regret as the expectation of step 6 over all possible values of $\mathbf{u}$ and using the distribution for $\mathbf{u}$ from step 3. See eqs. (15) and (16).

5. Examples and Associated Derivations

In this section, the valuation of information using the maximum entropy techniques of §3 is demonstrated in various application settings. For each setting, numerical examples along with instructive derivations and insights are discussed.

5.1. Newsvendor Model

Our first example is that of the newsvendor problem; a canonical model in inventory management (**?**). The newsvendor faces a stochastic demand D with $P(D=j)=p_j$ and must pre-order i newspapers where $i,j\in\{1,2,\ldots,N\}$. Mismatches in order quantity and demand result in a loss, which is represented by a loss matrix $\mathbf{L}$ with components

$$L_{ij}=\begin{cases}(i-j)c_o, & \text{if } i\geq j,\\ (j-i)c_u, & \text{if } i<j,\end{cases} \tag{18}$$

and the underages and overages c_u and c_o are given. Given $\{p_1,\ldots,p_N\}$, the classical newsvendor problem is to determine the optimal order quantity and associated loss.

Now suppose that $(p_1,\ldots,p_N)$ are unknown, but N, c_o and c_u are known and the mean demand $u=p_1+2p_2+\ldots Np_N$ can be determined by surveying the market (for example). How much should one pay for this extra information? We perform our analysis for the particular case where $N=3$ and $c_u=c_o=\$50$.

We first solve the classical newsvendor problem, given an arbitrary belief in the demand (p_1,p_2,p_3). The loss is $50\min\left[2-2p_1-p_2,1-p_2,2p_1+p_2\right]$, and the associated optimal decision is

$$K^*=\begin{cases}1, & p_1>1/2,\\ 2, & p_1+p_2>1/2, p_1<1/2,\\ 3, & p_1+p_2<1/2.\end{cases} \tag{19}$$

where $p_1+p_2\leq 1$ and $p_1\geq 0$, $p_2\geq 0$. We now go through the seven steps in section 4 to value the mean.

Steps 1 and 2 - Form the ignorant belief distribution, the ignorant operational belief and associated decision: The differential entropy functional (3) with $m=1$, $B_1=C_1=1$ is

$$S^*[\phi]=-\int_0^1\int_0^{1-x_2}\phi(x_2,x_3)\mathrm{d}x_3\mathrm{d}x_2+\lambda_1\left[-1+\int_0^1\int_0^{1-x_2}\phi(x_2,x_3)\mathrm{d}x_3\mathrm{d}x_2\right]. \tag{20}$$

We wish to find ϕ that maximizes this functional. Introducing a small perturbation in ϕ, which we call $\delta\phi$, we have

$$S^*[\phi+\delta\phi]-S^*[\phi]=\int_0^1\int_0^{1-x_2}\delta\phi(-\log\phi-1+\lambda_1)\mathrm{d}x_3\mathrm{d}x_2-\int_0^1\int_0^{1-x_2}\frac{\delta\phi^2}{2\phi}\mathrm{d}x_3\mathrm{d}x_2+O(\delta\phi^3). \tag{21}$$

The first integral represents the first variation which must vanish for all $\delta\phi$ if ϕ is an extremizing function. Therefore $\phi=\exp(\lambda_1-1)=C$, a constant which must satisfy $\int_0^1\int_0^{1-x_2}C\mathrm{d}x_3\mathrm{d}x_2=1\Rightarrow C=2$. It follows that $\phi(x_2,x_3)$ is uniform over V_2^*, that is

$$\begin{aligned}\phi(x_2,x_3)&=2, &&(x_2,x_3)\in V_2^*,\\ \Rightarrow f_{\mathbf{p}}(x_1,x_2,x_3)&=\frac{2}{\sqrt{3}}, &&(x_1,x_2,x_3)\in V_2,\end{aligned} \tag{22}$$

and furthermore, (22) is a maximizer of eq. (20) since the second variation in (21) is strictly negative when $\delta\phi\neq 0$.

The operational belief is

$$(\hat{p}_2,\hat{p}_3)=E[(p_2,p_3)]=\int_0^1\int_0^{1-p_3}(2p_2,2p_3)\mathrm{d}p_2\mathrm{d}p_3=\left(\frac{1}{3},\frac{1}{3}\right),$$

so that $\hat{\mathbf{p}}=(1/3,1/3,1/3)$. The corresponding loss is $50\min(2-2\hat{p}_1-\hat{p}_2,1-\hat{p}_2,2\hat{p}_1+\hat{p}_2)=\$100/3$, with optimal order quantity $K^*=2$ units.

Step 3 - Derive the information distribution: We compute the distribution of $u=p_1+2p_2+3p_3$ given that $(p_2,p_3)\sim\phi$ in (22) and $p_1=1-p_2-p_3$. Using Lemma 1 and in particular eq. (7) when $N=3$, we have

$$g_u(y;3)=\begin{cases} y-1, & 1\le y\le 2,\\ 3-y, & 2\le y\le 3,\\ 0 & \text{otherwise.}\end{cases}$$

so that the information distribution is triangular on $x\in[1,3]$. This is consistent with intuition since more randomly generated 3-tuples from $f_{\mathbf{p}}$ will have mean close to 2 rather than the extreme values 1 and 3.

Step 4 - Characterize the informed belief distributions: Eliminating p_1 from

$$p_1+p_2+p_3=1, \tag{23}$$

$$p_1+2p_2+3p_3=u, \tag{24}$$

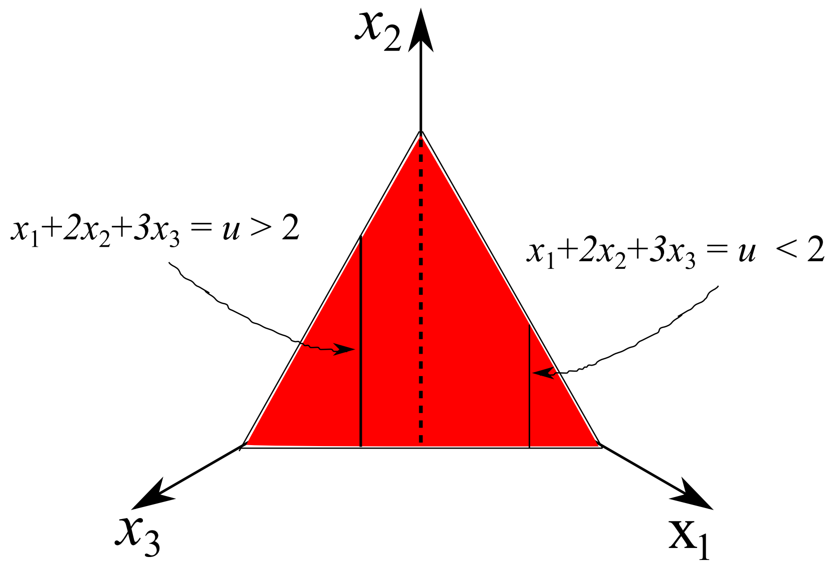

Figure 4 **Given extra information about mean demand, the ignorant belief space V_2, is restricted to the informed belief space (vertical lines). The informed belief space consists of the intersection of planes $x_1 + 2x_2 + 3x_3 = u$ with V_2. The dashed line shows possible informed beliefs when $u = 2$.**

we find that $p_2 = u - 1 - 2p_3$. For a given u, the informed belief space consists of (x_1, x_2, x_3) that lie on the intersection of (23) and (24); see Fig. 4. The (x_2, x_3) coordinates of the vertical lines in the figure satisfy $x_2 = u - 1 - 2x_3$. Therefore, the informed belief distribution is $q_3 \sim U(0, (u-1)/2)$ when $1 \le u \le 2$ and $q_3 \sim U(u-2, (u-1)/2)$ when $2 \le u \le 3$.

Step 5 - Find all informed operational beliefs and decisions: Taking expectations over $f_{\mathbf{q}}$, we find that

$$\hat{q}_3(u) = \begin{cases} (u-1)/4, & 1 \le u \le 2, \\ (3u-5)/4, & 2 \le u \le 3. \end{cases}$$

Since $\hat{q}_2 = u - 1 - 2\hat{q}_3$ and $\hat{q}_1 = 1 - \hat{q}_2 - \hat{q}_3$, we find that

$$\hat{q}_1(u) = \begin{cases} (7-3u)/4, & 1 \le u \le 2, \\ (3-u)/4, & 2 \le u \le 3, \end{cases}$$

$$\hat{q}_2(u) = \begin{cases} (u-1)/2, & 1 \le u \le 2, \\ (3-u)/2, & 2 \le u \le 3. \end{cases}$$

A geometric interpretation of $(\hat{q}_1, \hat{q}_2, \hat{q}_3)$ is that they are the coordinates of the centers of mass of the vertical lines in Fig. 4. Under these beliefs, the loss is

$$50 \min(2 - 2\hat{q}_1 - \hat{q}_2, 1 - \hat{q}_2, 2\hat{q}_1 + \hat{q}_2) = \begin{cases} 50 \min\left(u-1, \frac{3-u}{2}, 3-u\right), & 1 \le u \le 2, \\ 50 \min\left(u-1, \frac{u-1}{2}, 3-u\right), & 2 \le u \le 3. \end{cases}$$

Minimizing loss as a function of u implies that the optimal order quantity is

$$k^*(u) = \begin{cases} 1, & 1 \le u \le \frac{5}{3}, \\ 2, & \frac{5}{3} \le u \le \frac{7}{3}, \\ 3, & \frac{7}{3} \le u \le 3. \end{cases} \tag{25}$$

Steps 6 and 7- Compute the regret function and expected regret: The regret function is given by eq. (12) where $\mathbf{L} = \begin{bmatrix} 0 & 50 & 100 \\ 50 & 0 & 50 \\ 100 & 50 & 0 \end{bmatrix}$, and k^* and K^* are given by (25) and (19) respectively. The expected regret is computed stochastically by generating $(p_2^{(i)}, p_3^{(i)})$ from a Dirichlet(1,1,1) distribution, taking $p_1^{(i)} = 1 - p_2^{(i)} - p_3^{(i)}$ and computing

$$u^{(i)} = p_1^{(i)} + 2p_2^{(i)} + 3p_3^{(i)}, \quad i = 1, \ldots, M,$$
$$E[R(u)] \approx \frac{1}{M} \sum_{i=1}^{M} R(u^{(i)}) = \$7.41,$$

using $M = 10,000$ draws. If the newsvendor's quest for the mean is expected to cost more than \$7.41, then he acts using currently available information. Otherwise, pursuit of a belief in mean demand is advisable. This calculation illustrates the methodology's tremendous potential; valuation of an uncertainty reduction effort is possible without knowing the exact information that will be revealed! In addition, one can see from the loss function that the maximum value of regret for $K^* = 2$ (i.e. the EVDI) would be \$50 (i.e. if $u = 1$ or $u = 3$). Thus, the upper bound on regret is not tight and would be misleading as a proxy for valuing uncertainty reduction efforts.

5.2. Quick Response Model

Given more time, demand predictions for a short lifecycle product can be made with more certainty based upon early market feedback (see for example **?**). However, the need to start production before the selling season leads to two possible strategies to overcome the inherent forecast uncertainty. One supply strategy is to produce enough in the initial production run to accommodate any potential scenario for demand. Alternatively, a supply manager can employ a quick response strategy where shortened leadtimes allow for an additional production run after a demand signal is made more clear. The first production run is made to ensure enough supply is available to start the season and the additional production run ensures a closer match between supply and demand. Under a quick response strategy, total production is chosen with greater precision in forecasted demand.

To value the benefits of quick response, we assume that a history of forecasting situations, similar to those presented in **?**, provides the ignorant manager with an expectation of forecast accuracy. For our modeling purposes, the ignorant manager knows that beliefs exhibiting this historical forecast accuracy, as measured by the observed variance of previous forecast errors, are superior to other beliefs. The only problem is that the first moment of demand (i.e. mean demand) is simply unknowable without employing quick response. The ignorant manager's current information is that he knows the demand support and has an expectation of the demand variance exhibited by more plausible beliefs. His valuation question is whether to use quick response so that more information, in the form of a belief in expected demand, can be acquired. Our valuation of more information in this setting is similar to, albeit less restrictive, than that used by **?**. In their study, quick response valuation is accomplished by assuming an additional production run can be made using a perfectly accurate demand forecast. In contrast, the methodology presented here allows for shaping of the belief space in such a way that a manager's distributional uncertainty is modeled; highly plausible distributions can be differentiated from less plausible distributions through constraining (e.g. specifying moments of the demand distribution) or shaping (e.g. specifying moments of the belief distribution) of the belief space.

In the quick response model, the manager must pre-order i units in the face of stochastic demand D. As in the newsvendor model, assume $P(D=j)=p_j,\ i,j\in\{1,2,\ldots,N\}$. A mismatch between order quantity and demand results in a loss, represented by a loss matrix $\mathbf{L}$ whose components are given by eq. (18). The quick response model differs from the newsvendor model in that the ignorant and informed managers are able to forecast the variance of the demand distribution. There are two potential ways of modeling this knowledge. We could constrain the belief distributions to have non-zero density only for those beliefs that have variance σ_0^2; this is a stringent requirement. Alternatively, we can shape the entire population of plausible beliefs to have an expected variance, σ_0^2, such that

$$E[\sigma^2(p_1,p_2,\ldots,p_N)]=\sigma_0^2, \tag{26}$$

where $\sigma^2(p_1,\ldots,p_N)=\sum_{k=1}^N k^2p_k-\left(\sum_{k=1}^N kp_k\right)^2$ is the variance of the demand distribution. By specifying that ignorant beliefs have an *expected* variance, all N-tuples $(p_1,\ldots,p_N)$ are admissible, but those with $\sigma^2(p_1,\ldots,p_N)$ close to σ_0^2 are assigned greater probability density. The problem now is to value the mean demand $\sum_{k=1}^N kp_k$ given that N, c_o and c_u are known, and the $p_1,\ldots,p_N$ are unknown but constrained by (26).

To facilitate both analytic and graphical illustration of the methodology, we again consider a 3-decision problem. with $c_u=\$80$, $c_0=\$20$. Also assume that i and j are measured in hundreds of units. The loss matrix is

$$\mathbf{L}=\begin{bmatrix} 0 & 8,000 & 16,000 \\ 2,000 & 0 & 8,000 \\ 4,000 & 2,000 & 0 \end{bmatrix}.$$

Step 1 - Form the ignorant belief distribution: To find the ignorant belief distribution, we find the maximum entropy distribution subject to the additional constraint. We set $\phi(x_2,x_3)=f_{\mathbf{p}}(1-x_2-x_3,x_2,x_3)$, and find the ϕ that maximizes the entropy functional

$$\begin{aligned} S[\phi] = &-\int_0^1\int_0^{1-x_2}\phi(x_2,x_3)\log\phi(x_2,x_3)\mathrm{d}x_3\mathrm{d}x_2+\lambda_1\left[-1+\int_0^1\int_0^{1-x_2}\phi(x_2,x_3)\mathrm{d}x_3\mathrm{d}x_2\right]+\\ &\lambda_2\left[-\sigma_0^2+\int_0^1\int_0^{1-x_2}(4x_3+x_2-4{x_3}^2-{x_2}^2-4x_3x_2)\phi(x_2,x_3)\mathrm{d}x_3\mathrm{d}x_2\right] \end{aligned} \tag{27}$$

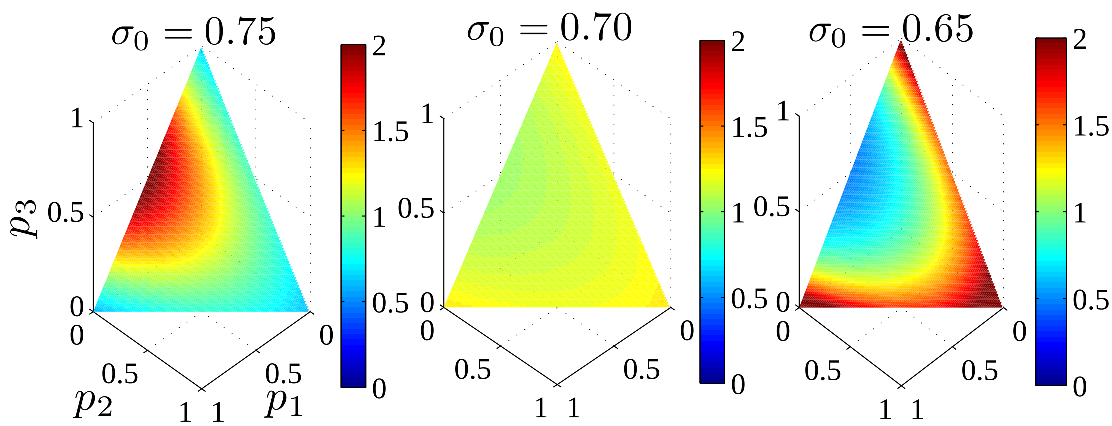

Figure 5 **Heat maps on the 2-simplex V_2 showing the maximum entropy distribution subject to a variance constraint** $(p_1 + 4p_2 + 9p_3) - (p_1 + 2p_2 + 3p_3)^2 = \sigma_0^2$.

The first term is the differential entropy term; the term proportional to λ_1 constrains ϕ to integrate to 1, and the term proportional to λ_2 imposes the constraint (26). To maximize S, note that

$$S[\phi + \delta\phi] - S[\phi] = \int_0^1 \int_0^{1-x_2} \delta\phi \left\{(-1 - \ln\phi) + \lambda_1 + \lambda_2[4x_3' + x_2' - 4{x_3'}^2 - {x_2'}^2 - 4x_3'x_2']\right\} \mathrm{d}x_3'\mathrm{d}x_2'$$
$$-\frac{1}{2}\int_0^1 \int_0^{1-x_2} \frac{\delta\phi^2}{\phi} \mathrm{d}x_3'\mathrm{d}x_2' + O(\delta\phi^3).$$

The second order term containing $\delta\phi^2$ is strictly negative. Therefore if

$$\phi(x_2, x_3) = a \exp\left[\lambda_2(4x_3 + x_2 - 4x_3^2 - x_2^2 - 4x_3x_2)\right], \tag{28}$$

then the functional (27) is maximized. The constants a and λ_2 satisfy

$$\sigma_0^2 = 1 - \frac{3}{2\lambda_2} + \frac{3e^{\lambda_2/4}\mathrm{Erf}(\sqrt{\lambda_2}/2)}{2e^{\lambda_2}\mathrm{Erf}(\sqrt{\lambda_2}) - 4e^{\lambda_2/4}\mathrm{Erf}(\sqrt{\lambda_2}/2)}, \tag{29}$$

$$a^{-1} = \sqrt{\pi}\lambda_2^{-3/2}\left[\frac{e^{\lambda_2}}{2}\mathrm{Erf}\left(\sqrt{\lambda_2}\right) - e^{\lambda_2/4}\mathrm{Erf}\left(\frac{\sqrt{\lambda_2}}{2}\right)\right], \tag{30}$$

where $\mathrm{Erf}(z)$ is the error function with (possibly complex argument): $\mathrm{Erf}(z) = \sqrt{\frac{2}{\pi}}\int_0^z e^{-t^2}\mathrm{d}t$. For a given σ_0, eqs. (29) and (30) can be solved numerically. Once a and λ_2 are known, they determine the ignorant manager's belief distribution through (28). In Fig. 5, the ignorant belief distribution is shown for three values of expected variance. As intuition suggests, reducing variance concentrates the density of beliefs around the 3-tuples representing certainty, namely (1,0,0), (0,1,0) and (0,0,1). Note that these extremal 3-tuples, corresponding to the vertices of the simplex, have the minimum variance $\sigma^2(\cdot,\cdot,\cdot)$ among all 3-tuples. Thus, in cases where σ_0 is small, a belief such as $\mathbf{p} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ is a far less plausible demand distribution than say a belief of $\mathbf{p} = (1, 0, 0)$.

Step 2 - Find the ignorant operational belief and associated decision: The ignorant manager's operational belief is

$$\begin{pmatrix}\hat{p}_2\\ \hat{p}_3\end{pmatrix} = a\int_0^1\int_0^{1-p_2}\begin{pmatrix}x_2\\ x_3\end{pmatrix}\exp\left[\lambda_2(4x_3+x_2-4x_3^2-x_2^2-4x_2x_3)\right]\mathrm{d}x_3\mathrm{d}x_2,$$
$$= \begin{pmatrix}\dfrac{1}{\lambda_2}-\dfrac{e^{\lambda_2/4}\mathrm{Erf}(\sqrt{\lambda_2}/2)}{e^{\lambda_2}\mathrm{Erf}(\sqrt{\lambda_2})-2e^{\lambda_2/4}\mathrm{Erf}(\sqrt{\lambda_2}/2)}\\ \dfrac{1}{2}-\dfrac{1}{2\lambda_2}+\dfrac{e^{\lambda_2/4}\mathrm{Erf}(\sqrt{\lambda_2}/2)}{2e^{\lambda_2}\mathrm{Erf}(\sqrt{\lambda_2})-4e^{\lambda_2/4}\mathrm{Erf}(\sqrt{\lambda_2}/2)}\end{pmatrix}. \quad (31)$$

with corresponding loss $2000\min\left[4\hat{p}_2+8\hat{p}_3, 1-\hat{p}_2+3\hat{p}_3, 2-\hat{p}_2-2\hat{p}_3\right]$, and optimal order quantity

$$K^* = \begin{cases}1, & \text{if } p_2+p_3<1/5,\\ 2, & \text{if } p_2+p_3>1/5 \text{ and } p_3<1/5,\\ 3, & \text{if } p_3>1/5.\end{cases}$$

and of course p_i are further subjected to the constraints $p_1+p_2\le 1$ and $p_i\ge 0$.

Note that in eq. (31), $(\hat{p}_2,\hat{p}_3)\to(0,1/2)$ as $\lambda_2\to\infty$ and $\to(2/3,1/6)$ as $\lambda_2\to-\infty$. Further analysis reveals that $\lambda\to\infty\Leftrightarrow\sigma_0\to 1$ and $\lambda\to-\infty\Leftrightarrow\sigma_0\to 0$. The locus of ignorant operational beliefs forms a finite line segment, as shown by the blue line in Fig. 6 with a the ignorant operational belief, $\hat{\mathbf{p}}$, and the associated optimal decision, K^*, determined by σ_0. We see that based on the operational belief, the only admissible order quantities are $K^*=2$ and $K^*=3$. The switch occurs when $(\hat{p}_2(\lambda_2),\hat{p}_3(\lambda_2))=(3/5,1/5)$ so that $\lambda_2\approx-22.078180$ and

$$K^* = \begin{cases}2, & \text{if } \sigma_0<\sigma_0^*,\\ 3, & \text{if } \sigma_0>\sigma_0^*.\end{cases} \quad (32)$$

where $\sigma_0^*\approx 0.316228$.

Step 3 - Derive the information distribution: In this example, the analytic form of g_u is more challenging to derive, so numerical realizations of g_u are used instead. Briefly, we sample M 3-tuples from (28) using a rejection-acceptance algorithm and then form the sum $u^{(i)}=p_1^{(i)}+2p_2^{(i)}+3p_3^{(i)}$ for $i=1,2,\ldots,M$. These $u^{(i)}$ are used to approximate the probability density function. Figure 7 shows six instances of the density function for various σ_0. The information distribution changes in interesting, but predictable ways. When σ_0 is small, the probabability density of u is concentrated at 1, 2, and 3. This is consistent with the heat map in Fig. 5(c). As σ_0 is increased, the distribution becomes closer to a triangular distribution. In particular, when $\sigma_0=0.7$, the heat map in (5) essentially has uniform density so it is not surprising that g_u resembles the one in Fig. 2(a) so closely. As σ_0 is further increased, $u=2$ becomes most likely since the most likely 3-tuple as $\sigma_0\to 1$ is $(p_1,p_2,p_3)=(1/2,0,1/2)$.

Steps 4 and 5- Calculate the informed belief distributions, operational beliefs and associated decisions: By restricting possible beliefs to satisfy $u=p_1+2p_2+3p_3$, we find that the informed belief distribution can be written purely as a function of x_3: $f_{\mathbf{q}}(x_3;u)=$

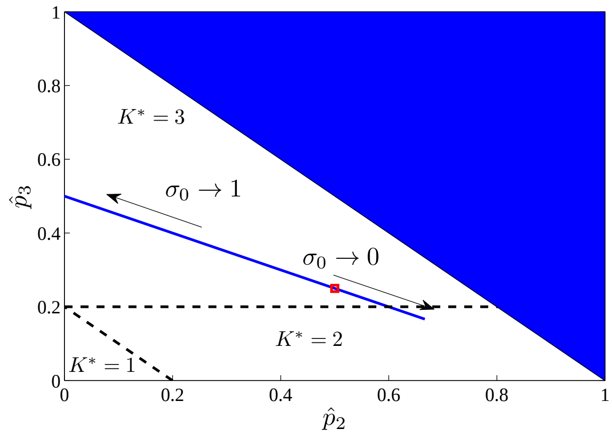

Figure 6 **Optimal order quantities K^* of the ignorant manager (measured in hundreds of units) vary depending on the manager's operational belief. The dashed lines partition the belief space. The blue line shows the locus of all ignorant operational beliefs. The red square corresponds to the point (0.5,0.25) corresponding to $\sigma_0 = 0.5$.**

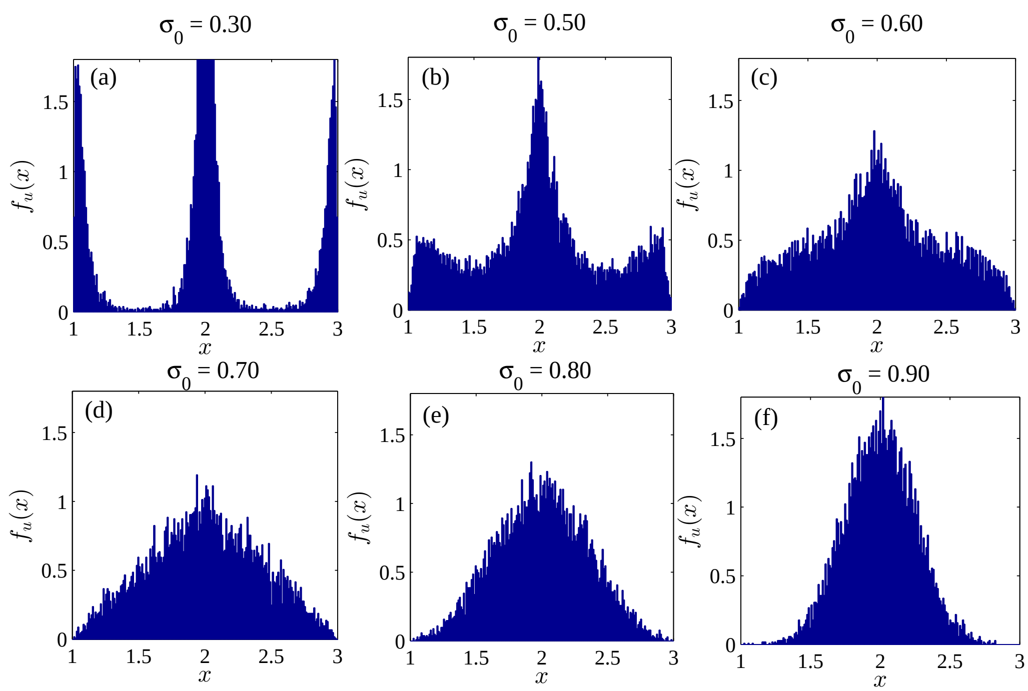

Figure 7 **Numerical realizations of f_u for six values of σ_0.**

$\tilde{a}\exp\left[\lambda_2\left(2x_3 - u^2 + 3u - 2\right)\right]$, where $\tilde{a} = (2\lambda_2 e^{\lambda_2(u^2-3u+3)})/(e^{\lambda_2 u} - e^{\lambda_2})$ if $1 \le u \le 2$ and $\tilde{a} = (2\lambda_2 e^{\lambda_2(u^2-4u+6)})/(e^{3\lambda_2} - e^{u\lambda_2})$, if $2 \le u \le 3$.

The informed manager's operational beliefs are found by taking expectations over $f_{\mathbf{q}}$. Closed form solutions for $\hat{\mathbf{q}}(\lambda_2, u)$ are available and are given in Appendix B. The corresponding loss is:

$$2000 \min \left[4\hat{q}_2(\lambda_2,u) + 8\hat{q}_3(\lambda_2,u), \hat{q}_1(\lambda_2,u) + 4\hat{q}_3(\lambda_2,u), 2\hat{q}_1(\lambda_2,u) + \hat{q}_2(\lambda_2,u)\right]. \tag{33}$$

Steps 6 and 7- Form the regret function and compute expected regret: Using (32), the regret function is

$$R(u) = -2000 \min \left[4\hat{q}_2 + 8\hat{q}_3, \hat{q}_1 + 4\hat{q}_3, 2\hat{q}_1 + \hat{q}_2\right] + \begin{cases} 2000(\hat{q}_1 + 4\hat{q}_3), & \text{if } \sigma_0 < \sigma_0^*, \\ 2000(2\hat{q}_1 + \hat{q}_2), & \text{if } \sigma_0 \geq \sigma_0^*. \end{cases} \tag{34}$$

The expectation is approximated using the sample mean and Monte Carlo simulation:

$$E[R(u)] \approx \frac{1}{M} \sum_{i=1}^{M} R(p_1^{(i)} + 2p_2^{(i)} + 3p_3^{(i)}),$$

where for $i = 1, \ldots, M$, $(p_2^{(i)}, p_3^{(i)})$ are drawn from the density (28) and $p_1^{(i)} = 1 - p_2^{(i)} - p_3^{(i)}$.

Fig. 8 shows the value of u as a function of σ_0. Thus, in situations where the mean is valued with prior knowledge of forecast accuracy (σ_0), one can determine if the benefits of a quick response initiative justify the costs. For example, at $\sigma_0 = 0.7$, the extra costs to enable quick response should be less than $\approx$ \$250 for those costs to be justified. To find a worst-case distribution for regret as a point of comparison, one has two choices. First, we can assume $\sigma_0 = 0.7$ to be the actual variance of the demand distribution (i.e. $\hat{q}_1 + 4\hat{q}_2 + 9\hat{q}_3) - (\hat{q}_1 + 2\hat{q}_2 + 3\hat{q}_3)^2 = 0.7$), then find the maximum regret (34) subject to this variance constraint. The answer turns out to be $\max_u[R(u)] = \$0$. All feasible distributions of the given variance lead to an unchanged decision. Alternatively, one can pursue maximizing regret without the variance constraint (as all beliefs are assumed possible) and discover that $\max_u[R(u)] = \$4,000$ when $u = 1$. In either case, the upper bound on the valuation of uncertainty reduction proves quite different from the results presented here. As a manager, one should be leery of using maximum regret as an information-valuation methodology and recognize that the value of quick response strategy is highly dependent upon the reduction uncertainty that can be achieved.

5.3. Revenue Management Model

We now examine the value of information for a pricing decision as opposed to the previously mentioned inventory decisions. In addition, this example is used to explore a larger decision/outcome space. Assume a manager faces a revenue management problem where his decision is to choose a price to optimize revenue over a finite selling period. This manager seeks to understand the potential value of information regarding the market size parameter of a linear demand curve (see **?**, and references therein for discussion of similar problems). The choice of a linear demand model

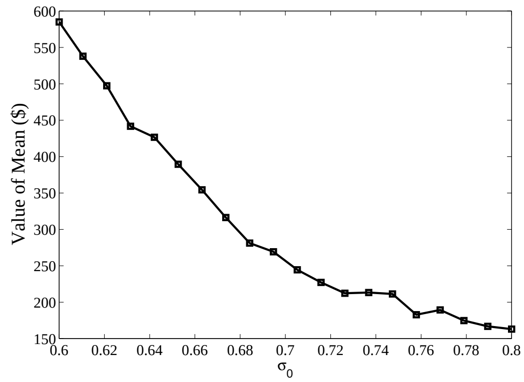

Figure 8 **Price of u as a function of σ_0 for the quick response problem. 10,000 trials were used in the expectation for each σ_0 value.**

for analysis is common in the literature (e.g. **??**) because they are often used in practice and are shown to provide good performance even when the model of demand is misspecified (**?**).

Assume demand to be a linearly decreasing function of price such that demand $D = A - Br$ where r is a non-negative price to be determined, and constants A and B are independent of r. $A > 0$ is the market potential (or size) and is a pure integer. $B > 0$ is the slope of the demand curve. For exposition, we assume inventory is sufficient and pricing decisions are made such that demand satisfies $0 \leq D \leq I$ for all relevant values of A and r. Suppose the slope of the demand curve B is known, but an ignorant manager's information regarding market potential A is unknown. How should information regarding the mean of A be valued?

We now state the problem more precisely. The demand follows a function of the form $D(r) = A - Br$. While B is known, A is unknown in the sense that its value follows some probability distribution with known support:

$$\text{Prob}(A = N_0 + i) = p_i, \qquad i = 1, 2, \ldots, N,$$
$$A \in \{N_0 + 1, N_0 + 2, \ldots, N_0 + N\},$$

where N and N_0 are known, but the probabilities p_i (which quantify the manager's belief in A) are *unknown*. Under a belief $\mathbf{p}$, the manager chooses a price r in order to maximize expected revenue $\mathcal{L}(\mathbf{p}, r) \equiv E[D(r)] \times r$. The problem is to calculate the additional revenue when the mean of A is known to the manager and therefore compute a fair price for it. Let us now follow the 7 steps in section 4 to price the mean of A.

Step 1 - Form the ignorant belief distribution: As in the previous examples, we may represent all possible beliefs with the ignorant belief distribution $f_{\mathbf{p}}$. Given that no information about $f_\mathbf{p}$ is known apart from the fact that it is a density on V_{N-1}, all beliefs are equally likely and the maximum entropy belief distribution is uniform over V^*_{N-1} (**?**):

$$f_\mathbf{p}(1-x_2-\ldots-x_N, x_2,\ldots,x_{N-1},x_N) = \begin{cases} (N-1)!, & (x_2,x_3,\ldots,x_N)\in V^*_{N-1}, \\ 0, & \text{otherwise.} \end{cases}$$

Steps 2 and 3 - Find the ignorant operational belief, associated decision and information distribution: Expected revenue $\mathcal{L}$ is given by the product of expected demand and price:

$$\mathcal{L}(\mathbf{p},r) = E[A]r - Br^2 = r\sum_{i=1}^{N} p_i(N_0+i) - Br^2, \tag{35}$$

and we see that $\mathcal{L}$ depends on the manager's beliefs $\{p_i\}$, specifically on his belief in the mean $E[A]$. His operational belief is $\hat{p}_i = E[p_i] = \int_{V_{N-1}} x_i f_\mathbf{p}(x_1,\ldots,x_N)\mathrm{d}x_1\ldots\mathrm{d}x_N$, so that $\hat{\mathbf{p}} = (1/N, 1/N, \ldots, 1/N)$. The optimal decision for this problem is to choose the price r_0 that maximizes the revenue $\mathcal{L}(\hat{\mathbf{p}},r) = r[N_0+(N+1)/2] - Br^2$:

$$r_0 = r \mid \mathcal{L}(\hat{\mathbf{p}},r) = \max \qquad \Longrightarrow \qquad r_0 = \frac{N_0+(N+1)/2}{2B}. \tag{36}$$

Consider now the introduction of new information, the mean of A: $v \equiv N_0 + \sum_{i=1}^N ip_i$. Then the information distribution for v is

$$g_v(y;N) = g_u(y-N_0;N). \tag{37}$$

where $g_u(\cdot;N)$ is given by eq. (7).

Steps 4 and 5 - Characterize the informed belief distributions, informed operational beliefs and associated decisions: In light of new information v, the informed belief distribution is found from eq. (5):

$$f_\mathbf{q}(x_1,\ldots,x_N,y) = \frac{f_\mathbf{p}(x_1,\ldots,x_N)I_{\Omega(y)}(x_1,\ldots,x_N)}{g_v(y;N)}, \tag{38}$$

where $\Omega(y) = \{\mathbf{x}\in V_{N-1} : \sum_{i=1}^N ix_i = y\}$. The operational beliefs are

$$\begin{aligned}(\hat{q}_1(y),\ldots,\hat{q}_N(y)) &= \int_{\Omega(y)} (x_1,\ldots,x_N) f_\mathbf{q}(x_1,\ldots,x_N,y)\mathrm{d}x_1\ldots\mathrm{d}x_N, \\ &= [f_u(y)]^{-1}\int_{\Omega(y)} (x_1,\ldots,x_N) f_\mathbf{p}(\mathbf{x}) I_{\Omega(y)}(\mathbf{x})\mathrm{d}x_1\ldots\mathrm{d}x_N. \end{aligned} \tag{39}$$

The results of steps 4 and 5 are not directly used because revenue (35) depends on the belief in the mean $E[A]$ rather than the individual probabilities q_i. Nevertheless, for sake of completeness, we have outlined how the informed belief distribution and informed operational beliefs could be calculated through (38) and (39). If the individual informed operational probabilites $\hat{q}_i$ were needed, the multi-dimensional integrals in (39) can be computed using monte-carlo integration.

Step 6 - Form the regret function: The expected revenue for the informed manager who acts on knowledge of v is

$$\mathcal{L}(\hat{\mathbf{q}}, r) = r\sum_{i=1}^{N}\hat{q}_i(N_0+i) - Br^2 = rv - Br^2. \tag{40}$$

Note that since the informed probabilities satisfy the constraint $\mathcal{M}(\mathbf{q}) = v$, so must the operational belief $\hat{\mathbf{q}}$ and we have $N_0 + \sum_{i=1}^{N} i\hat{q}_i = v$.

An informed manager would maximize his revenue by setting the price as $r^* = v/(2B)$ (maximizer of (40)) in which case his revenue is

$$\mathcal{L}(v, r^*) = \frac{v^2}{4B}. \tag{41}$$

We emphasize that in this problem, the revenue can be found directly from the new information, as opposed to finding an operational belief that is consistent with this information and using it to find the revenue. In other words, we use v to calculate $\mathcal{L}$ in (40) instead of the $\{\hat{q}_i\}$. (Accordingly, we have abused notation in (41) by replacing the first argument of (40) with v).

Now consider a manager who knows v but chooses to act as if he was ignorant of this information. His operational belief is $\hat{p}_i = 1/N$ for $i = 1, \ldots, N$, his operational belief in the mean is $N_0 + (N+1)/2$ and despite his knowledge of v, he sets a price r_0 as given by eq. (36). His corresponding revenue is

$$\mathcal{L}(v, r_0) = r_0 v - Br_0^2. \tag{42}$$

Therefore, the regret function is the difference between the revenues (41) and (42):

$$R(v) = \frac{1}{4B}\left[v - N_0 - \frac{N+1}{2}\right]^2.$$

Step 7 - Compute expected regret: Using (37) and the asymptotic property of g_u when $N \gg 1$ (relation (8)), we have

$$\begin{aligned}
E[R(v)] &= \int_{N_0+1}^{N_0+N} R(v')g_v(v';N)\mathrm{d}v', \\
&= \int_{N_0+1}^{N_0+N} R(v')g_u(v'-N_0;N)\mathrm{d}v', \\
&\sim \frac{1}{\sqrt{N}}\int_{-N/2}^{N/2} R(y+N_0+(N+1)/2)H(y/\sqrt{N})\mathrm{d}y, \\
&= \frac{1}{\sqrt{N}}\int_{-N/2}^{N/2} \frac{y^2}{4B}H\left(\frac{y}{\sqrt{N}}\right)\mathrm{d}y, \\
&\sim \frac{N}{4B}\int_{-\infty}^{\infty} z^2 H(z)\mathrm{d}z,
\end{aligned}$$

and recall that $H(\cdot)$ is a known even function which we determined empirically in Fig. 3(b). In other words, the value of u increases linearly with N and decreases linearly in B. From a manager's perspective, the value of market size information (i.e. the mean) is most useful for products/services with uncertain market potential and relatively inelastic demand. For example, let us examine the role of uncertainty in pricing airplane seats. Investing in reducing the demand uncertainty for routes with more inelastic demand (e.g. business travel routes) is more important than information pertaining to routes with elastic demand (e.g. pleasure travel); it would be of greater benefit to forecast the impact of large business conventions than say the impact of cruise ship locations on market potential.

6. CONCLUDING REMARKS

Our method of valuing information purely lives within the traditional notions of Bayesian statistical inference (see **?**). A Bayesian would not assert that his operational belief $\hat{\mathbf{p}}$ represents the "true" demand distribution, rather his operational belief is a construct purely used for decision making at a specific point in time. As additional information is obtained, the Bayesian's operational belief would then change as well. Historically, the laborious numerical tasks of working with and updating arbitrary belief distributions (and associated operational beliefs) led decision makers to jump to the use of conjugate distributions. The most relevant conjugate relationship to our work is that between a Dirichlet prior and information in the form of sample realizations generated from a multinomial distribution. The posterior distribution would then also be from the Dirichlet family of distributions (see **?**, §9.8).

In an information valuation context, conjugacy is leveraged as part of a preposterior analysis to estimate the expected value of sample information (EVSI) prior to knowing what sample might be collected (see example in **?**). Our methodology is analogous in philosophy to preposterior analysis, both methods use prior information to create a probability distribution over potential posterior distributions. Differences of our work from the more traditional preposterior analysis are: 1) distributional assumptions are not required; construction of maximum-entropy belief distributions enable prior beliefs to be made consistent with current knowledge, and 2) more general information distributions arise naturally from belief distributions; the information is not restricted to being a realization of demand, rather information about the demand distribution itself can also be valued.

Our methodology's reliance on potentially arbitrary prior distributions, without necessarily cojugate relationships, is less of a concern given the advances in both MCMC methods and computing power (**??**). As such, leveraging maximum entropy priors is not only theoretically justified, but practically useful as well. Reliance on maximum regret or expected value of perfect information (EVPI) estimates to serve as a proxy for information value is no longer required. Through examples (see §5.1 and §5.2), we have shown that these other methods of valuing information may be quite misleading in their information valuation; they can only find an upper bound on the value of uncertainty reduction. More realistic valuations of partial distributional information are possible with our methodology. As far as we know, it is the only methodology leading to *expected* information value without making distributional assumptions.

In our examples, pricing of the mean value of an unknown variable is performed; this proves to be the most interesting/relevant information for the examples presented. However, the methodology can value any function of the demand distribution including fractiles, other moments, and other more obscure functions. In addition, the methodology's use of entropy maximization to form belief distributions for information valuation is new and novel, but other definitions of entropy or other techniques to form belief distributions may also lead to fruitful results using similar techniques to those presented here. Additional future research pursuing other relevant and tractable (or easily computed) models is encouraged. Certainly, using different start and end states of information to represent practical scenarios can lead to context-specific insights. In addition, the information signals that we value, like the mean, are assumed to be accurate representations of the underlying demand distribution. In reality, the information distributions will include some noise and exploring the impact of noisy information signals is being taken up in a future paper. Another future path for expanding the work is to address the more general tradeoff between learning and revenue/cost optimization in multi-stage settings starting from the concept of a maximum entropy belief distribution. This work could easily serve as a springboard leading to new insights/models for the demand learning and revenue management literature streams.

Appendix A: Proof of Lemmas

Proof of Lemma 1 First we note that from the constraint $\sum_{i=1}^{N} p_i = 1$ that we can write u as a sum of $p_2, p_3, \ldots, p_N$ only since

$$u = 1 + \sum_{j=2}^{N} (j-1) p_j \equiv 1 + q,$$

and it is clear that $p_2 + p_3 + \ldots + p_N \leq 1$. We can now leverage the result from **?** to find the distribution function of q. Specifically, these authors derive the cumulative distribution function for a linear combination of random variables $\sum_{j=1}^{n} c_j Y_j$ where $c_1 > c_2 > \ldots > c_n > 0$ and Y_j are uniformly distributed over the simplex: $Y_j \geq 0$, $j = 1, \ldots, n$, $Y_1 + Y_2 + \ldots + Y_n \leq 1$. Adapting their result with $n = N-1$, we find that the cumulative distribution function of q is

$$F_q(x) = \frac{x^{N-1}}{c_1 c_2 \ldots c_{N-1}} - \sum_{j=r+1}^{N-1} \frac{(x - N + j)^{N-1}}{(N-j) \prod_{i=1 \neq j}^{N-1} (j-i)}$$

with $c_j = N - j$, $r = N - \lceil x \rceil$ and $0 \leq x \leq N-1$. The pdf of u is therefore $F_q'(y-1)$ for $1 \leq y \leq N$, which is exactly Eq.(7). Q.E.D.

Proof of Lemma 2

$$\begin{aligned}
E_{\mathbf{u}}[\mathbf{e}_{K^*}^T \mathbf{L}\hat{\mathbf{q}}(\mathbf{u})] &= \mathbf{e}_{K^*}^T \mathbf{L} E_{\mathbf{u}}[\hat{\mathbf{q}}(\mathbf{u})], \\
&= \mathbf{e}_{K^*}^T \mathbf{L} E_{\mathbf{u}}\left\{E[\mathbf{q}(\mathbf{u})]\right\}, \\
&= \mathbf{e}_{K^*}^T \mathbf{L} E_{\mathbf{u}}\left\{E[\mathbf{p}|\mathcal{M}(\mathbf{p}) = \mathbf{u}]\right\}, \\
&= \mathbf{e}_{K^*}^T \mathbf{L} E[\mathbf{p}], \\
&= \mathbf{e}_{K^*}^T \mathbf{L}\hat{\mathbf{p}},
\end{aligned}$$

where we have used the law of total expectation.

Appendix B: Deriving the Informed Operational Belief for the Quick-Response Model

These are calculated as:

$$\hat{\mathbf{q}}(\lambda_2, u) = \begin{pmatrix} \dfrac{e^{\lambda_2 u}(\lambda_2 u - 3\lambda_2 + 1) + e^{\lambda_2}(-2\lambda_2 u + 4\lambda_2 - 1)}{2\lambda_2(e^{\lambda_2} - e^{\lambda_2 u})} \\[2ex] \dfrac{e^{\lambda_2 u} - e^{\lambda_2}(\lambda_2 u - \lambda_2 + 1)}{\lambda_2(e^{\lambda_2 u} - e^{\lambda_2})} \\[2ex] \dfrac{e^{\lambda_2} + e^{\lambda_2 u}(\lambda_2 u - \lambda_2 - 1)}{2\lambda_2(e^{\lambda_2 u} - e^{\lambda_2})} \end{pmatrix},$$

if $1 \leq u \leq 2$ and

$$\hat{\mathbf{q}}(\lambda_2, u) = \begin{pmatrix} \dfrac{e^{\lambda_2 u} - e^{3\lambda_2}(\lambda_2 u - 3\lambda_2 + 1)}{2\lambda_2(e^{3\lambda_2} - e^{\lambda_2 u})} \\[2ex] \dfrac{e^{3\lambda_2} + e^{\lambda_2 u}(\lambda_2 u - 3\lambda_2 - 1)}{\lambda_2(e^{3\lambda_2} - e^{\lambda_2 u})} \\[2ex] \dfrac{e^{\lambda_2 u}(1 + 4\lambda_2 - 2\lambda_2 u) + e^{3\lambda_2}(\lambda_2 u - \lambda_2 - 1)}{2\lambda_2(e^{3\lambda_2} - e^{\lambda_2 u})} \end{pmatrix},$$

if $2 \leq u \leq 3$, and are used in eq. (33).