

A Closed-Form EVSI Expression for a Multinomial Data-Generating Process

Adam Fleischhacker*, Pak-Wing Fok†, Mokshay Madiman‡, Nan Wu§

January 3, 2023

Abstract

This paper derives analytic expressions for the expected value of sample information (EVSI), the expected value of distribution information (EVDI), and the optimal sample size when data consists of independent draws from a bounded sequence of integers. Due to challenges of creating tractable EVSI expressions, most existing work valuing data does so in one of three ways: 1) analytically through closed-form expressions on the *upper bound* of the value of data, 2) calculating the expected value of data using numerical comparisons of decisions made using simulated data to optimal decisions where the underlying data distribution is known, or 3) using variance reduction as proxy for the uncertainty reduction that accompanies more data. For the very flexible case of modelling integer-valued observations using a multinomial data-generating process with Dirichlet prior, this paper develops expressions that 1) generalize existing beta-Binomial computations, 2) do not require prior knowledge of some underlying “true” distribution, and 3) can be computed prior to the collection of any sample data.

1 Introduction

The seminal work of [34] introduced preposterior analysis, a Bayesian recipe for estimating the value of information (VOI) prior to knowing the informa-

*Department of Business Administration, University of Delaware, Newark, DE 19716, email: ajf@udel.edu

†Department of Mathematical Sciences, University of Delaware, Newark, DE 19716, email: pakwing@udel.edu

‡Department of Mathematical Sciences, University of Delaware, Newark, DE 19716, email: madiman@udel.edu

§Institute for Financial Services Analytics, University of Delaware, Newark, DE 19716, email: nanw@udel.edu

tion’s content. The *expected value of sample information* (EVSI), a particularly valuable VOI computation, values the information contained in sample observations prior to their collection. [34] include many closed-form and oft-used expressions for calculating EVSI under the assumption of quadratic loss. One such expression is for a Bernoulli data-generating process with beta prior distribution (a.k.a. a beta-Binomial model); each observation being either zero or one [34, Table 6.2, p. 191]. In this paper, we generalize the beta-binomial EVSI expression beyond binary-valued observations to the case where each data point is drawn from a bounded sequence of integers. These results expand the availability of tractable VOI expressions to a useful scenario where previously value could only be approximated or bounded when a closed-form expression was needed.

Depending on a modeler’s choices of actions, states of uncertainty, loss (or utility) functions, and probability models, tractable calculations of VOI may exist, but intractable formulations, especially for EVSI, are much more common. In fact, reputed statistician Dennis Lindley has remarked that the question of sample size “is embarrassingly difficult to answer” due to difficulties calculating EVSI [26]. More generally, [14] shows that simply characterizing the relationship between information and value is challenging; [14]’s work dispels the idea that information value will reliably exhibit monotonic relationships with information value determinants such as action flexibility, risk aversion, or a decision maker’s wealth.

While for some EVSI and VOI problems, closed-form solutions are attainable [34, 5, 4], value of information solutions are often difficult to formulate. Hence, many papers are known for their ability to characterize aspects of VOI expressions such as the distributional properties of the expected value of perfect information (EVPI) [28], the impact of an exogenous variable on EVPI [20], and the additivity of information value when multiple sources of uncertainty exist [21]. EVSI calculations, in particular, often result in intractable expressions of multiple integrals where only numerical methods can yield results [25]. Even then, many numerical methods still require further simplifying assumptions (see, e.g., [36]). While it is possible to approximate VOI computations via normal approximations (see, e.g., [30, 19]) or using a computationally intense simulation-based methodology (see, e.g., [10, 37]), closed-form expressions yield instantaneous and accurate value computations with more interpretable insights regarding the effects of prior beliefs and sample sizes.

In this paper, we provide a new EVSI calculation for a flexible (i.e. multinomial) data-generating process that adheres to three desiderata outlined in [34, p.44]:

Tractable	EVSI is easily calculated using a closed-form expression.
Rich	A decision maker’s prior beliefs and information are readily incorporated as part of the calculation.
Interpretable	The expression for EVSI provides insight as to the effects of prior beliefs and sample size choices on the expected value of a sample.

Generating Process	Conjugate Prior	Source
Bernoulli(θ)	$(\theta) \sim \text{Beta}$	[34] [32]
Poisson(λ)	$\lambda \sim \text{gamma}$	[34]
Normal(μ, σ)	$\mu \sim \text{Normal}, \sigma \text{ known}$	[34]
	$\mu \text{ known}, \sigma^2 \sim \text{inv. Gamma}$	[34]
	$\sigma^2 \sim \text{inv. Gamma}, \mu \sigma^2 \sim \text{Normal}$	[34]
Multinomial(t) ¹	$t \sim \text{Dirichlet}$	This Paper

Table 1: Position of this paper in comparison to other tractable EVSI calculations.

Shown in Table 1, our point of departure is generalizing the EVSI calculation for a Bernoulli data-generating process with beta prior (a.k.a. a beta-binomial model) to the case of a multinomial data generating process with Dirichlet prior. Rich treatment and illustrative examples surrounding EVSI calculations for the beta-binomial conjugacy can be found in [15]. Additionally, [32] provide explicit closed-form value of information computations for the beta-binomial case and is very close in spirit to this work, but does not investigate the Dirichlet-multinomial setting. In relation to the multinomial sampling process we explore in this paper, existing work has focused on non-utility based approaches where data is valued based on its ability to bound a parameter of interest within a certain level of precision [1, 6]. Our approach, in contrast, extends the utility-based valuation of sampling to a multinomial sampling environment to yield closed-form expressions for both EVSI and the expected value of distribution information (EVDI). Publication of analytically tractable expressions will be able

to supplant the still-present usage of Monte Carlo simulation in multinomial settings (see, e.g., [38]).

When closed-form EVSI expressions are unavailable, quantification of value created through uncertainty reduction typically relies on one of three techniques: 1) closed-form expressions on the *upper bound* of the value of data, 2) simulated comparisons between valuing decisions made by an oracle who knows the underlying data distribution to decisions made by a less-informed decision maker, or 3) using variance-reduction as a proxy for how data reduces underlying uncertainty in the data-generating process. For examples of the first type, [27] bound EVPI for a risk-averse decision maker and [40] place an upper bound on the value of knowing the true distribution when one already knows the mean and variance of that distribution. Examples of the second type often compare a Bayesian updating procedure to a known optimal solution [8, 29, 7, 35]. Lastly, computing the value of variance reduction independent of the specific quantity of data is also seen within the literature [11, 22].

2 Problem Setup

Despite substantial efforts, notation for preposterior analysis has not been standardized and is often a matter of personal taste [33]. To aid the reader with this paper’s notation surrounding its random variables and their realizations, we present the following summary breaking the notation into three levels of analysis:

1. Data/Sample. Data is an integer-valued random variable with support $\{0, 1, \dots, M\}$. Sample is a random vector referring to either a sequence of n data observations or a vector of counts representing the number of occurrences of each potential data value recorded in n observations.

	D :	A random variable representing a single data observation.
	d :	A single realization of D with integer valued support: $d \in \{0, 1, \dots, M\}$.
$X \equiv (X_1, \dots, X_n)$:		A random <i>vector</i> of n observations of D .
$x \equiv (x_1, \dots, x_n)$:		A realization of data vector X .
	\mathcal{D}^n :	The support of X when n realizations are observed.
	n_k :	the number of times that $k \in \{0, 1, \dots, M\}$ appears in x .
(n_0, n_1, \dots, n_M) :		A vector of counts of occurrences for each potential data value.

2. Data/Sampling Distributions. Data and sampling distributions are identical terms referring to the probability distribution governing the data-generating process. Data distribution refers to generating individual data points and sampling

¹With support interpreted as a sequence of integer values.

distribution preferred when talking about a sequence of observations.

- $T \equiv (T_0, T_1, \dots, T_M)$: A random vector representing a data distribution. Random elements T_k are data distribution parameters representing the probability of data realization being k .
- $t \equiv (t_0, t_1, \dots, t_M)$: A realization of random vector T such that $t_k = p(D = k)$ for $k \in \{0, 1, \dots, M\}$.
- t^* : The “true” data distribution or sampling distribution; only knowable by an oracle.
- \mathcal{T} : The space or set of all possible data distributions. $T, t, t^* \in \mathcal{T}$.

3. Prior/Posterior Distributions. Continuous multivariate probability distributions with domain of all possible data distributions.

- π : A prior from which data distributions are generated.
- π_X : A posterior that updates π in light of data X .

2.1 Modelling Data and Loss

Consider a *data-generating process* that generates independent and identically distributed samples from a bounded sequence of $M + 1$ integers. For notational simplicity, we rescale the sequence to be $[M] \equiv \{0, 1, \dots, M\}$. For practical motivation, the data could represent product demand and the goal is to make accurate predictions for inventory control [39]. For the specific case of demand uncertainty, we note that there are asymmetric and other loss functions that would be preferred to the quadratic loss function used here, but closed-form expressions are not forthcoming for those cases.

The data-generating process is governed by an unknown *data distribution*, t , with discrete-finite support $[M]$. Thus the statistical model for the data-generating process is parameterized by the standard M -dimensional simplex of probabilities

$$\mathcal{T} = \{t = (t_0, \dots, t_M) \in \mathbb{R}_+^{M+1} : t_0 + \dots + t_M = 1\};$$

this infinite (but finite-dimensional) parameter space describes how we are labeling the potential data distributions. If the sample size of the data is n , we have n values $x_1, \dots, x_n \in [M]$ being generated by the data-generating process. For a given $t \in \mathcal{T}$, the associated data-generating process $p_t^{(n)}$ assigns probability

$$p_t^{(n)}(x_1, \dots, x_n) = \prod_{i=1}^n t_{x_i} \tag{1}$$

to this particular sequence of data values. In particular, if the sample size is 1, the data-generating process is simply given by

$$p_t(d) \equiv p_t^{(1)}(d) = t_d, \quad d \in [M].$$

It is clear that the number of occurrences of particular data values in the sample is a sufficient statistic for the model described, and that the sampling distribution for this sufficient statistic is just the multinomial model. Specifically, if $n_d = |\{1 \leq i \leq n : x_i = d\}|$, then (n_0, \dots, n_M) is a sufficient statistic, and we have, with obvious abuse of notation,

$$p_t(n_0, \dots, n_M) = \binom{n}{n_0 \dots n_M} \prod_{d=0}^M t_d^{n_d}. \quad (2)$$

Note that $n_0 + \dots + n_M = n$ by definition; so we do not write the superscript (n) when using the sufficient statistic to represent the data.

When making predictions for future data, ideally the action (or prediction) is close to the actual data realization. For tractability, we consider a quadratic terminal opportunity *loss function* for a single prediction to be of the following form:

$$\ell(d, a) = k(d - a)^2 \quad (3)$$

where $k > 0$ is a known constant, a is the action/prediction, and $d \in [M]$ is the actual data realization.

To briefly make the above notation more concrete, let's imagine forecasting product demand for a product that will sell between 0 and 5 units ($M = 5$). Each period's i.i.d demand, $d \in \{0, 1, \dots, 5\}$, has an associated probability of occurrence, $p_t(0), p_t(1), \dots, p_t(5)$, which is represented more compactly as t_0, t_1, \dots, t_5 . The effectiveness of any action will be measured using quadratic loss scaled by a factor k such that if $k = 5$, $d = 4$, and $a = 1$, then $\ell(4, 1) = 45$. The decision maker is contemplating the value of $n = 3$ observations where generated data, (x_1, x_2, x_3) , might be something like $(0, 5, 0)$ and the associated sufficient statistic of counts, (n_0, \dots, n_5) , would be $(2, 0, 0, 0, 0, 1)$. Note that $t \equiv t_0, t_1, \dots, t_5$ parameterizes both the data-generating process of eq. (1) yielding (x_1, x_2, x_3) and the equivalent sampling process of eq. (2) yielding (n_0, \dots, n_5) . As a result, we refer to t as both data distribution and sampling distribution depending on context.

2.2 Preposterior Analysis

For any data distribution t , define the expectation of loss as:

$$R(t, a) = \mathbb{E}_{D|T=t} [\ell(D, a)] = \sum_{d=0}^M p_t(d) \ell(d, a). \quad (4)$$

where $R(t, a)$ is known as the Bayes risk. Since a decision maker (DM) does not know the underlying “true” $t^* \in \mathcal{T}$ data distribution, the minimum Bayes risk, $\min_a R(t^*, a)$, is likely unachievable.

For a DM, risk is evaluated on an average basis based on the probability distribution the DM places over the simplex \mathcal{T} . Without any sample observations, this distribution is the prior π over all possible data distributions in \mathcal{T} . The average risk of taking action a using prior π is

$$\bar{R}(\pi, a) = \mathbb{E}_T [R(T, a)], \quad (5)$$

with $T \sim \pi$. The Bayes action for π is

$$a^*(\pi) = \arg \min_{a \in \mathcal{A}} \bar{R}(\pi, a). \quad (6)$$

The Bayes risk for π is

$$\bar{R}(\pi, a^*(\pi)) = \min_{a \in \mathcal{A}} \bar{R}(\pi, a). \quad (7)$$

Access to a sample $X \equiv (X_1, \dots, X_n)$ results in a different decision with different risk. With sample observations, the DM applies Bayes’ rule to update π to π_X (the posterior) and calculates the associated optimal Bayes action $a^*(\pi_X)$. Since X is unknown prior to actually collecting the sample, the Bayes risk for π_X is itself a random variable. Hence, we evaluate the DM’s prior expectation of loss with sample information over all possible samples X ,

$$\mathbb{E}_X [\bar{R}(\pi_X, a^*(\pi_X))] = \mathbb{E}_T \mathbb{E}_{X|T} [R(T, a^*(\pi_X))], \quad (8)$$

with $T \sim \pi$ and the right-hand side expression derived by substituting π_X for π in eq. (5) and applying the law of total expectation.

Thus, the *expected value of a sample of information* (EVSI), $V_n(\pi)$, is the difference between the prior expectations of loss with and without sample X under prior π :

$$V_n(\pi) = \bar{R}(\pi, a^*(\pi)) - \mathbb{E}_X [\bar{R}(\pi_X, a^*(\pi_X))] \quad (9)$$

$$= \mathbb{E}_T [R(T, a^*(\pi))] - \mathbb{E}_T \mathbb{E}_{X|T} [R(T, a^*(\pi_X))] \quad (10)$$

where $T \sim \pi$ and eq. (10) follows from eqs. (5) and (8). Proposition 2.1 formalizes our intuition that this expected value of sample information should be non-negative.

Proposition 2.1. *Suppose data distribution $T \equiv (T_0, \dots, T_M)$ is drawn from a given prior π . Assume further that a DM is given n samples $X \equiv (X_1, \dots, X_n)$ and updates his/her prior to the posterior π_X . Then, under quadratic loss, the expected value of these n samples is non-negative, i.e.*

$$V_n(\pi) = \mathbb{E}_T [R(T, a^*(\pi))] - \mathbb{E}_T \mathbb{E}_{X|T} [R(T, a^*(\pi_X))] \geq 0. \quad (11)$$

Proof. See Appendix. □ □

Because the ordering within the sample X does not matter, the inner expectation in (11) is performed over $(n_0, n_1, \dots, n_M) \sim \text{Multinomial}(t)$ conditioned on $T = t$ where n_j is the number of times that $j \in [M]$ appears in the sample, and the outer expectation is performed over $T \sim \pi$.

3 Tractable Valuation of Sample Information

To arrive at a tractable valuation for (10), we leverage the Dirichlet distribution as a prior for three reasons: 1) it is a conjugate prior to categorical/multinomial outcomes, 2) its support is the M -dimensional simplex \mathcal{T} , and 3) it has flexibility to model many types of prior information for the decision maker. With the Dirichlet assumption, the main result of this paper, Theorem 3.1, can be presented:

Theorem 3.1. *For data distribution T with support $[M]$ and prior $\pi = \text{Dirichlet}(\alpha_0, \alpha_1, \dots, \alpha_M)$, the expected reduction in quadratic loss after observing n data samples, also called the expected value of sample information (EVSI), is given by:*

$$V_n(\pi) = \frac{kn(c_2 - c_1^2)}{(n + \alpha)(1 + \alpha)}. \quad (12)$$

where $\alpha = \sum_{d=0}^M \alpha_d$ is the precision/concentration parameter of the Dirichlet distribution (see [16]) and $c_1 = \frac{1}{\alpha} \sum_{d=0}^M d\alpha_d$ and $c_2 = \frac{1}{\alpha} \sum_{d=0}^M d^2\alpha_d$ are the first and second moments of the data under the marginal likelihood $(\alpha_1, \alpha_2, \dots, \alpha_M)/\alpha$.

Proof. See Appendix. □ □

Theorem 3.1 gives the expected value of observing an n -trial multinomial sample with Dirichlet prior where support of the underlying data-generating process is the bounded sequence of integers $[M] = \{0, 1, \dots, M\}$. This is a natural generalization of valuing an n -trial binomial sample with beta prior where support of the underlying data-generating process is restricted such that $[M] = \{0, 1\}$. With just a slight change of notation, we know from [32] that EVSI for the beta-binomial case in closed-form is:

$$\frac{kn}{n + \alpha_0 + \alpha_1} \cdot \frac{\alpha_0 \alpha_1}{(\alpha_0 + \alpha_1)^2 (\alpha_0 + \alpha_1 + 1)} \quad (13)$$

where $\pi \sim \text{Beta}(\alpha_0, \alpha_1)$. Replacing this prior with the equivalent Dirichlet parameterization of $\pi \sim \text{Dirichlet}(\alpha_0, \alpha_1)$ and using Theorem 3.1 yields an identical result:

$$\begin{aligned} V_n(\pi) &= \frac{kn(c_2 - c_1^2)}{(n + \alpha)(1 + \alpha)} \\ &= \frac{kn}{(n + \alpha_0 + \alpha_1)} \cdot \frac{\frac{\alpha_1}{\alpha_0 + \alpha_1} - \frac{\alpha_1^2}{(\alpha_0 + \alpha_1)^2}}{(\alpha_0 + \alpha_1 + 1)} \\ &= \frac{kn}{(n + \alpha_0 + \alpha_1)} \cdot \frac{\alpha_0 \alpha_1}{(\alpha_0 + \alpha_1)^2 (\alpha_0 + \alpha_1 + 1)} \end{aligned} \quad (14)$$

As a direct consequence of Theorem 3.1, when $n \rightarrow \infty$, we have an expression for the *expected value of distribution information* (EVDI), as an infinite sample gives the data distribution exactly:

$$V_\infty(\pi) = \lim_{n \rightarrow \infty} V_n(\pi) = \frac{k(c_2 - c_1^2)}{1 + \alpha}. \quad (15)$$

Lastly, we can express the efficiency η of the sample information as a function of the number of sample points using the ratio of (12) to (15) as:

$$\eta = \frac{n}{n + \alpha}. \quad (16)$$

Hence, the percentage of value obtained through sampling is given by the ratio of the number of data points n to the sum of the n data points and the concentration parameter α of a Dirichlet distribution. This sampling efficiency calculation directly simplifies to the known formula of the beta-binomial case from [34](in our notation): $\eta = n/(\alpha_0 + \alpha_1)$ where $\pi \sim \text{Beta}(\alpha_0, \alpha_1)$.

Again, we make the notation more concrete, by revisiting our forecasting product demand example from the end of §2.1. Recall, we have a product

that will sell between 0 and 5 units ($M = 5$) and loss is scaled by $k = 5$. The decision maker is contemplating the value of $n = 3$ observations. Introducing a zero-inflated prior $\pi \sim \text{Dirichlet}(\frac{10}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$ means $\alpha = \frac{15}{6}$, $c_1 = \frac{6}{15} \cdot (0 \cdot \frac{10}{6} + 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6}) = 1$, $c_2 = \frac{6}{15} \cdot (0 \cdot \frac{10}{6} + 1 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 9 \cdot \frac{1}{6} + 16 \cdot \frac{1}{6} + 25 \cdot \frac{1}{6}) = \frac{11}{3}$. Plugging into eq. (12) yields EVSI $V_3(\pi) = \frac{160}{77} \approx 2.08$. and EVDI $V_\infty(\pi) = \frac{80}{21} \approx 3.81$. From eq. (16) we get $\eta = \frac{6}{11} \approx 54.5\%$, so the learning from $n = 3$ samples is expected to provide more than half of the maximum possible reduction in loss. Following from eqs. (26) - (31), $a^*(\pi) = 1$ and prior expected loss $\bar{R}(\pi, a^*(\pi)) = 5 \cdot (-1^2 \cdot \frac{10}{15} + 0^2 \cdot \frac{1}{15} + 1^2 \cdot \frac{1}{15} + 2^2 \cdot \frac{1}{15} + 3^2 \cdot \frac{1}{15} + 4^2 \cdot \frac{1}{15}) = \frac{40}{3} \approx 13.33$. And thus, we can also get the prior expectation of posterior loss $\mathbb{E}_X [\bar{R}(\pi_X, a^*(\pi_X))] = \bar{R}(\pi, a^*(\pi)) - V_3(\pi) = \frac{40}{3} - \frac{160}{77} \approx 11.26$.

4 Notes on Richness and Interpretability of Modeling Assumptions

In the previous section, we showed one of the three EVSI desiderata, tractability, can be achieved for a multinomial data-generating process with Dirichlet prior. The multinomial distribution is flexible enough to model any discrete (finite) data distribution. Its prior, the Dirichlet distribution, is also flexible in its ability to model a wide range of distributions over a simplex. Yet, some sacrifice of richness in modeling prior beliefs is made in the name of tractability. Most notably, a more rich/flexible alternative prior over a simplex is the logistic-normal distribution [3, see discussion in]. The most glaring weakness of the Dirichlet distribution is in modeling prior beliefs where there is some type of correlation structure between data observations. For example, observing a high data value, say 100, would make one think values of 101 and 99 are also more likely to occur than data values further away. However, the Dirichlet distribution, as a prior distribution to multinomial data, is unable to capture this structure. Notably, the distribution-free underpinnings of the Kaplan-Meier estimator also ignore this potential correlation among data observations, yet shows favorable results in a similar repeated newsvendor setting [17].

The richness of the Dirichlet prior is best seen through the lens of its intuitive reparameterization [16]. Let the concentration parameter $\alpha = \sum_{i=0}^M \alpha_i$ and let the vector $\mathbf{m} = (\frac{\alpha_0}{\alpha}, \frac{\alpha_1}{\alpha}, \dots, \frac{\alpha_M}{\alpha})$ represent the mean where the expected mean of the data observations is given as $c_1 = \frac{1}{\alpha} \sum_{i=0}^M i \alpha_i = \sum_{i=0}^M i m_i$. When α is small, say $\alpha \leq M$, the prior distribution over the simplex can differ greatly from \mathbf{m} and reflect a decision maker's uncertainty

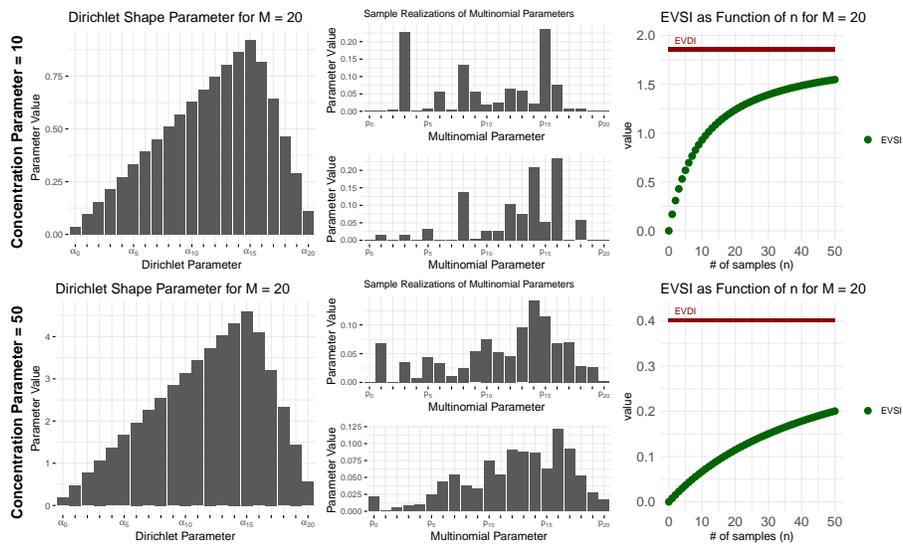


Figure 1: Graphical depiction of the Dirichlet prior parameters, potential realizations for that prior (i.e. the multinomial parameters), and the EVSI/EVDI calculations as a function of n samples for the given prior. Top row for concentration parameter $\alpha = 10$ and bottom row for concentration parameter $\alpha = 50$

around their expectation. As α is made larger, the prior distribution will concentrate probability density near \mathbf{m} and reflect greater confidence. We present a graphical overview of this in Figure 1 for two different concentration parameters. As seen, when α is smaller (top row of Figure 1) the realized multinomial parameters (middle-top plot) can be further away from the mean \mathbf{m} (which is proportional to the parameters in the top-left plot). As α increases (bottom-row) the prior distribution becomes much more informative and multinomial parameters will most likely mirror the prior Dirichlet parameters.

In terms of interpretability, Theorem 3.1 formalizes our intuition about what drives the value of data. Specifically, data is valuable when 1) the sample contains a lot of data (high n), 2) the expected variance of the data distribution is large (high $c_2 - c_1^2$), and 3) there is a lot of uncertainty regarding the true data distribution (α is small). Additionally, the calculation for EVDI (eq. 15) gives an interpretable upper bound on the value of data where high variance pushes to make samples more valuable and a high concentration parameter makes samples less valuable. Lastly, the equation for efficiency (16) adds further insight by stating how quickly the upper bound on the value of data is approached; basically, the smaller the Dirichlet concentration parameter, the more quickly EVDI is approached with each subsequent data point.

5 Illustrative Examples

In this section, we demonstrate how the tractable formulation for EVSI, equation (12), can serve as a building block inside of other research initiatives. The first example explores sample size optimization and the second example shows how a tractable EVSI calculation can lead to a tractable decision policy in a two-stage production planning problem. In the third/last example, the EVSI formula provides a foundation from which to benchmark heuristic updating procedures that seek to estimate an underlying unknown data distribution.

5.1 The Choice of Sample Size

We now explore a decision maker’s objective to choose the number of sample points to collect in such a way as to minimize his expected loss when assuming expected sampling cost, $C_s(n)$, is a linear function of the number of sampled points n :

$$C_s(n) = K + sn \tag{17}$$

where s is the cost of one sample and K represents the fixed costs of sampling.

The loss function to be minimized, $\ell_s(n)$, combines equations (12) and (17):

$$\ell_s(n) = -\frac{kn(c_2 - c_1^2)}{(n + \alpha)(1 + \alpha)} + K + sn \quad (18)$$

And assuming for practical purposes that n can be treated continuously, we get the optimal sample size:

$$n^* = \sqrt{\frac{\alpha}{(1 + \alpha)} \frac{k}{s} (c_2 - c_1^2)} - \alpha \quad (19)$$

for cases where n^* is positively valued and the fixed costs of sampling K can be recovered, i.e. $V_n(\pi) > C_s(n^*)$. In all other cases, $n^* = 0$. Equation (19) has a nice economic interpretation where the three terms represent the strength of the prior, the ratio between the scaling of the quadratic loss costs and the unit sampling costs, and the predicted variance of the data distribution.

5.2 Two-Stage Production Planning

The example shown here is a simple two-stage production planning problem (see, e.g., [9]) where the decision maker seeks to optimally schedule the 2nd production run.

Assume J periods make up a selling season. Each period, $j \in J$ faces independent and identical categorical demand with Dirichlet prior and quadratic loss (i.e. a repeated newsvendor setting with quadratic loss) with identical shipments scheduled for each period. A decision maker can choose either 1) to schedule the delivery quantity for each period in the entire selling season or, 2) at cost K can specify a period j^* after which the scheduled delivery quantity can be changed. Assuming this change date will be contractually set in advance of the selling season, find j^* to minimize expected net costs over the entire season J .

The net cost function for this problem is:

$$C(j) = \begin{cases} 0, & \text{if } j = 0, \\ K - (J - j) \frac{kj(c_2 - c_1^2)}{(j + \alpha)(1 + \alpha)}, & \text{if } j \in (0, J] \end{cases} \quad (20)$$

When $j \in (0, J]$, the net cost function $C(\cdot)$ is strictly convex and has a unique global minimum value. The optimal period j^* is

$$j^* = \arg \min_{j \in \{0, 1, \dots, J\}} C(j)$$

When $\min C(j) = 0$ for $0 < j \leq J$, we choose $j^* = 0$.

For the case when $\min C(j) < 0$, we have

$$j^* = \sqrt{\alpha(J + \alpha)} - \alpha$$

Considering that j^* must be a non-negative integer, summarizing different cases we have the optimal j^* as

$$j^* = \begin{cases} 0, & \text{if } \min_{j \in [0, J]} C(j) = 0, \\ \arg \min_{j \in \{\lfloor j_0 \rfloor, \lceil j_0 \rceil\}} C(j), & \text{if } \min_{j \in [0, J]} C(j) < 0. \end{cases} \quad (21)$$

where $j_0 = \sqrt{\alpha(J + \alpha)} - \alpha$.

5.3 Benchmarking Data-Driven Algorithms

An active area of research is to propose algorithms for decisions in repeated settings where minimal assumptions about the underlying data distribution are known. These approaches include Sample Average Approximation (SAA) [24, 23], concave adaptive value estimation (CAVE) [12], and Second Order Belief Maximum Entropy (SOBME) [35]. When benchmarking these algorithms, it is customary to pick a handful of “true” distributions where the algorithm competes against a known optimal solution.

With the introduction of a closed-form EVSI calculation in the context of a Dirichlet prior, a more robust benchmarking scenario can be achieved. Instead of picking a “true” data distribution, we pick a “true prior” from the Dirichlet family with support matching the problem of interest. This prior can be used to then simulate “true” data distributions (as many as we want) by which we can estimate the reduction in squared loss as a function of n , the number of data samples. Given this setup, a comparison of a proposed algorithm can be made against a known optimal *updating procedure*. After all, it is the updating procedure that we are seeking to validate, and the optimal updating procedure to benchmark new algorithms against is, therefore, the Bayesian one detailed in the proof of Theorem 3.1 (see appendix).

As a proof of concept, Figure 2 is an example benchmarking the well-known sample average approximation (SAA) (see [24]) against the known optimal Bayesian updating procedure (BAYES) using a Dirichlet($\alpha_0, \alpha_1, \dots, \alpha_M$)

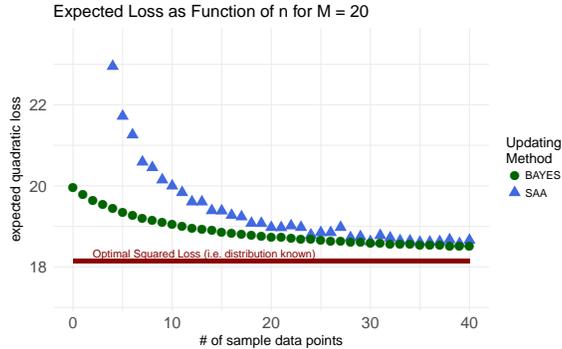


Figure 2: Comparing the sample average approximation(SAA) updating procedure to the known Bayesian (BAYES) optimal updating procedure.

prior with $M = 20$, $\alpha = 10$, and $\mathbf{m} \propto \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 13, 11, 9, 7, 5, 3, 1\}$ (chosen to be slightly skewed). In this scenario, we see the value of prior information in small data settings as BAYES outperforms SAA. It also shows how as the amount of data increases, the non-parametric SAA algorithm’s performance improves and closely mimics that of the optimal Bayesian updating procedure.

6 Conclusion

The use of preposterior analysis in this paper provides a formal method for valuing data prior to its collection and as such, should serve as a building block in many systems and models going forward. By expanding the support of the underlying data-generating process from $[M] = \{0, 1\}$ to $[M] = \{0, 1, \dots, M\}$, the beta-binomial EVSI calculations are successfully generalized to a Dirichlet-multinomial setting. Using this new EVSI computation, three illustrative examples valuing data prior to its collection are shown, there are potentially many other contexts where this tractable formulation might also prove useful. Researchers in two particular areas, medical decision making and active (machine) learning are known to be interested in EVSI types of calculations (see, e.g., [2, 13, 18, 31]). And we look forward to hearing of other useful deployments for this method of valuing data prior to its collection.

A Proof of Proposition 2.1 and Theorem 3.1

A.1 Proof of Proposition 2.1

The expected value of sample information is

$$V_n(\pi) = \mathbb{E}_T [R(T, a^*(\pi))] - \mathbb{E}_T \mathbb{E}_{X|T} [R(T, a^*(\pi_X))] . \quad (22)$$

For the first term in eq. (22), we have

$$\begin{aligned} \mathbb{E}_T [R(T, a^*(\pi))] &= k \mathbb{E}_T \left[\mathbb{E}_{D|T} \left[(D - a^*(\pi))^2 \right] \right] \\ &= k \mathbb{E}_T \left[\mathbb{E}_{D|T} \left[(D - \mathbb{E}[D])^2 \right] \right] \\ &= k \mathbb{E}_D \left[(D - \mathbb{E}[D])^2 \right] \\ &= k \text{Var} [D] . \end{aligned} \quad (23)$$

The second line is due to the optimal action under squared loss being the mean (see eq. (30)). The third line of equation (23) follows from the *law of total expectation*. Thus, the optimal Bayes risk without sample information under quadratic loss (3) is the marginal variance of D scaled by a factor k .

Similarly, for the second term in eq. (22) we find

$$\begin{aligned} \mathbb{E}_T \left\{ \mathbb{E}_{X|T} [R(T, a^*(\pi_X))] \right\} &= k \mathbb{E}_T \left\{ \mathbb{E}_{X|T} \left[\mathbb{E}_{D|T} \left[(D - a^*(\pi_X))^2 \right] \right] \right\} \\ &= k \mathbb{E}_T \left\{ \mathbb{E}_{X|T} \left[\mathbb{E}_{D|T} \left[(D - \mathbb{E}_{D|X} [D])^2 \right] \right] \right\} \\ &= k \mathbb{E}_X \left\{ \mathbb{E}_{D|X} \left[(D - \mathbb{E}_{D|X} [D])^2 \right] \right\} \\ &= k \mathbb{E}_X [\text{Var}_{D|X} [D]] . \end{aligned} \quad (24)$$

The optimal Bayes risk under quadratic loss (3) if a sample of size n is to be collected is the expected variance of the predictive posterior distribution of D scaled by a factor k .

Combining (22), (23), and (24), we complete the proof:

$$\begin{aligned} V_n(\pi) &= \mathbb{E}_T [R(T, a^*(\pi))] - \mathbb{E}_T \left\{ \mathbb{E}_{X|T} [R(T, a^*(\pi_X))] \right\} \\ &= k \text{Var} [D] - k \mathbb{E}_X [\text{Var}_{D|X} [D]] \\ &= k \left\{ \text{Var} [D] - \mathbb{E}_X [\text{Var}_{D|X} [D]] \right\} \\ &= k \text{Var}_X [\mathbb{E}_{D|X} [D]] \\ &\geq 0 . \end{aligned} \quad (25)$$

The last equal sign in equation (25) follows from the *law of total variance*. Since $k > 0$ and $\text{Var}_X [\mathbb{E}_{D|X} [D]] \geq 0$ for any X , we have $V_n(\pi) \geq 0$ for any sample size n . \square

A.2 Proof of Theorem 3.1

Consider the prior distribution for the data-generating process

$$\pi = \text{Dirichlet}(\alpha_0, \alpha_1, \dots, \alpha_M).$$

Suppose our information consists of n samples of the data distribution. Let n_j , $j \in [M]$ be the frequency of the data being j so that n_j are integers such that $\sum_{j=0}^M n_j = n$. Then, because the multinomial and Dirichlet distributions are conjugate,

$$\pi_X = \text{Dirichlet}(\alpha_0 + n_0, \alpha_1 + n_1, \dots, \alpha_M + n_M).$$

Because π and π_X both belong to the same class of distributions, we derive closed-form valuations for the information X . The corresponding marginal likelihoods for π and π_X are

$$\begin{aligned} q_\pi(d) &= \frac{\alpha_d}{\alpha}, \\ q_{\pi_X}(d) &= \frac{\alpha_d + n_d}{\alpha + n}, \end{aligned}$$

where $\alpha = \sum_{i=0}^M \alpha_i$. If the information happens to occur in such a way that $n_j \propto \alpha_j$ for each j , then the updated marginal likelihood is unchanged: $q_d(\pi) = q_d(\pi_X)$, $d \in [M]$.

For convenience, define the quantities

$$\begin{aligned} Z &= \frac{1}{n} \sum_{d=0}^M d n_d, \\ c_1 &= \frac{1}{\alpha} \sum_{d=0}^M d \alpha_d, \\ c_2 &= \frac{1}{\alpha} \sum_{d=0}^M d^2 \alpha_d, \end{aligned}$$

where Z represents the average frequency of the sample, c_1 the prior expectation for a sample value, and c_2 the prior second moment for the sample value.

Given the loss function in (3), the Bayes risk and action without sample information can be explicitly calculated

$$\bar{R}(\pi, a) = \mathbb{E}_{T \sim \pi}[R(T, a)], \quad (26)$$

$$= \mathbb{E}_{T \sim \pi} \left[\sum_{d=0}^M p_T(d) \ell(d, a) \right], \quad (27)$$

$$= \sum_{d=0}^M \ell(d, a) \mathbb{E}_{T \sim \pi}[p_T(d)], \quad (28)$$

$$= \sum_{d=0}^M \ell(d, a) q_\pi(d), \quad (29)$$

where $\{q_\pi(0), q_\pi(1), \dots, q_\pi(M)\}$ is the *marginal likelihood*. The Bayes action minimizes eq. (29):

$$\frac{\partial \bar{R}(\pi, a)}{\partial a} = -2k \sum_{d=0}^M (d - a) q_\pi(d) = -2k \left(\sum_{d=0}^M d q_\pi(d) - a \sum_{d=0}^M q_\pi(d) \right) = 0,$$

$$\begin{aligned} \Rightarrow a^*(\pi) &= \sum_{d=0}^M d q_\pi(d), \\ &= \mathbb{E}_{q_\pi}[D], \end{aligned} \quad (30)$$

$$= c_1. \quad (31)$$

the mean data outcome under the prior marginal likelihood. The corresponding Bayes Risk is

$$\begin{aligned} \bar{R}(\pi, a^*(\pi)) &= k \sum_{d=0}^M (d - a^*(\pi))^2 q_\pi(d), \\ &= k \text{Var}_{q_\pi}[D], \\ &= k(c_2 - c_1^2). \end{aligned}$$

Similarly, with sample information we have

$$\begin{aligned}
\frac{\partial \bar{R}(\pi_X, a)}{\partial a} &= -2k \sum_{d=0}^M (d - a) q_{\pi_X}(d), \\
&= -2k \left(\sum_{d=0}^M d q_{\pi_X}(d) - a \sum_{d=0}^M q_{\pi_X}(d) \right) = 0, \\
\Rightarrow a^*(\pi_X) &= \sum_{d=0}^M d q_{\pi_X}(d), \\
&= \mathbb{E}_{q_{\pi_X}}[D], \\
&= \frac{\alpha c_1 + nZ}{\alpha + n}, \tag{32}
\end{aligned}$$

which is the mean data outcome under the posterior marginal likelihood. Now expressing EVSI as

$$V_n(\pi) = \bar{R}(\pi, a^*(\pi)) - \mathbb{E}_T \mathbb{E}_{X|T} R(T, a^*(\pi_X)), \tag{33}$$

note the inner expectation is taken over the data frequency, which follows a multinomial distribution: $(n_0, \dots, n_M) \sim \text{Multinomial}(p_t(0), \dots, p_t(M))$, and the outer expectation is taken over all possible distributions $p_{t^*} \sim \text{Dir}(\alpha_0, \dots, \alpha_M)$.

The first term in (33) has already been evaluated as $k(c_2 - c_1^2)$. We now calculate the second term.

$$\begin{aligned}
R(t, a^*(\pi_X)) &= k \sum_{d=0}^M p_t(d) (d - a^*(\pi_X))^2 \\
&= k \sum_{d=0}^M p_t(d) \left[d - \frac{\alpha c_1 + nZ}{\alpha + n} \right]^2 \\
\Rightarrow \mathbb{E}_{X|T=t} [R(t, a^*(\pi_X))] &= k \sum_{d=0}^M p_t(d) \left[d^2 - \left(\frac{2nd}{\alpha + n} - \frac{2n\alpha c_1}{(\alpha + n)^2} \right) \mathbb{E}_X[Z] \right. \\
&\quad \left. - \frac{2d\alpha c_1}{\alpha + n} + \frac{\alpha^2 c_1^2}{(\alpha + n)^2} + \frac{n^2}{(\alpha + n)^2} \mathbb{E}_X[Z^2] \right]. \tag{34}
\end{aligned}$$

Since $Z(n_0, \dots, n_M) = \frac{1}{n} \sum_{d=0}^M dn_d$,

$$\begin{aligned}\mathbb{E}_{X|T=t}[Z] &= \sum_{d=0}^M dp_t(d), \\ \mathbb{E}_{X|T=t}[Z^2] &= \text{Var}_{X|T=t}[Z] + (\mathbb{E}_{X|T=t}[Z])^2 \\ &= \frac{1}{n} \sum_{d=0}^M d^2 p_t(d) + \frac{(n-1)}{n} \left(\sum_{d=0}^M dp_t(d) \right)^2,\end{aligned}$$

where the last line follows from the fact

$$\begin{aligned}\text{Var}_{X|T=t}[Z] &= \text{Var}_{X|T=t} \left[\frac{1}{n} \sum_{d=0}^M dn_d \right] \\ &= \frac{1}{n^2} \text{Var}_{X|T=t} \left[\sum_{d=0}^M dn_d \right] \\ &= \frac{1}{n^2} \left\{ \sum_{d=0}^M d^2 \text{Var}_{X|T=t}[n_d] + 2 \sum_{0=i<j}^M ij \text{Cov}_{X|T=t}(n_i, n_j) \right\} \\ &= \frac{1}{n^2} \left\{ \sum_{d=0}^M d^2 np_t(d) (1 - p_t(d)) - 2 \sum_{0=i<j}^M ij np_t(i) p_t(j) \right\} \\ &= \frac{1}{n} \sum_{d=0}^M d^2 p_t(d) - \frac{1}{n} \left\{ \sum_{d=0}^M d^2 p_t^2(d) + 2 \sum_{0=i<j}^M ij p_t(i) p_t(j) \right\} \\ &= \frac{1}{n} \sum_{d=0}^M d^2 p_t(d) - \frac{1}{n} \left(\sum_{d=0}^M dp_t(d) \right)^2.\end{aligned}\tag{35}$$

Eq. (34) becomes

$$\begin{aligned}\mathbb{E}_{X|T=t}[R(t, a^*(\pi_X))] &= k \left\{ \left(1 + \frac{n}{(\alpha+n)^2} \right) \sum_{d=0}^M d^2 p_t(d) + \left(\frac{2\alpha n c_1}{(\alpha+n)^2} - \frac{2\alpha c_1}{\alpha+n} \right) \sum_{d=0}^M dp_t(d) \right. \\ &\quad \left. + \left[\frac{n(n-1)}{(\alpha+n)^2} - \frac{2n}{\alpha+n} \right] \left[\sum_{d=0}^M dp_t(d) \right]^2 + \frac{\alpha^2 c_1^2}{(\alpha+n)^2} \right\}.\end{aligned}$$

The final step is to take the expectation over all possible beliefs $p_t \sim$

Dirichlet($\alpha_0, \dots, \alpha_M$). Using the fact that

$$\begin{aligned}
\mathbb{E}_{T \sim \pi}[p_T(i)] &= \frac{\alpha_i}{\alpha}, \\
\mathbb{E}_{T \sim \pi}[p_T(i)^2] &= \text{Var}[p_T(i)] + \mathbb{E}_T[p_T(i)]^2 \\
&= \frac{\alpha_i(\alpha - \alpha_i)}{\alpha^2(\alpha + 1)} + \frac{\alpha_i^2}{\alpha^2} \\
&= \frac{\alpha_i(\alpha_i + 1)}{\alpha(\alpha + 1)}, \\
\mathbb{E}_{T \sim \pi}[p_T(i)p_T(j)] &= \text{Cov}[p_T(i), p_T(j)] + \mathbb{E}_T[p_T(i)]\mathbb{E}_T[p_T(j)], \quad i \neq j, \\
&= -\frac{\alpha_i\alpha_j}{\alpha^2(\alpha + 1)} + \frac{\alpha_i\alpha_j}{\alpha^2} \\
&= \frac{\alpha_i\alpha_j}{\alpha(\alpha + 1)},
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}_{T \sim \pi} \left[\left(\sum_{d=0}^M dp_T(d) \right)^2 \right] &= \mathbb{E}_{T \sim \pi} \left[\sum_{d=0}^M d^2 p_T^2(d) + 2 \sum_{0=i<j}^M ij p_T(i)p_T(j) \right] \\
&= \sum_{d=0}^M d^2 \mathbb{E}_{T \sim \pi} [p_T^2(d)] + 2 \sum_{0=i<j}^M ij \mathbb{E}_{T \sim \pi} [p_T(i)p_T(j)] \\
&= \sum_{d=0}^M d^2 \frac{\alpha_d(\alpha_d + 1)}{\alpha(\alpha + 1)} + 2 \sum_{0=i<j}^M ij \frac{\alpha_i\alpha_j}{\alpha(\alpha + 1)} \\
&= \frac{1}{\alpha + 1} \sum_{d=0}^M \frac{d^2 \alpha_d}{\alpha} + \frac{1}{\alpha(\alpha + 1)} \left(\sum_{d=0}^M d^2 \alpha_d^2 + 2 \sum_{0=i<j}^M ij \alpha_i \alpha_j \right) \\
&= \frac{c_2}{\alpha + 1} + \frac{1}{\alpha(\alpha + 1)} \left(\sum_{d=0}^M d \alpha_d \right)^2 \\
&= \frac{c_2}{\alpha + 1} + \frac{\alpha c_1^2}{\alpha + 1}, \tag{36}
\end{aligned}$$

we obtain

$$\begin{aligned}
\mathbb{E}_T \mathbb{E}_{X|T}[R(T, a^*(\pi_X))] &= k \left\{ \left(1 + \frac{n}{(\alpha+n)^2} \right) \sum_{d=0}^M \frac{d^2 \alpha_d}{\alpha} + \left(\frac{2\alpha n c_1}{(\alpha+n)^2} - \frac{2\alpha c_1}{\alpha+n} \right) \sum_{d=0}^M \frac{d \alpha_d}{\alpha} \right. \\
&\quad \left. + \left[\frac{n(n-1)}{(\alpha+n)^2} - \frac{2n}{\alpha+n} \right] \left[\frac{c_2}{\alpha+1} + \frac{\alpha c_1^2}{\alpha+1} \right] + \frac{\alpha^2 c_1^2}{(\alpha+n)^2} \right\} \\
&= k(c_2 - c_1^2) \frac{\alpha(1+\alpha+n)}{(1+\alpha)(n+\alpha)}.
\end{aligned}$$

The value of n samples from the data distribution is therefore

$$\begin{aligned}
V_n(\pi) &= k(c_2 - c_1^2) - k(c_2 - c_1^2) \frac{\alpha(1+\alpha+n)}{(1+\alpha)(n+\alpha)} \\
&= \frac{kn(c_2 - c_1^2)}{(n+\alpha)(1+\alpha)}. \tag{37}
\end{aligned}$$

□

References

- [1] C. J. Adcock. An improved bayesian procedure for calculating sample sizes in multinomial sampling. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 42(2):91–95, 1993.
- [2] AE Ades, G Lu, and K Claxton. Expected value of sample information calculations in medical decision modeling. *Medical Decision Making*, 24(2):207–227, 2004.
- [3] J. Aitchison and S. M. Shen. Logistic-normal distributions: Some properties and uses. *Biometrika*, 67(2):261–272, 1980.
- [4] Debarun Bhattacharjya, Jo Eidsvik, and Tapan Mukerji. The value of information in portfolio problems with dependent projects. *Decision Analysis*, 10(4):341–351, 2013.
- [5] J Eric Bickel. The relationship between perfect and imperfect information in a two-action risk-sensitive problem. *Decision Analysis*, 5(3):116–128, 2008.
- [6] Jing Cao, J Jack Lee, and Susan Alber. Comparison of bayesian sample size criteria: Acc, alc, and woc. *Journal of statistical planning and inference*, 139(12):4111–4122, 2009.

- [7] Li Chen and Erica L Plambeck. Dynamic inventory management with learning about the demand distribution and substitution probability. *Manufacturing & Service Operations Management*, 10(2):236–256, 2008.
- [8] Gary D. Eppen and Ananth. V. Iyer. Improved fashion buying with bayesian updates. *Operations Research*, 45(6):805–819, 1997.
- [9] Marshall Fisher, Kumar Rajaram, and Ananth Raman. Optimizing inventory replenishment of retail fashion products. *Manufacturing & Service Operations Management*, 3(3):230–241, 2001.
- [10] Adam J. Fleischhacker and Pak-Wing Fok. An entropy-based methodology for valuation of demand uncertainty reduction. *Decision Sciences*, 46(6):1165–1198, 2015.
- [11] Yigal Gerchak and David Mossman. On the effect of demand randomness on inventories and costs. *Operations Research*, 40(4):804–807, 1992.
- [12] Gregory A. Godfrey and Warren B. Powell. An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution. *Management Science*, 47:1101–1112, 2001.
- [13] Robbie A Haertel, Kevin D Seppi, Eric K Ringger, and James L Carroll. Return on investment for active learning. In *Proceedings of the NIPS Workshop on Cost-Sensitive Learning*, volume 72, 2008.
- [14] Ronald W Hilton. The determinants of information value: Synthesizing some general results. *Management Science*, 27(1):57–64, 1981.
- [15] Ronald A Howard. Decision analysis: Perspectives on inference, decision, and experimentation. *Proceedings of the IEEE*, 58(5):632–643, 1970.
- [16] Jonathan Huang. Maximum likelihood estimation of dirichlet distribution parameters. *CMU Technique Report*, 2005. <http://jonathan-huang.org/research/dirichlet/dirichlet.pdf> accessed on 2017-09-19.
- [17] Woonghee Tim Huh, Retsef Levi, Paat Rusmevichientong, and James B. Orlin. Adaptive data-driven inventory control with cen-

- sored demand based on kaplan-meier estimator. *Operations Research*, 59(4):929–941, 2011.
- [18] Christopher Jackson, Anne Presanis, Stefano Conti, and Daniela De Angelis. Value of information: Sensitivity analysis and research design in bayesian evidence synthesis. *Journal of the American Statistical Association*, 2019.
- [19] Hawre Jalal and Fernando Alarid-Escudero. A gaussian approximation approach for value of information analysis. *Medical Decision Making*, 38(2):174–188, 2018.
- [20] Jeffrey Keisler. Comparative static analysis of information value in a canonical decision problem. *The Engineering Economist*, 49(4):339–349, 2004.
- [21] Jeffrey M Keisler. Additivity of information value in two-act linear loss decisions with normal priors. *Risk Analysis: An International Journal*, 25(2):351–359, 2005.
- [22] Jin Kyung Kwak and Srinagesh Gavirneni. Retailer policy, uncertainty reduction, and supply chain performance. *International Journal of Production Economics*, 132(2):271 – 278, 2011.
- [23] Retsef Levi, Georgia Perakis, and Joline Uichanco. The data-driven newsvendor problem: new bounds and insights. *Operations Research*, 63(6):1294–1306, 2015.
- [24] Retsef Levi, Robin O. Roundy, and David B. Shmoys. Provably Near-Optimal Sampling-Based Policies for Stochastic Inventory Control Models. *Mathematics of Operations Research*, 32(4):821–839, 2007.
- [25] Chi-Yuan Lin. Numerical techniques for evaluating sample information. *Technometrics*, 16(3):447–454, 1974.
- [26] Dennis V Lindley. The choice of sample size. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 46(2):129–138, 1997.
- [27] Abraham Mehrez. The effect of risk aversion on the expected value of perfect information. *Operations research*, 33(2):455–458, 1985.
- [28] Avraham Mehrez and Alan Stulman. Some aspects of the distributional properties of the expected value of perfect information (evpi). *Journal of the Operational Research Society*, 33(9):827–836, 1982.

- [29] Joseph M. Milner and Panos Kouvelis. Order quantity and timing flexibility in supply chains: The role of demand characteristics. *Management Science*, 51(6):970–985, 2005.
- [30] Satoshi Morita, Peter F Thall, and Peter Müller. Determining the effective sample size of a parametric prior. *Biometrics*, 64(2):595–602, 2008.
- [31] Jeremy Muesing, Nisar Ahmed, Luke Burks, Michael Iuzzolino, and Danielle Albers Szafr. Fully bayesian human–machine data fusion for robust online dynamic target characterization. *Journal of Aerospace Information Systems*, 18(2):26–49, 2021.
- [32] T. Pham-Gia and N. Turkkan. Sample size determination in bayesian analysis. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 41(4):389–397, 1992.
- [33] Howard Raiffa and Stephen E Fienberg. The early statistical years: 1947-1967 a conversation with howard raiffa. *Statistical Science*, pages 136–149, 2008.
- [34] Howard Raiffa and Robert Schlaifer. *Applied Statistical Decision Theory*. MIT press, Cambridge, Massachusetts, 1961.
- [35] Soroush Saghafian and Brian Tomlin. The newsvendor under demand ambiguity: Combining data with moment and tail information. *Operations Research*, 64(1):167–185, 2016.
- [36] Mark Strong, Jeremy E. Oakley, Alan Brennan, and Penny Breeze. Estimating the expected value of sample information using the probabilistic sensitivity analysis sample: A fast, nonparametric regression-based method. *Medical Decision Making*, 35(5):570–583, 2015.
- [37] Zhiyuan Wang, Zhiqiang Zheng, Wei Jiang, and Shaojie Tang. Blockchain-enabled data sharing in supply chains: Model, operationalization, and tutorial. *Production and Operations Management*, 30(7):1965–1985, 2021.
- [38] Wei Xiang and Wenxing Zhou. Optimal sample size determination based on bayesian reliability and value of information. In *13th International Conference on Applications of Statistics and Probability in Civil Engineering*, 2019.

- [39] Phillip M Yelland. Bayesian forecasting for low-count time series using state-space models: An empirical evaluation for inventory management. *International Journal of Production Economics*, 118(1):95–103, 2009.
- [40] Jinfeng Yue, Bintong Chen, and Min-Chiang Wang. Expected value of distribution information for the newsvendor problem. *Operations Research*, 54(6):1128–1136, 2006.