Proceedings of the 2016 Winter Simulation Conference
T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, eds.

LOT-SIZING IN SEQUENTIAL AUCTIONS WHILE LEARNING BID AND DEMAND DISTRIBUTIONS

Mahshid Salemi Parizi
Archis Ghate

Department of Industrial and Systems Engineering
University of Washington, Box 352650
3900 Northeast Stevens Way
Seattle, WA 98195, USA

ABSTRACT

Sellers often need to decide lot-sizes in sequential, multi-unit auctions, where bidder demand and bid distributions are not known in their entirety. We formulate a Bayesian Markov decision process (MDP) to study a profit maximization problem in this setting. We assume that the number of bidders is Poisson distributed with a Gamma prior on its mean, and that the bid distribution is categorical with a Dirichlet prior. The seller updates these beliefs using data collected over auctions while simultaneously making lot-sizing decisions until all inventory is depleted. Exact solution of our Bayesian MDP is intractable. We propose and numerically compare three approximation methods via extensive numerical simulations.

1 INTRODUCTION

Sellers often use sequential, multi-unit, online auctions of identical items as a revenue generation and inventory clearing tool (Chen, Ghate, and Tripathi 2011, Pinker, Seidmann, and Vakrat 2003, Pinker, Seidmann, and Vakrat 2010, van Ryzin and Vulcano 2004, Vulcano, van Ryzin, and Maglaras 2002).

Lot-size is a key design variable in multi-unit auctions. A large lot-size reduces bidder competition and hence may reduce the clearing price. The total revenue might still be high because the number of units sold is large. A small lot-size increases bidder competition and hence the clearing price, but the total revenue might still be limited because just a few units are sold and a higher holding cost is incurred. Bidders in different auctions also compete indirectly through the opportunity cost of inventory. Owing to the uncertainty in the number of bidders and also in the bidders' posted bids, inventory evolves stochastically and hence optimal lot-sizes should vary dynamically. The cost of holding inventory also affects optimal lot-sizes. Finally, in auctions of fashion goods, of seasonal products, of new products, and where the seller is new to the market, the seller may not know *a priori* the distributions of the number of bidders and their bids, in their entirety (Araman and Caldentey 2009, Aviv and Pazgal 2005a, Aviv and Pazgal 2005b, Ghate 2015, Farias and Roy 2010). The seller could then learn these sequentially by observing the number and values of posted bids. We emphasize here that the interaction between learning and decision-making in this problem is somewhat different from the standard multi-armed bandit problems. Here, the lot-size decision in an auction does not affect the number of bids and the values of the bids received in that auction. Instead, a higher lot-size in one auction may correspond to a fewer total number of auctions and hence fewer opportunities to learn. Thus, the learning and decision-making aspects of the problem interact across auctions but not so much within an auction. We present a Bayesian Markov decision process (MDP) (Kumar 1985) model of this simultaneous learning and optimization problem.

Pinker et al. (2010) considered multi-unit, second-price sequential auctions with a fixed, deterministic constant number of bidders in each auction, and uniformly distributed bids. They were able to derive a closed-form formula for an optimal lot-size using dynamic programming. They also presented a Bayesian framework to learn the spread of the uniform bid distributions.

Tripathi et al. (2009) also assumed a fixed, deterministic constant number of bidders in each auction and uniformly distributed bids. They worked with a multi-unit Dutch mechanism. They assumed that the lot-size did not change across auctions and derived a simple closed-form formula for an optimal lot-size. They presented a goal programming method to learn bid distributions from online auction data.

Chen et al. (2011) generalized the models in Pinker et al. and Tripathi et al. in several ways. First, they allowed for a random number of bidders in each auction and arbitrary bid distributions. They also did not restrict their formulation to any specific auction mechanism. Instead, they provided a sufficient condition (that involved the expected revenue function and the probability mass function (pmf) of the number of bidders) under which a staircase with unit jumps policy was optimal for lot-sizes. According to this policy, if it is optimal to auction x units when the inventory on hand is i, then it is optimal either to auction x units or to auction $x+1$ units when the inventory level is $i+1$. Our clairvoyant MDP model in Section 2 below is similar to the model in Chen et al. barring minor idiosyncratic differences. Our research in this paper extends that clairvoyant model to accommodate learning over a sequence of auctions.

There is a parallel body of literature on sequential auctions, where the seller does not announce lot-sizes, but instead uses reserve prices for releasing units. Examples include Vulcano et al. (2002) and Vulcano and van Ryzin (2004). These papers did not incorporate learning. Ghate (2015) recently incorporated demand learning in sequential, *single-unit*, second-price auctions where the seller optimizes her reserve price for this unit in each auction. Ghate assumed that the number of bidders in each auction is Poisson distributed with an unknown mean. The seller then uses a mixture-of-Gamma prior on this mean. This resulted in a high-dimensional Bayesian MDP, that was solved approximately by using semi-stochastic certainty equivalent control (CEC) and Q-function approximation.

Several papers in the inventory control and dynamic pricing areas have studied demand learning. Examples on the inventory control side include Scarf (1959), Iglehart (1964), Azoury (1985), Lovejoy (1990), Lariviere and Porteus (1999). For instance, Scarf generalized the classic inventory model of Arrow et al. (1951) to accommodate demand distributions from an exponential family with one unknown parameter; other papers listed above studied variations of this learning problem. Examples on the dynamic pricing side include Araman and Caldentey (2009), Aviv and Pazgal (2005b), and Farias and Van Roy (2010). For instance, Aviv and Pazgal (2005b) included demand learning in the Poisson intensity control model of Gallego and van Ryzin (1994). They assumed that the seller's prior belief about the Poisson rate parameter was Gamma, and used the CEC heuristic to approximately solve the resulting problem. Farias and Van Roy (2010) revisited the Poisson intensity control problem, where the seller's prior belief on the Poisson rate parameter was a mixture-of-Gamma distributions. They proposed a new approximation approach called load-balancing, and showed that it outperformed CEC.

We next begin with a clairvoyant MDP model similar to Chen et al. (2011), where the seller is assumed to know the demand and the bid distributions.

2 CLAIRVOYANT MDP

Consider a seller who initially holds $I \geq 1$ units in her inventory. The seller uses a sequence of online, multi-unit auctions to clear inventory and generate revenue. A cost of $h \geq 0$ is incurred per unit held in inventory over the duration of an auction and is charged at the beginning of the auction. At the end of each auction, the inventory level drops by a quantity that either equals the lot-size (if the number of bids received is strictly larger than the lot-size) or the number of bids received (if the number of bids is less than or equal to the lot-size). This process continues until all inventory is cleared. The discount factor over the duration of each auction is $0 < \delta < 1$. The seller's goal is to find a rule for lot-size decisions to maximize total discounted expected profit over all auctions.

We assume that the numbers of bidders across different auctions are independent and identically distributed (iid) random variables. Similarly, bids across different auctions are iid random variables. In addition, bids of different bidders in one auction are also iid random variables and are independent of the number of bids posted in that and in other auctions. These assumptions are standard in the aforementioned literature on sequential auctions.

The random number of bids in any auction is denoted by N. Let $p(n)$ denote the Poisson pmf of N with mean $\lambda > 0$. Moreover, for $n = 0, 1, \ldots$, we use $\bar{P}(n)$ to denote the probability that more than n bids are received.

A bid is denoted by the random variable Y, which takes values from the finite set of non-negative integers $\{0, 1, 2, \ldots, B\}$. Let $f_y \triangleq P(Y = y)$, for $y \in \{0, 1, 2, \ldots, B\}$, denote the categorical pmf of Y. We write it compactly as $\vec{f} \triangleq (f_0, f_1, \ldots, f_B)$.

Our first task is to obtain an expression for the expected revenue in a single auction with lot-size x. When $n \geq x+1$ bids are posted, the seller's expected revenue in a second-price auction with x units equals the expected value of the $x+1$st largest among n iid categorical random variables defined above. We denote this expected value of the $x+1$st largest among n iid categorical random variables by $\psi(x;n)$.

Lemma 1 The expected value of the $x+1$st largest among n iid categorical random variables is given by

$$\psi(x;n) \triangleq \sum_{y=0}^{B} \sum_{j=0}^{n-x-1} \binom{n}{j} (\eta(y))^j (1-\eta(y))^{n-j}, \tag{1}$$

where $\eta(y) = P(Y < y)$, for $y = 0, 1, \ldots, B$.

Proof. For any positive integer $n \geq 1$, and $1 \leq k \leq n$, let $Y_{(k)}(n)$ denote the kth smallest among n iid categorical random variables as defined above. Then, $E[Y_{(k)}(n)] = \sum_{y=0}^{B} P(Y_{(k)}(n) \geq y)$ by Problem 1.1 in Ross (1996). Note that $P(Y_{(k)}(n) \geq y)$, which represents the probability that the kth smallest among n iid random variables is at least y, is equal to the probability that at most $k-1$ among these random variables are strictly smaller than y. Thus, $P(Y_{(k)}(n) \geq y) = \sum_{j=0}^{k-1} \binom{n}{j} (\eta(y))^j (1-\eta(y))^{n-j}$. The result then follows by observing that the $x+1$st largest among n numbers is the $(n-x)$th smallest among these numbers. ☐

Thus, the seller's expected revenue with a Poisson distributed number of bids equals $\phi(x) \triangleq x \sum_{n=x+1}^{\infty} p(n)\psi(x;n)$. Here, we consider second-price auctions for concreteness; as in Chen et al. (2011), other variations such as first-price auctions can also be handled by our approach by using the appropriate expected order statistic formula in (1).

We now model the seller's lot-size decision problem as a stationary, discounted MDP (Puterman 1994, Ross 1983). The state of this MDP equals the inventory on hand. For such an MDP, a stationary policy is optimal. That is, it suffices to find, irrespective of the stochastic history of inventory evolution and of the auction number, the lot-size x_i in each inventory level $1 \leq i \leq I$. For $0 \leq i \leq I$, we let $V(i)$ denote the maximum total discounted expected profit earned from all future auctions starting with inventory i on hand. In dynamic programming parlance, function V is called the optimal value function.

For inventory levels $i = 0, 1, \ldots, I$, the optimal values $V(i)$ uniquely satisfy Bellman's equations given by

$$V(i) = \max_{x \in \{0,1,\ldots,i\}} \Big[-hi + \delta\phi(x) + \delta \sum_{n=0}^{x} p(n)V(i-n) + \delta\bar{P}(x)V(i-x) \Big]. \tag{2}$$

Here, the right hand side of (2) includes three terms. The first term $-hi$ is the cost of holding i units in inventory. The second term, $\delta\phi(x)$, accounts for the seller's discounted expected revenue from a single auction with x units. The third term corresponds to the optimal discounted expected profit earned through all future auctions given that the seller announced a lot-size x in the current auction but only up to x bids were posted. The forth term accounts for the optimal discounted expected profit generated through all remaining auctions given that the seller announced a lot-size x in the current auction and at least $x+1$ bids were posted. The lot-sizes x_i, for $i = 0, 1, \ldots, I$, that achieve the maxima in (2) define an optimal stationary policy (Puterman 1994, Ross 1983).

Value iteration (Puterman 1994, Ross 1983) is a standard algorithm for solving Bellman's equations and obtaining the corresponding optimal policy for MDPs. In our case, value iteration would proceed as follows. We start with $V^1(i) = 0$ for all $0 \le i \le I$, and use the update formula

$$V^{t+1}(i) = \max_{x \in \{0,1,\ldots,i\}} \left[-hi + \delta\phi(x) + \delta \sum_{n=0}^{x} p(n) V^t(i-n) + \delta \bar{P}(x) V^t(i-x) \right],$$

for $t = 1, 2, 3, \ldots$. We generate these updates until $\max_{0 \le i \le I} |V^{t+1}(i) - V^t(i)|$ drops below a tolerance. From the general theory of MDPs (Puterman 1994, Ross 1983), we know that $\lim_{t \to \infty} V^t(i) = V(i)$ for all $0 \le i \le I$. The function V^t thus provides an approximation of the optimal value function V and we obtain an approximate optimal policy by substituting V^t as a surrogate for V in the right hand side of (2) on termination.

3 BAYESIAN LEARNING WITH GAMMA AND DIRICHLET PRIORS

The goal in this paper is to study a variation of the above clairvoyant problem wherein the seller is uncertain about (i) the mean λ of the Poisson demand distribution, and (ii) the categorical pmf $\vec{f}$ of the bids. To emphasize the dependence of quantities such as $p(n)$, $\bar{P}(n)$, $\psi(x;n)$, and $\phi(x)$ from Section 2 on λ and $\vec{f}$, we denote them instead by $p_\lambda(n)$, $\bar{P}_\lambda(n)$, $\psi_{\vec{f}}(x;n)$, and $\phi_{\lambda,\vec{f}}(x)$, respectively, in the rest of this paper. We similarly write the clairvoyant value function as $V_{\lambda,\vec{f}}$.

3.1 A Gamma Prior on Mean Poisson Demand

We assume that (i) the seller views λ as a random variable, which we denote by Λ, and (ii) the seller has a Gamma prior belief about Λ. The shape and rate parameters of this prior are denoted by α, β, respectively. In particular, the seller believes *a priori* that the probability density function $r(\lambda;\alpha,\beta)$ of Λ is given by $r(\lambda;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$. A benefit of this approach is that a Gamma prior is conjugate to Poisson. That is, the seller's posterior belief about the unknown mean of the Poisson number of bidders after observing the number of bids posted in one auction is also Gamma. Specifically, if $n \ge 0$ bids are posted in an auction, then the seller's belief about Λ is updated according to $\alpha^n \leftarrow \alpha + n$ and $\beta^n \leftarrow \beta + 1$. This conjugate property and the resulting straightforward parameter update considerably simplify the Bayesian MDP formulation of the seller's decision problem.

3.2 A Dirichlet Prior on categorical Bids

We assume that (i) the seller views the probabilities $\vec{f} = (f_0, f_1, \ldots, f_B)$ as random variables, and (ii) the seller has a Dirichlet prior belief with hyperparameters $\vec{a} \triangleq (a_0, a_1, \ldots, a_B) \in \Re_{++}^{B+1}$ about these random variables. Specifically, the prior probability density function of the categorical pmf $\vec{f}$ is given by $q(\vec{f};\vec{a}) \triangleq \frac{1}{\mathscr{B}(\vec{a})} \prod_{j=0}^{B} (f_j)^{a_j - 1}$, where $\mathscr{B}(\vec{a})$ is the multinomial Beta function that is expressed in terms of Gamma functions as $\mathscr{B}(\vec{a}) \triangleq \frac{\prod_{j=0}^{B} \Gamma(a_j)}{\Gamma(\sum_{j=0}^{B} a_j)}$. Recall here that the Gamma function is defined by $\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$ (Abramowitz and Stegun 1972). Again, a benefit of this approach is that the Dirichlet prior is conjugate to categorical. That is, the seller's posterior belief about the unknown pmf of the categorical bid distribution after observing the bids posted in one auction remains Dirichlet. Specifically, if bids $\vec{y}^n \triangleq (y_1, y_2, \ldots, y_n)$ are posted in an auction, then the seller's belief about $\vec{f}$ is updated according to $a_j \leftarrow a_j + \sum_{k=1}^{n} \mathscr{I}_j(y_k),\ j = 0,1,\ldots,B$. Here, $\mathscr{I}_j(y_k)$ is the indicator function that equals one when $y_k = j$ and equals zero otherwise. We compactly write the new Dirichlet hyperparameters as $\vec{a}(\vec{y}^{\,n})$. Again, this conjugate property and the resulting straightforward parameter update considerably simplify the Bayesian MDP formulation of the seller's decision problem.

3.3 Bayesian MDP Formulation

We use $z \triangleq (\alpha,\beta) \in \Re^2_{++}$ and $\vec{a} \in \Re^{B+1}_{++}$ to denote the information states at the beginning of an auction. The seller's decision problem can now be formulated as a Bayesian MDP that is described next.

Given the information state $(z,\vec{a})$, the seller believes that her expected revenue from one auction with lot-size x is given by

$$\begin{aligned}\phi(x;z,\vec{a}) &\triangleq \int\int r(\lambda;z)q(\vec{f};\vec{a})\phi_{\lambda,\vec{f}}(x)d\lambda df_0 df_1\ldots df_B \\ &= \int\int r(\lambda;z)q(\vec{f};\vec{a})\left(\sum_{n=x+1}^{\infty} p_\lambda(n)\psi_{\vec{f}}(x;n)\right)d\lambda df_0 df_1\ldots df_B. \end{aligned} \tag{3}$$

Note that in the above double integrals, the first integral in over $\{\vec{f} \geq 0 : \sum_{j=0}^{B} f_j = 1\}$ and the second one is over $\lambda \geq 0$. Similarly, the seller believes that the probability that $n \geq 0$ bids will be posted is given by $p(n;z) \triangleq \int_{\lambda \geq 0} r(\lambda;z)p_\lambda(n)d\lambda$. Also, when the categorical pmf of bids is $\vec{f}$, let $\omega_{\vec{f}}(\vec{y}^{\,n})$ denote the probability that the posted bids will equal $\vec{y}^{\,n} = (y_1,y_2,\ldots,y_n) \in \{0,1,\ldots,B\}^n$, given that $n \geq 1$ bids are posted. Here, $\{0,1,\ldots,B\}^n$ denotes the set of all permutations of n elements from the set $\{0,1,\ldots,B\}$. Thus, if $n \geq 1$ bids are posted, the seller believes that they will equal $\vec{y}^{\,n} \in \{0,1,\ldots,B\}^n$ with probability $\omega(\vec{y}^{\,n};\vec{a}) \triangleq \int_{\{\vec{f}\geq 0:\ \sum_{j=0}^{B} f_j = 1\}} \omega_{\vec{f}}(\vec{y}^{\,n})q(\vec{f};\vec{a})df_0 df_1\ldots df_B$. We assume, as a matter of convention, that $\sum_{\vec{y}^{\,0}\in\{0,1,\ldots,B\}^0} \omega(\vec{y}^{\,0};\vec{a}) = 1$. For all $0 \leq i \leq I$, all $z \in \Re^2_{++}$, and all $\vec{a} \in \Re^{B+1}_{++}$, let the optimal value function $\mathscr{V}(i;z,\vec{a})$ denote the maximum total discounted expected profit earned from all future auctions when the inventory at the beginning of an auction is i and the information state is $z,\vec{a}$. We use z^n to denote the updated information state (α^n,β^n) as obtained from the information state $z = (\alpha,\beta)$, when $n \geq 0$ is the number of bids posted in an auction. Similarly, we use $\vec{a}(\vec{y}^{\,n})$ to denote the updated information state obtained from the information state $\vec{a}$, when bids $\vec{y}^{\,n}$ are posted in an auction. Then, Bellman's equations for the seller's Bayesian MDP are given by

$$\begin{aligned}\mathscr{V}(i;z,\vec{a}) = \max_{x\in\{0,1,\ldots,i\}} \Big\{ &-hi + \delta\phi(x;z,\vec{a}) + \delta\sum_{n=0}^{x} p(n;z) \sum_{\vec{y}^{\,n}\in\{0,1,\ldots,B\}^n} \omega(\vec{y}^{\,n};\vec{a})\mathscr{V}(i-n;z^n,\vec{a}(\vec{y}^{\,n})) \\ &+ \delta\sum_{n=x+1}^{\infty} p(n;z) \sum_{\vec{y}^{\,n}\in\{0,1,\ldots,B\}^n} \omega(\vec{y}^{\,n};\vec{a})\mathscr{V}(i-x;z^n,\vec{a}(\vec{y}^{\,n}))\Big\}. \end{aligned} \tag{4}$$

Exact solution of this MDP is computationally intractable owing to the multi-dimensional state-space. We therefore investigate three methods for its approximate solution in the next section.

4 ALGORITHMS FOR APPROXIMATE SOLUTION OF THE BAYESIAN MDP

4.1 Semi-stochastic Certainty Equivalent Control

In semi-stochastic certainty equivalent control (CEC), (some of) the random variables in a decision problem are replaced by their expected values (Bertsekas 2007). In our context, if the information state is $z,\vec{a}$, then the seller makes a lot-size decision by assuming that (i) the number of bidders is Poisson distributed with mean equal to the mean of the Gamma belief; and (ii) the bids are distributed according to a categorical. distribution whose pmf is equal to the component means of the Dirichlet belief. We denote these means by $\lambda(z)$ and $\mu_j(\vec{a})$, for $j = 0,1,\ldots,B$, respectively. Note that $\lambda(z) = \alpha/\beta$, and $\mu_j(\vec{a}) = a_j/(\sum_{j=0}^{B} a_j)$. We use the shorthand $\vec{\mu}(\vec{a}) \triangleq (\mu_0(\vec{a}),\mu_1(\vec{a}),\ldots,\mu_B(\vec{a}))$ to denote the vector of means. The inventory level and the information state then stochastically evolve to their new values and this process repeats until the end of all auctions. In particular, in state $(i;z,\vec{a})$, the lot-size is obtained by solving

$$\max_{x\in\{0,1,\ldots,i\}} -hi + \delta\phi_{\lambda(z),\vec{\mu}(\vec{a})}(x) + \delta\sum_{n=0}^{x} p_{\lambda(z)}(n)V_{\lambda(z),\vec{\mu}(\vec{a})}(i-n) + \delta\bar{P}_{\lambda(z)}(x)V_{\lambda(z),\vec{\mu}(\vec{a})}(i-x).$$

The certainty equivalent policy depends on the seller's beliefs only through their expectations $\lambda(z)$ and $\vec{\mu}(\vec{a})$. We therefore propose alternative approximation methods in the next two sections.

4.2 Knowledge Gradient Algorithm

The knowledge gradient algorithm is often employed to balance the exploration-exploitation tradeoff in sequential information collection problems (Frazier, Powell, and Dayanik 2008, Ryzhov, Powell, and Frazier 2012). Roughly speaking, the basic idea is to choose an alternative that would be optimal if the current period were the final opportunity for learning although profits would continue to accrue in future periods. In our case, this amounts to choosing lot-sizes in state $(i;z,\vec{a})$ by solving

$$\begin{aligned}\max_{x\in\{0,1,\ldots,i\}} & -hi + \delta\phi(x;z,\vec{a}) + \delta\sum_{n=0}^{x} p(n;z)\sum_{\vec{y}^{\,n}\in\{0,1,\ldots,B\}^n}\omega(\vec{y}^{\,n};\vec{a})\tilde{V}(i-n;z^n,\vec{a}(\vec{y}^{\,n})) \\ & + \delta\sum_{n=x+1}^{\infty} p(n;z)\sum_{\vec{y}^{\,n}\in\{0,1,\ldots,B\}^n}\omega(\vec{y}^{\,n};\vec{a})\tilde{V}(i-x;z^n,\vec{a}(\vec{y}^{\,n})),\end{aligned} \tag{5}$$

where

$$\tilde{V}(i;z^n,\vec{a}(\vec{y}^{\,n})) \triangleq \int\int V_{\lambda,\vec{f}}(i)r(\lambda;z^n)q(\vec{f};\vec{a}(\vec{y}^{\,n}))d\lambda df_0 df_1\ldots df_B, \tag{6}$$

for $0 \le i \le I$ and the first and second integrals are over $\{\vec{f} \ge 0 : \sum_{j=0}^{B} f_j = 1\}$ and $\lambda \ge 0$ respectively. In our implementation of this method, we estimate $\phi(x;z,\vec{a})$ via Monte Carlo simulation. Specifically, we sample K values of λ, and K values of the vector $\vec{f}$ independently according to density functions $r(\lambda;z)$ and $q(\vec{f};\vec{a})$, respectively. We denote these values by $\hat{\lambda}_k$ and $\hat{\vec{f}}(k)$, for $k = 1,2,\ldots,K$. We then estimate $\phi(x;z,\vec{a})$ using the sample average $\phi(x;z,\vec{a}) \approx \frac{\sum_{k=1}^{K}\phi_{\hat{\lambda}_k,\hat{\vec{f}}(k)}(x)}{K}$. The expectations in (5) are also estimated using sample average approximation as follows. For every $\hat{\lambda}_k$ and $\hat{\vec{f}}(k)$ we sample an n and a $\vec{y}^{\,n}$ from a Poisson and a categorical distribution, and then calculate the associated $(z^n,\vec{a}(\vec{y}^{\,n}))$. Using K values of $(z^n,\vec{a}(\vec{y}^{\,n}))$, we then use sample average to approximate the expectations in (5). Equation (6) is approximated using $\tilde{V}(i;z^n,\vec{a}(\vec{y}^{\,n})) \approx V_{\lambda(z^n),\vec{f}(\vec{a}(\vec{y}^{\,n}))}(i)$, where $\lambda(z^n)$ and $\vec{f}(\vec{a}(\vec{y}^{\,n}))$ are expectations of densities $r(\lambda;z^n)$ and $q(\vec{f};\vec{a}(\vec{y}^{\,n}))$. Finally, these approximations are employed in the sample average calculation. Then equation (5) will be approximated by

$$\max_{x\in\{0,1,\ldots,i\}} -hi + \delta\frac{\sum_{k=1}^{K}\phi_{\hat{\lambda}_k,\hat{\vec{f}}(k)}(x)}{K} + \delta\frac{\sum_{0\le n_k\le x} V_{\lambda(z^{n_k}),\vec{f}(\vec{a}(\vec{y}^{\,n_k}))}(i-n_k) + \sum_{n_k>x} V_{\lambda(z^{n_k}),\vec{f}(\vec{a}(\vec{y}^{\,n_k}))}(i-x)}{K}. \tag{7}$$

4.3 Thompson Sampling

The Thompson sampling algorithm is typically employed to balance the exploration-exploitation tradeoff in multi-armed bandit problems (Thompson 1933). It has also been extended to Bayesian MDPs (Strens 2000). The basic idea is to choose an action according to the probability that it maximizes the expected reward. In practice, this is implemented via sampling. That is, at each decision epoch, problem parameters are sampled according to the decision-maker's belief, and an action that maximizes the expected reward (given these sampled parameter values) is chosen. In our context, this reduces to the following approach. When the inventory is i and the information state is $(z,\vec{a})$, the seller samples a λ according to the Gamma density $r(\lambda;z)$ and the bid pmf $\vec{f}$ according to the Dirichlet density $q(\vec{f};\vec{a})$. Let $\hat{\lambda}$ and $\hat{\vec{f}}$ denote these sampled values. The seller then chooses a lot-size that maximizes the profit assuming that the demand and bid distribution parameters are $\hat{\lambda}$ and $\hat{\vec{f}}$, respectively.

The computational effort required for the above three approximation methods is roughly as follows. For finding the optimal lot-size in every state, the Bellman equation (2) needs to be solved using value iteration when CEC and Thompson sampling approximations are applied. For knowledge gradient, the Bellman equation (2) needs to be solved K times in every state in contrast. This causes knowledge gradient to be slower than CEC and Thompson sampling. This behavior was also observed in our numerical results in the next section.

5 COMPUTATIONAL RESULTS

We created different problem instances by changing the true λ (columns of our tables), and initial inventory I (rows). The discount factor was $\delta = 0.99$. The sample size K was set to 50 in all sample average approximations. Parameters of the initial Gamma prior were fixed at $\alpha = 5$ and $\beta = 1$. The initial Dirichlet parameters were set to $\vec{a} = (1,1,\ldots,1)$. The maximum bid value B was set to 430 and the holding cost was $h = 10$. The true categorical bid distribution was set to equal a discretized, truncated Weibull distribution. The scale parameter of this Weibull distribution was fixed at $B/2$. Its shape parameter was 2 for Tables 1-4, and 4 for Tables 5-7. We refer to these two shapes as wide and narrow bid distributions, respectively. In each row of each table, we report averages over 50 independent simulations. Each simulation terminated when the inventory was completely depleted. Semi-stochastic CEC, knowledge gradient, and Thompson sampling need, as subroutines, value functions of MDPs where the mean demand and the bid pmf are fixed at various values. These were approximated by the value function of the staircase with unit jumps policy to significantly speed up our simulations. In all our simulations, the clairvoyant optimal policy is a hypothetical, ideal policy that is assumed to know the true λ and the true categorical bid distribution.

The numbers in Table 1 equals the profit made by a lot-sizing policy that does not update the seller's beliefs, reported as a percentage of the clairvoyant optimal profit. This "no learning" policy finds lot-sizes by solving Bellman's equations where the expected revenue and the demand pmf are calculated by taking expectations with respect to the initial Gamma and Dirichlet priors. In our approximate implementation of this process, these expectations were estimated as sample averages. The boldface numbers in Table 1 indicate that the profit made by the no learning policy was statistically different from the clairvoyant optimal profit (as inferred from a t-test at the significance threshold of 0.05).

The first number in each column of Tables 2, 3, and 4 equals the profits made by semi-stochastic CEC, knowledge gradient, and Thompson sampling, reported as a percentage of the clairvoyant optimal profit, respectively. The boldface numbers in these tables indicate that the profits made by these three learning methods were statistically different from those made by the no learning policy (again, as inferred from a t-test at the significance threshold of 0.05).

The second number in each column of Tables 2, 3, and 4 refers to the increase or decrease in profit compared to the no learning policy. This is reported as 0.00 when the learning and no learning policies are not statistically different; otherwise, it is reported by subtracting from the first number either the first number in Table 1 or 100 depending on whether or not the profit without learning is statistically different form the clairvoyant optimal profit.

Table 1: The percentage of optimal profit reached with no learning for the wide bid distribution.

	λ			
I	5	10	15	20
20	104.31	**95.71**	**96.49**	**94.64**
25	99.09	**96.71**	**95.79**	**93.16**
30	96.08	**96.29**	**92.24**	**92.61**
35	96.38	**94.42**	**92.97**	**91.89**
40	96.99	**95.21**	**90.07**	**91.75**
45	97.60	**91.60**	**91.43**	**90.30**
50	91.54	**93.42**	**89.64**	**89.28**
55	89.52	**90.86**	**90.81**	**87.21**
60	112.86	**92.37**	**89.89**	**86.10**

Table 2: The percentage of optimal profit reached by semi-stochastic CEC for the wide bid distribution.

	λ			
I	5	10	15	20
20	**102.41**,(2.41)	**98.15**,(2.44)	**98.17**,(1.68)	**98.39**,(3.75)
25	**96.68**,(-3.32)	**98.06**,(1.35)	**99.28**,(3.49)	**98.40**,(5.24)
30	95.92,(0.00)	**98.24** ,(1.95)	**98.03**,(5.79)	**97.77**,(5.16)
35	95.07,(0.00)	**96.28**,(1.86)	**97.33**,(4.36)	**97.28**,(5.39)
40	**92.92**,(-7.08)	**97.20**,(1.99)	**96.08**,(6.01)	**97.12**,(5.37)
45	**94.18**,(-5.82)	**94.76**,(3.16)	**96.62**,(5.19)	**96.99**,(6.69)
50	**86.52**,(-13.48)	**95.18**,(1.76)	**96.31**,(6.67)	**96.73**,(7.45)
55	90.74,(0.00)	**93.59**,(2.73)	**96.93**,(6.12)	**96.06**,(8.85)
60	**91.69**,(-8.31)	93.49,(0.00)	**95.24**,(5.35)	**95.66**,(9.56)

Table 3: The percentage of optimal profit reached by knowledge gradient for the wide bid distribution.

	λ			
I	5	10	15	20
20	**102.04**,(2.04)	**98.03**,(2.32)	**98.08**,(1.59)	**98.15**,(3.51)
25	**97.44**,(-2.56)	97.74,(0.00)	**99.50**,(3.71)	**98.04**,(4.88)
30	**95.36**,(-4.64)	**98.17**,(1.88)	**97.90**,(5.66)	**97.64**,(5.03)
35	94.56,(0.00)	**96.05**,(1.63)	**97.92**,(4.95)	**97.79**,(5.9)
40	**93.15**,(-6.85)	**96.74**,(1.53)	**96.01**,(5.94)	**96.97**,(5.22)
45	**94.01**,(-5.99)	**94.68**,(3.08)	**96.47**,(5.04)	**96.88**,(6.58)
50	**85.45**,(-14.55)	**94.71**,(1.29)	**96.06**,(6.42)	**96.39**,(7.11)
55	90.21,(0.00)	**93.56**,(2.7)	**96.76**,(5.95)	**96.01**,(8.8)
60	**93.65**,(-6.35)	**93.97**,(1.6)	**95.39**,(5.5)	**95.87**,(9.77)

Table 4: The percentage of optimal profit reached by Thompson sampling for the wide bid distribution.

	λ			
I	5	10	15	20
20	**100.95**,(0.95)	**97.82**,(2.11)	**98.22**,(1.73)	**98.20**,(3.56)
25	97.37,(0.00)	**97.89**,(1.18)	**99.14**,(3.35)	**97.85**,(4.69)
30	96.08,(0.00)	**98.26**,(1.97)	**97.35**,(5.11)	**97.88**,(5.27)
35	95.41,(0.00)	**97.17**,(2.75)	**97.11**,(4.14)	**97.68**,(5.79)
40	94.11,(0.00)	**96.74**,(1.53)	**96.29**,(6.22)	**96.01**,(4.26)
45	94.45,(0.00)	**94.50**,(2.9)	**96.66**,(5.23)	**95.94**,(5.64)
50	**87.32**,(-12.68)	**95.28**,(1.86)	**96,.55**,(6.91)	**96.82**,(7.54)
55	85.16,(0.00)	**94.09**,(3.23)	**96.43**,(5.62)	**96.68**,(9.47)
60	**94.59**,(-5.41)	**95.48**,(3.11)	**95.49**,(5.6)	**96.15**,(10.05)

Recall from the first column of Table 1 that the profit made by the no learning policy is not significantly different from the clairvoyant optimal policy. This is intuitive because the mean α/β of the initial Gamma prior matches the true value of $\lambda = 5$. The profits in the first columns of Tables 2-4 are much more "noisy" — sometimes statistically worse than the no learning profit and sometimes better. This is perhaps to be expected because the estimates of the mean demand as computed by the learning algorithms can oscillate around the true λ depending on the observed stochastic demands.

The last three columns from Table 1 suggest that the profit of the no learning policy is statistically worse than the clairvoyant optimal profit. This is intuitive since the true λ does not match the initial Gamma prior mean. In each of these last three columns of Table 1 the no learning policy roughly seems to make decreasing relative profits as the initial inventory I increases. This is perhaps because the temporal accumulation of error induced by the mean mismatch increases with increasing initial inventory. This error accumulation phenomenon is also observed in each of the last three columns of Tables 2, 3, and 4, where the percentage (first number in each column) of the optimal clairvoyant profit reached by the learning algorithms seems to decrease with increasing inventory. The last three columns in Tables 2, 3, and 4 show that the learning algorithms are able to improve (third number in every cell) upon the profit of the no learning policy. The range of this improvement seems to increase as the true λ drifts away from the mean of the initial Gamma prior. This value of active learning as a function of the mean mismatch also seems to increase with increasing inventory. Tables 2-4 suggest that there is no significant difference among the profits made by the three learning algorithms.

In Tables 1-4, the true bid distribution was similar to a centered Weibull. We also tried right- and left-skewed Weibull bids (not reported in this paper for brevity), and were able to draw qualitatively similar conclusions.

Tables 5-7 report results of simulations similar to the previous tables, but this time with the narrow bid distribution. For this narrow bid distribution, we only tried semi-stochastic CEC and Thompson sampling. This decision was made based on our observation (from simulations and tables for the wide bid distribution) that the knowledge gradient algorithm was computationally slower but did not provide a statistically significant improvement in profit.

The profit reached by the no learning policy is generally smaller in Table 5 than in Table 1, for each column. This is because the narrow bid distribution is more different from the flat Dirichlet prior that in the wide one. Even for the first column of Table 5, although the true λ matches the Gamma prior mean, the profit of the no learning policy can be statistically worse than the clairvoyant optimal profit. The learning algorithms are still noisy for the same reason as in Table 1. Consistent with these points, the value of active learning over no learning in the last three columns of Tables 6-7 seems to be generally larger than the corresponding tables for the wide bid distribution. Finally, other qualitative observations from the tables for the wide bid distribution also hold for the narrow bid distribution.

Table 5: The percentage of optimal profit reached with no learning for the narrow bid distribution.

	λ			
I	5	10	15	20
20	**90.66**	**88.88**	**86.64**	**88.83**
25	**93.13**	**88.65**	**85.47**	**83.90**
30	**90.93**	**87.10**	**80.76**	**83.17**
35	98.13	**85.82**	**82.94**	**81.09**
40	**90.9**	**85.59**	**79.15**	**80.34**
45	94.57	**85.13**	**77.71**	**78.37**
50	**86.09**	**84.38**	**77.24**	**75.04**
55	94.57	**82.82**	**74.86**	**75.80**
60	91.45	**84.52**	**73.95**	**74.87**

Table 6: The percentage of optimal profit reached by semi-stochastic CEC for the narrow bid distribution.

	λ			
I	5	10	15	20
20	90.52,(0.00)	**93.58**,(4.69)	**93.44**,(6.8)	**94.29**,(5.45)
25	**92.02**,(-1.11)	**91.74**,(3.09)	**92.29**,(6.82)	**92.96**,(9.06)
30	90.21,(0.00)	**90.68**,(3.58)	**91.35**,(10.59)	**92.48**,(9.30)
35	**94.8**,(-5.2)	**90.70**,(4.88)	**90.92**,(7.98)	**91.45**,(10.36)
40	**88.35**,(-2.55)	**91.57**,(5.97)	**89.85**,(10.7)	**91.1**,(10.8)
45	**91.22**,(-8.78)	**90.14**,(5.01)	**89.12**,(11.41)	**90.75**,(12.38)
50	**80.42**,(-5.67)	**88.56**,(4.18)	**89.40**,(12.16)	**88.98**,(13.94)
55	94.64,(0.00)	**85.93**,(3.10)	**87.76**,(12.9)	**88.98**,(13.18)
60	**83.12**,(-16.88)	**86.91**,(2.39)	**87.55**,(13.6)	**87.71**,(12.84)

Table 7: The percentage of optimal profit reached by Thompson sampling for the narrow bid distribution.

	λ			
I	5	10	15	20
20	90.78,(0.00)	**93.89**,(5)	**93.67**,(7.03)	**94.31**,(5.48)
25	92.38,(0.00)	91.61,(2.96)	**92.56**,(7.09)	**92.91**,(9.01)
30	90.31,(0.00)	**90**,(2.9)	**91.24**,(10.48)	**92.89**,(9.72)
35	**94.93**,(-5.07)	**90.73**,(4.91)	**91.81**,(8.87)	**92.37**,(11.28)
40	**87.10**,(-3.8)	**90.19**,(4.59)	**90.25**,(11.1)	**89.76**,(9.41)
45	92.62,(0.00)	**89.95**,(4.82)	**89.64**,(11.93)	**90.58**,(12.21)
50	**81.09**,(-5)	**89.11**,(4.73)	**89.63**,(12.39)	**89.38**,(14.34)
55	93.51,(0.00)	**86.64**,(3.81)	**87.46**,(12.6)	**88.49**,(12.69)
60	**83.29**,(-16.71)	**87.32**,(2.8)	**88.45**,(14.5)	**87.36**,(12.49)

In summary, we conclude that active learning may provide higher profits as compared to not learning, especially when the initial priors are off. Among the three learning algorithms we implemented, semi-stochastic CEC and Thompson sampling were computationally faster and yet produced profits comparable to knowledge gradient.

It might be interesting in the future to apply other general approaches to handle parameteric uncertainties in MDPs to our auction model. Examples include the robust optimization approach from (Iyengar 2005, Nilim and El Ghaoui 2005), the less conservative approach from (Delage and Mannor 2010), and the data-driven method from (Jiang and Shanbhag 2015).

ACKNOWLEDGMENT

This research was funded in part by the National Science Foundation through grant CMMI 1536717.

REFERENCES

Abramowitz, M., and I. A. Stegun. 1972. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York, New York, USA: Dover.

Araman, V. F., and R. Caldentey. 2009. "Dynamic Pricing for Nonperishable Products with Demand Learning". *Operations Research* 57 (5): 1169–1188.

Arrow, K. J., T. E. Harris, and J. Marschak. 1951. "Optimal Inventory Policy". *Econometrica* 19 (3): 250–272.

Aviv, Y., and A. Pazgal. 2005a. "A Partially Observed Markov Decision Process for Dynamic Pricing". *Management Science* 51 (9): 1400–1416.

Aviv, Y., and A. Pazgal. 2005b. "Dynamic Pricing of Short Life-Cycle Products Through Active Learning". *Working paper available at http://apps.olin.wustl.edu/faculty/aviv/papers/Pricing_MS_Dec_2005.pdf*.

Azoury, K. S. 1985. "Bayes Solution to Dynamic Inventory Models under Unknown Demand Distribution". *Management Science* 31 (9): 1150–1160.

Bertsekas, D. P. 2007. *Dynamic Programming and Optimal Control*. third ed, Volume 1 and 2. Nashua, New Hampshire, USA: Athena Scientific.

Chen, X., A. Ghate, and A. K. Tripathi. 2011. "Dynamic Lot-Sizing in Sequential Online Retail Auctions". *European Journal of Operational Research* 215 (1): 257–267.

Delage, E., and S. Mannor. 2010. "Percentile Optimization for Markov Decision Processes with Parameter Uncertainty". *Operations Research* 58 (1): 203–213.

Farias, V. F., and B. V. Roy. 2010. "Dynamic Pricing with a Prior on Market Response". *Operations Research* 58 (1): 16–29.

Frazier, P. I., W. B. Powell, and S. Dayanik. 2008. "A Knowledge-Gradient Policy for Sequential Information Collection". *SIAM Journal on on Control and Optimization* 47 (5): 2410–2439.

Gallego, G., and G. van Ryzin. 1994. "Optimal Dynamic Pricing of Inventories with Stochastic Demand over Finite Horizons". *Management Science* 40 (8): 999–1020.

Ghate, A. 2015. "Optimal Minimum Bids and Inventory Scrapping in Sequential, Single-Unit, Vickrey Auctions with Demand Learning". *European Journal of Operational Research* 245 (2): 555–570.

Iglehart, D. L. 1964. "The Dynamic Inventory Problem with Unknown Demand Distribution". *Management Science* 10 (3): 429–440.

Iyengar, G. N. 2005. "Robust Dynamic Programming". *Mathematics of Operations Research* 30 (2): 257–280.

Jiang, H., and U. V. Shanbhag. 2015. "Data-Driven Schemes for Resolving Misspecified MDPs: Asymptotics and Error Analysis". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 3801–3812. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Kumar, P. R. 1985. "A Survey of Some Results in Stochastic Adaptive Control". *SIAM Journal on Control and Optimization* 23 (3): 329–380.

Lariviere, M. A., and E. L. Porteus. 1999. "Stalking Information: Bayesian Inventory Management with Unobserved Lost Sales". *Management Science* 4 (3): 346–363.

Lovejoy, W. S. 1990. "Myopic Policies for Some Inventory Models with Uncertain Demand Distributions". *Management Science* 36 (6): 724–738.
Nilim, A., and L. El Ghaoui. 2005. "Robust Control of Markov Decision Processes with Uncertain Transition Matrices". *Operations Research* 53 (5): 780–798.
Pinker, E., A. Seidmann, and Y. Vakrat. 2003. "Managing Online Auctions: Current Business and Research Issues". *Management Science* 49 (11): 1457–1484.
Pinker, E., A. Seidmann, and Y. Vakrat. 2010. "Using Bid Data for the Management of Sequential, Multi-Unit, Online Auctions with Uniformly Distributed Bidder Valuations". *European Journal of Operational Research* 202:574–583.
Puterman, M. 1994. *Markov Decision Processes*. New Jersey: John Wiley and Sons.
Ross, S. M. 1983. *Introduction to Stochastic Dynamic Programming*. New York, New York, USA: Academic Press.
Ross, S. M. 1996. *Stochastic Processes*. New York, New York, USA: Wiley.
Ryzhov, I. O., W. B. Powell, and P. I. Frazier. 2012. "The Knowledge Gradient Algorithm for a General Class of Online Learning Problems". *Operations Research* 60 (1): 180–195.
Scarf, H. 1959. "Bayes Solutions of the Statistical Inventory Problems". *Annals of Mathematical Statistics* 30 (2): 490–508.
Strens, M. 2000. "A Bayesian Framework for Reinforcement Learning". In *Proceedings of the Seventeenth International Conference on Machine Learning*. San Francisco, California, USA.
Thompson, W. R. 1933. "On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples". *Biometrika* 25 (3-4): 285–294.
Tripathi, A. K., S. K. Nair, and G. Karuga. 2009. "Optimal Lot Sizing Policies for Sequential Online Auctions". *IEEE Transactions of Knowledge and Data Engineering* 21 (4): 554–567.
van Ryzin, G., and G. Vulcano. 2004. "Optimal Auctioning and Ordering in an Infinite Horizon Inventory-Pricing System". *Operations Research* 52 (3): 347–367.
Vulcano, G., G. van Ryzin, and C. Maglaras. 2002. "Optimal Dynamic Auctions for Revenue Management". *Management Science* 48 (11): 1388–1407.

AUTHOR BIOGRAPHIES

ARCHIS GHATE is an Associate Professor in the Department of Industrial and Systems Engineering at the University of Washington, Seattle. He holds a PhD in Industrial and Operations Engineering from University of Michigan. His research interests lie in stochastic and dynamic optimization. His email address is archis@uw.edu.

MAHSHID SALEMI PARIZI is a PhD student in Industrial and Systems Engineering at the University of Washington, Seattle. She completed her undergraduate education in Industrial Engineering at Sharif University of Technology. Her research interest is in approximate dynamic programming. Her email address is msalemip@uw.edu.