

LEARNING LOGISTIC DEMAND CURVES IN BUSINESS-TO-BUSINESS PRICING

Huashuai Qu

Department of Mathematics
University of Maryland
College Park, MD 20742, USA

Ilya O. Ryzhov
Michael C. Fu

Robert H. Smith School of Business
University of Maryland
College Park, MD 20742, USA

ABSTRACT

This work proposes an approximate Bayesian statistical model for predicting the win/loss probability for a given price in business-to-business (B2B) pricing. This model allows us to learn parameters in logistic regression based on binary (win/loss) data and can be quickly updated after each new win/loss observation. We also consider an approach for recommending target prices based on the approximate Bayesian model, thus integrating uncertainty into decision-making. We test the statistical model and the target price recommendation strategy with synthetic data, and observe encouraging empirical results.

1 INTRODUCTION

The problem of business-to-business (B2B) pricing arises in high-volume commercial transactions between businesses. For example, consider the problem faced by a supplier of raw materials negotiating a long-term contract with a large manufacturing concern. After a period of negotiation, the seller quotes a price, which can be accepted or rejected. If the pricing offer is rejected, the seller loses a substantial amount of revenue, but it may not be clear exactly how much lower the offer should have been. If the offer is accepted, the seller makes a profit, but is left wondering whether a somewhat higher offer would still have been accepted. The seller's goal is to maximize total revenue from a sequence of contracts, in the face of uncertainty about buyer behavior.

Dynamic pricing in general is subject to uncertainty. Classic models in revenue management often assume stochastic demand for a product (Gallego and Van Ryzin 1994, Feng and Gallego 1995), or uncertain customer valuations of it (Lazear 1986). Recent work, however, has considered the additional dimension that the uncertainty may be *environmental*, that is, the seller does not even know the distribution from which customer valuations are drawn. In practice, this distribution must be estimated, and the estimate must be adjusted over time as new transactions are observed. This gives rise to the problem of "learning and earning," in which the seller does not always prefer the decision that appears to be optimal based on the current demand model (referred to as the "myopic" decision), but rather may engage in more exploratory or experimental behavior. For example, an online retailer may increase or decrease some prices for a period of time, simply to observe the effect on sales. Although this behavior may result in lost revenue, it provides new information that produces a more accurate demand model, enabling better pricing decisions in the future.

The literature has used Bayesian statistics to model environmental uncertainty (Cope 2007, Farias and Van Roy 2010), and different pricing strategies have been proposed to optimize the balance between revenue and information. For example, Carvalho and Puterman (2005) proposes a one-step look-ahead strategy for problems with logistic revenue curves, while Chhabra and Das (2011) presents an approach based on multi-armed bandit theory. A recent stream of work, represented by Besbes and Zeevi (2009), den Boer and Zwart (2011), Harrison et al. (2012), and Broder and Rusmevichientong (2012), has focused

on establishing long-run convergence rates for policies that are mostly myopic, with occasional periods of exploration spaced increasingly further apart. However, in the specific context of B2B pricing, individual transactions typically have high volume (for example, the seller may be negotiating the price of a year’s supply of raw materials) and incur high costs (e.g. the time and money spent during negotiations), making it important to obtain good performance quickly.

We consider an application where information arrives in the form of binary win/loss observations, representing customers’ yes/no responses to the seller’s pricing offers (or “bids”). A common demand model in this setting (used e.g. by Carvalho and Puterman 2005) assumes that these binary outcomes follow a logistic distribution, which also allows us to relate the win probability to a set of regression features representing additional information about the product or customer type. Although this is a fairly natural choice of demand model (essentially just an instance of logistic regression), it is quite challenging to connect to the Bayesian way of representing new information and using it to update the seller’s beliefs. While, for linear regression, the multivariate normal distribution offers an intuitive and easy-to-use conjugate prior (Minka 2000), no such model is available for logistic regression, making it difficult to represent a belief over a continuous space of logistic curves.

We approach this problem with approximate Bayesian inference, using the technique of density projection to create a multivariate normal posterior distribution that is “approximately conjugate,” in the sense of minimizing the Kullback-Leibler divergence from the actual posterior. See e.g. Qu et al. (2012) for an application of this technique to the problem of learning unknown correlation structures in ranking and selection. In the context of logistic regression, our approach is similar to the variational approximation by Jaakkola and Jordan (2000), but involves an additional optimization step using infinitesimal perturbation analysis (see e.g. Fu 2006 or Fu 2008) to further improve the quality of the approximation. Using this statistical technique to efficiently update a multivariate normal prior on the parameters of the logistic demand curve, we then apply a policy that optimizes a myopic estimate of the expected revenue curve (see e.g. Ch. 11 of Powell and Ryzhov 2012). Our numerical experiments provide evidence in favor of both the approximate Bayesian learning model and the Bayes-greedy pricing policy. Although our Bayesian model has numerous applications outside pricing, in this particular context it enables the seller to compactly model a set of beliefs about win probabilities for a wide range of customer and product segments, and then quickly update this belief in real time.

The paper is organized as follows. Section 2 develops a scheme for Bayesian learning in logistic regression using the framework of approximate Bayesian inference. Section 3 describes the Bayes-greedy pricing policy that works together with this learning model, as well as some related structural properties. Section 4 presents simulation results, and Section 5 concludes.

2 PROBLEM FORMULATION AND LEARNING MODEL

Section 2.1 introduces the demand and revenue curves optimized by a seller in the B2B pricing problem. In Section 2.2, we discuss the challenge of developing a Bayesian model for learning the parameters of the demand curve. Then, Sections 2.3 and 2.4 outline our proposed approach for overcoming this challenge.

2.1 Problem Formulation

Consider a seller who must quote prices for a sequence of corporate clients. The $(n + 1)$ st client will accept a price offer $p^n \geq 0$ with probability ρ , which may also depend on additional properties of the client or product. The function ρ is called the *demand curve*, and is not known exactly to the seller. However, the seller does observe the client’s response, modeled as a binary variable Y^{n+1} , where $Y^{n+1} = 1$ with probability ρ , representing a sale (or “win”), and $Y^{n+1} = 0$ represents a “loss.” The seller’s expected revenue from the client is

$$R(p^n) = p^n \rho, \quad p^n \geq 0, \quad (1)$$

where the demand curve ρ usually depends on the price p^n . In most applications, we need to consider the marginal cost c for the product, and work with the expected profit

$$\Pi(p^n) = (p^n - c)\rho, \quad p^n \geq c. \quad (2)$$

We assume that Y^n follows a logistic distribution, allowing us to write the demand curve as

$$\rho(\mathbf{x}^n) = \mathbb{P}(Y^{n+1} = 1) = \frac{1}{1 + e^{-\boldsymbol{\mu}^\top \mathbf{x}^n}}, \quad (3)$$

where \mathbf{x}^n is a vector of features, observed by the seller, providing relevant information for the $(n + 1)$ st pricing decision. In the simplest possible model, the customers are assumed to be homogeneous, $\mathbf{x}^n = [1, p^n]^\top$, and the parameter vector $\boldsymbol{\mu}$ consists only of an intercept and a slope term. We use this simple model in our examples throughout this paper. However, our analysis is readily applicable to the general case, where \mathbf{x}^n may also contain information about the product (type or volume) and the client (region, industry, history with the seller).

In all of these cases, the parameter vector $\boldsymbol{\mu}$ is unknown to the seller and must be inferred using a combination of prior knowledge and incoming win/loss results. The shape of the demand curve is extremely sensitive to the parameter values, making it important to obtain accurate estimates of the parameters as quickly as possible. We now propose a Bayesian framework for learning the demand curve.

2.2 Bayesian Model For Dynamic Pricing

We adopt the Bayesian view, and represent our uncertainty about the vector $\boldsymbol{\mu}$ using a multivariate normal prior distribution, that is,

$$\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{\Sigma}). \quad (4)$$

The multivariate normal distribution offers a compact and powerful way to model correlations between our beliefs about different components of $\boldsymbol{\mu}$. Because the observation Y^{n+1} provides information about an entire vector \mathbf{x}^n , our beliefs about different components of this vector should become correlated due to their dependence on the same observations. A second convenience of the multivariate normal distribution (important for computational purposes) is that the linear combination $\boldsymbol{\mu}^\top \mathbf{x}^n$ follows a univariate normal distribution.

In linear regression, where a continuous response variable is related to a linear combination of features, the multivariate normal prior possesses the property of *conjugacy*. That is, if the residual errors are i.i.d. normal, the posterior distribution of the regression parameters, conditional on a sequence of observations, will remain normal (Minka 2000). This model makes the learning process highly efficient computationally, as one only needs to recursively update the mean vector and covariance matrix of the belief distribution after each observation. Unfortunately, in logistic regression, there is no known prior distribution that is conjugate with logistic observations. To see this, we first assume that $\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\theta}^n, \boldsymbol{\Sigma}^n)$, and write the likelihood function of Y^{n+1} as

$$P(Y^{n+1}) = g(H^{n+1}(\boldsymbol{\mu})), \quad (5)$$

where $g(z) = (1 + e^{-z})^{-1}$ and $H^{n+1}(\boldsymbol{\mu}) = (2Y^{n+1} - 1)(\boldsymbol{\mu}^\top \mathbf{x}^n)$. Equation (5) allows us to represent the win/loss probability in a concise form. Applying Bayes' rule, the posterior distribution, given the bidding price p^n and the observation Y^{n+1} , can be written as

$$p(\boldsymbol{\mu} | p^n, Y^{n+1}) \propto g(H^{n+1}(\boldsymbol{\mu})) |\boldsymbol{\Sigma}^n|^{-1/2} \exp \left\{ -\frac{1}{2} (\boldsymbol{\mu} - \boldsymbol{\theta}^n)^\top (\boldsymbol{\Sigma}^n)^{-1} (\boldsymbol{\mu} - \boldsymbol{\theta}^n) \right\}, \quad (6)$$

which is clearly non-normal.

We would like to retain the multivariate normal prior due to its power in modeling correlated beliefs. However, we are now required to use the techniques of approximate Bayesian inference to develop a

multivariate normal posterior that is ‘‘approximately conjugate.’’ Several such approaches have been proposed, including approximation methods based on Laplace approximation (Spiegelhalter and Lauritzen 1990) and variational bounds (Jaakkola and Jordan 2000). We take a variational Bayesian approach to approximate the posterior distribution by minimizing the Kullback-Leibler divergence between the true posterior distribution and a multivariate normal distribution.

2.3 Variational Bayesian Approximation

Suppose that, after observing n responses, our beliefs about $\boldsymbol{\mu}$ are multivariate normal with parameters $(\boldsymbol{\theta}^n, \boldsymbol{\Sigma}^n)$. Let $P(\boldsymbol{\mu}|p^n, Y^n, \boldsymbol{\theta}^n, \boldsymbol{\Sigma}^n)$ be the likelihood function of this distribution. The variational Bayesian approach approximates the posterior distribution of $\boldsymbol{\mu}$, given Y^{n+1} , with a normal distribution $Q(\boldsymbol{\mu}|\boldsymbol{\theta}^{n+1}, \boldsymbol{\Sigma}^{n+1})$ by minimizing the Kullback-Leibler (KL) divergence. The KL divergence between $P(\boldsymbol{\mu}|p^n, Y^{n+1}, \boldsymbol{\theta}^n, \boldsymbol{\Sigma}^n)$ and $Q(\boldsymbol{\mu}|\boldsymbol{\theta}^{n+1}, \boldsymbol{\Sigma}^{n+1})$ is defined as

$$\mathcal{D}(Q \parallel P) := \mathbb{E}_Q \left(\log \frac{Q(\boldsymbol{\mu}|\boldsymbol{\theta}^{n+1}, \boldsymbol{\Sigma}^{n+1})}{P(\boldsymbol{\mu}|p^n, Y^{n+1}, \boldsymbol{\theta}^n, \boldsymbol{\Sigma}^n)} \right), \quad (7)$$

where the expectation is taken with respect to Q . This definition can be partially simplified, as stated in the following result.

Proposition 1 The KL divergence can be written as

$$\mathcal{D}(Q \parallel P) = \mathbb{E}_Q \left[\log \left(1 + e^{-H^{n+1}(\boldsymbol{\mu})} \right) \right] + h(\boldsymbol{\theta}^n, \boldsymbol{\Sigma}^n, \boldsymbol{\theta}^{n+1}, \boldsymbol{\Sigma}^{n+1}), \quad (8)$$

with the second component specified as

$$h(\boldsymbol{\theta}^n, \boldsymbol{\Sigma}^n, \boldsymbol{\theta}^{n+1}, \boldsymbol{\Sigma}^{n+1}) = \frac{1}{2} \left[\text{tr}((\boldsymbol{\Sigma}^n)^{-1} \boldsymbol{\Sigma}^{n+1}) + (\boldsymbol{\theta}^n - \boldsymbol{\theta}^{n+1})^\top (\boldsymbol{\Sigma}^n)^{-1} (\boldsymbol{\theta}^n - \boldsymbol{\theta}^{n+1}) - k - \ln \frac{|\boldsymbol{\Sigma}^{n+1}|}{|\boldsymbol{\Sigma}^n|} + C \right],$$

where C is a constant that does not depend on $\boldsymbol{\theta}^{n+1}$ and $\boldsymbol{\Sigma}^{n+1}$.

To minimize the KL divergence, the first step is to take the gradient of $\mathcal{D}(Q \parallel P)$ with respect to its parameter $\boldsymbol{\theta}^{n+1}$ and $\boldsymbol{\Sigma}^{n+1}$. Unfortunately, a closed-form expression for the gradient is not available, because the expectation in equation (8) is intractable. However, if our goal is to minimize an expected value, a connection to gradient-based stochastic search (Kim 2006) comes naturally to mind. The work by Blei et al. (2012) uses such an approach, where a likelihood ratio estimate (Spall 2005) of the gradient is constructed. However, this approach leads to a noisy simulation optimization problem, whose dimensionality is quadratic in the number of features, presenting substantial computational difficulties.

Instead of optimizing with respect to $(\boldsymbol{\theta}^{n+1}, \boldsymbol{\Sigma}^{n+1})$, we utilize a dimension reduction technique and propose the following form for $\boldsymbol{\theta}^{n+1}$ and $\boldsymbol{\Sigma}^{n+1}$:

$$\boldsymbol{\theta}^{n+1} = \boldsymbol{\Sigma}^{n+1} \left((\boldsymbol{\Sigma}^n)^{-1} \boldsymbol{\theta}^n + \left(Y^{n+1} - \frac{1}{2} \right) \mathbf{x}^n \right) \quad (9)$$

$$\boldsymbol{\Sigma}^{n+1} = \left((\boldsymbol{\Sigma}^n)^{-1} + \lambda \mathbf{x}^n (\mathbf{x}^n)^\top \right)^{-1} \quad (10)$$

Applying the Sherman-Morrison formula to (9) and (10), we obtain

$$\boldsymbol{\theta}^{n+1} = \boldsymbol{\theta}^n + \frac{Y^{n+1} - 1/2 - (\mathbf{x}^n)^\top \boldsymbol{\theta}^n}{\frac{1}{\lambda} + (\mathbf{x}^n)^\top \boldsymbol{\Sigma}^n \mathbf{x}^n} \boldsymbol{\Sigma}^n \mathbf{x}^n, \quad (11)$$

$$\boldsymbol{\Sigma}^{n+1} = \boldsymbol{\Sigma}^n - \frac{\boldsymbol{\Sigma}^n \mathbf{x}^n (\mathbf{x}^n)^\top \boldsymbol{\Sigma}^n}{\frac{1}{\lambda} + (\mathbf{x}^n)^\top \boldsymbol{\Sigma}^n \mathbf{x}^n}. \quad (12)$$

In this form, there is only one parameter λ to be determined. We minimize the KL divergence with respect to λ to find the optimal multivariate normal posterior distribution from the parametrized family in (11)-(12). Aside from the computational convenience of reducing the size of the problem, we choose precisely this form because it resembles the Kalman-filter-like equations used for Bayesian linear regression; the parameter λ is analogous to the precision of the residuals, while $\frac{Y^{n+1}-1/2}{\lambda}$ stands in for the continuous observation. In this way, our learning model for logistic regression makes an intuitive connection to the well-understood linear setting. Moreover, previous work on logistic regression, including Jaakkola and Jordan (2000) and Spiegelhalter and Lauritzen (1990), has derived updating rules with very similar form, based on different approximation techniques for the posterior likelihood function.

For additional convenience, we apply the transformation $v = \frac{1}{\lambda}$ and find

$$v^* = \underset{v}{\operatorname{argmin}} \mathcal{D}(Q \parallel P). \quad (13)$$

The parameter v is analogous to the variance of the residuals in a linear regression model. Since no such explicit parameter is given in logistic regression, we simply find the value that produces the most accurate approximation.

2.4 Minimizing the Kullback-Leibler Divergence

We now propose a stochastic approximation method to solve the minimization problem in (13), which requires estimations of the gradient of $\mathcal{D}(Q \parallel P)$ with respect to the single parameter v . This results in

$$\nabla_v \mathcal{D}(Q \parallel P) = \nabla_v \mathbb{E}_Q \left[\log \left(1 + e^{-H^{n+1}(\boldsymbol{\mu})} \right) \right] + \nabla_v h(\boldsymbol{\theta}^n, \boldsymbol{\Sigma}^n, \boldsymbol{\theta}^{n+1}, \boldsymbol{\Sigma}^{n+1}). \quad (14)$$

Since we do not have a close-form expression for $\nabla_v \mathbb{E}_Q \left[\log \left(1 + e^{-H^{n+1}(\boldsymbol{\mu})} \right) \right]$, we propose to use infinitesimal perturbation analysis (IPA) to obtain noisy samples of the gradient (see e.g. Kim 2006 or Fu 2008 for an introduction). First, we transform the expectation in (14) into an integration with respect to a standard univariate normal distribution, $\mathbb{E}[\bar{f}(Z)]$, where $Z \sim \mathcal{N}(0, 1)$ and

$$\bar{f}(z) = \log \left(1 + \exp \left\{ -(2Y^{n+1} - 1) \left[(\mathbf{x}^n(p^n(\boldsymbol{\mu}))^\top \boldsymbol{\Sigma}^{n+1} \mathbf{x}^n(p^n))^{1/2} z + (\boldsymbol{\theta}^{n+1})^\top \mathbf{x}^n(p^n) \right] \right\} \right).$$

The next result shows that the conditions for IPA (L'Ecuyer 1995) hold.

Proposition 2 $\nabla_v \mathbb{E}[\bar{f}(Z)] = \mathbb{E}[\nabla_v \bar{f}(Z)]$.

The IPA estimator itself is given as

$$\mathbb{E}[\nabla_v \bar{f}(Z)] \approx \frac{1}{N} \sum_{i=1}^N \nabla_v \bar{f}(Z^{(i)}),$$

where $Z^{(i)}$ are independent samples from a standard normal distribution. We denote the gradient estimator by $\widehat{\nabla}_v \mathbb{E}_Q \left[\log \left(1 + e^{-H^{n+1}(\boldsymbol{\mu})} \right) \right]$ and plug it into (14) for $\nabla_v \mathbb{E}_Q \left[\log \left(1 + e^{-H^{n+1}(\boldsymbol{\mu})} \right) \right]$. This produces an estimator of $\widehat{\nabla}_v \mathcal{D}(Q \parallel P)$, and we can apply the Robbins-Monro stochastic approximation algorithm

$$v_{n+1} = v_n - a_n \widehat{\nabla}_v \mathcal{D}(Q \parallel P),$$

for some suitably chosen stepsize a_n , to find the optimal v^* and thus the optimal λ^* . Then we can apply the updating rules in (11) and (12) to determine the approximate posterior distribution after collecting each observation Y^n .

3 DYNAMIC PRICING POLICY

We have shown a way in which the seller’s beliefs can be updated after observing customer response to a price. It remains to address how that price can be chosen in the first place. In this section, we expand upon the notion of a “Bayes-greedy” pricing policy introduced in Ch. 11 of Powell and Ryzhov (2012). Greedy and semi-greedy policies have been widely studied in the literature on dynamic pricing under environmental uncertainty (see e.g. Harrison et al. 2012), and our policy may also be viewed as part of that realm. However, in the setting of Bayesian logistic regression, the concept of “greedy” admits important nuances.

Ideally, the seller would like to choose the price that maximizes the true revenue curve,

$$p^* = \arg \max_p \frac{p}{1 + e^{-(\boldsymbol{\mu}^\top \mathbf{x}(p))}}, \quad (15)$$

where we emphasize that \mathbf{x} depends on p since p is typically one component of the vector of features. A simple “greedy” policy will simply replace $\boldsymbol{\mu}$ in (15) by the current posterior mean vector $\boldsymbol{\theta}^n$. This is typically the approach used in frequentist models (e.g. in Broder and Rusmevichientong 2012) where an MLE estimator is used in place of $\boldsymbol{\theta}^n$.

In the Bayesian setting, however, this approach will under-perform, because it does not use all of the available information. In particular, it does not account for the uncertainty in our beliefs, expressed by $\boldsymbol{\Sigma}^n$. The covariance matrix is important because it specifies a whole family of possible revenue curves, parametrized by $\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\theta}^n, \boldsymbol{\Sigma}^n)$. Thus, a *Bayes-greedy* policy will still myopically optimize the expected single-period revenue, but the expectation will be over the entire space of revenue curves. That is,

$$p^n = \operatorname{argmax}_p \mathbb{E}[R(p)] = \operatorname{argmax}_p \mathbb{E} \left[\frac{p}{1 + e^{-(\boldsymbol{\mu}^\top \mathbf{x}^n(p))}} \right], \quad (16)$$

where the expectation is taken with respect to the (approximate) posterior joint distribution of the parameters.

3.1 Computation of the Bayes-Greedy Policy

In order to use the Bayes-greedy policy, we require the ability to compute the expectation in (16). The approximate Bayesian model suggests that the posterior distribution is multivariate normal, which leads to another convenient dimension reduction. If $\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\theta}^n, \boldsymbol{\Sigma}^n)$ after collecting n observations, then

$$\boldsymbol{\mu}^\top \mathbf{x} \sim \mathcal{N}((\boldsymbol{\theta}^n)^\top \mathbf{x}, \mathbf{x}^\top \boldsymbol{\Sigma}^n \mathbf{x})$$

for arbitrary \mathbf{x} . Therefore, let $W = \boldsymbol{\mu}^\top \mathbf{x}$, noticing that W is actually a function of p , and rewrite (16) as

$$p^n = \operatorname{argmax}_p \mathbb{E} \left[\frac{p}{1 + e^{-W}} \right], \quad (17)$$

where the expectation is now taken with respect to a univariate normal distribution with appropriately chosen mean and variance.

The expectation in (17) is known as the *logistic-normal integral* (Demidenko 2005), which plays an important role in statistics. However, this integral is impossible to compute analytically. It may be computed using Monte Carlo simulation, in particular using IPA (it can be shown that the relevant conditions hold). However, Crooks (2009) offers a tractable approximation

$$\mathbb{E} \left[\frac{1}{1 + e^{-W}} \right] \approx \frac{1}{1 + e^{-\frac{\mathbb{E}(W)}{\gamma}}},$$

where

$$\gamma = \sqrt{1 + \frac{\pi}{8} \operatorname{Var}(W)}.$$

This leads to an approximate Bayes-greedy policy that can be written as

$$p^n = \operatorname{argmax}_p \frac{p}{1 + e^{-\frac{(\boldsymbol{\theta}^n)^\top \mathbf{x}^n(p)}{\gamma^n(p)}}}, \quad (18)$$

where

$$\gamma^n(p) = \sqrt{1 + \frac{\pi}{8} \mathbf{x}^n(p)^\top \boldsymbol{\Sigma}^n \mathbf{x}^n(p)}.$$

This approximation gives us a closed-form expression for the expected revenue function, so that making a pricing decision using (18) is computationally easier.

3.2 Analysis of the Bayes-Greedy Policy

It can be easily shown that the objective function optimized by the point-estimate policy,

$$R_{PE}(p) = \frac{p}{1 + e^{-(\boldsymbol{\theta}^n)^\top \mathbf{x}^n(p)}},$$

is log-concave (but not concave). As a consequence, this function has a single globally optimal price. We show that the Bayes-greedy objective function in (16) possesses the same property, whence it follows that the idea of a ‘‘Bayes-greedy price’’ is well-defined.

Theorem 3 The Bayes-greedy objective function

$$R_{BG}(p) = \mathbb{E}[R(p)] = \mathbb{E} \left[\frac{p}{1 + e^{-(\boldsymbol{\mu}^\top \mathbf{x}^n(p))}} \right]$$

is quasi-concave in p when $p > 0$.

An important consequence of Theorem 3 is that, if we apply IPA to optimize R_{BG} , we are guaranteed to converge to the optimal price. In general, IPA is only guaranteed to find a local optimum. However, in this case, we can apply stochastic approximation to solve the problem directly instead of using the approximation in (18). However, we are still interested in understanding the approximate problem, since it is easier to solve. One can observe that (18) resembles the point-estimate objective, but with an additional factor $\gamma^n(p)$ incorporating our uncertainty about the regression coefficients. The following proposition summarizes structural properties of this factor.

Proposition 4 The factor $\gamma^n(p) \geq 1$ is convex in p .

The variance factor can be viewed as the risk we have to take when choose a price. In most problem instances that we have observed, the factor $\gamma^n(p)$ is not only a convex function, but also an increasing function within the domain of bidding prices. This suggests that the risk is higher when we take a higher price, but the possible reward is also higher. Moreover, the probability of success decreases when we choose a higher price. However, since the factor $\gamma^n(p)$ is greater than 1, the Bayes-greedy policy tends to explore higher prices than the point-estimate policy, leading to possible higher profit.

4 NUMERICAL EXPERIMENTS

In this section, we present numerical experiments using the approximate Bayesian learning model with stochastic approximation proposed in Section 2 and the Bayes-greedy policy proposed in Section 3. We compare this with several alternative approaches, described as follows:

1. The standard frequentist logistic regression approach with the point-estimate policy. In this approach, logistic regression is reapplied after collecting each observation to estimate parameters in the demand function. Then, a pricing decision is made using the point-estimate policy mentioned in Section 3.

2. The variational lower bound approach in Jaakkola and Jordan (2000), using the point-estimate policy to make pricing decisions.
3. The variational lower bound approach in Jaakkola and Jordan (2000), with the proposed Bayes-greedy policy. This is used to show the advantage of choosing the parameter λ optimally using IPA.

Suppose that we are running a computer company, and one of our standard desktop computer models has production cost c . When we set a price for selling, we restrict the prices to be within the range $[p_l, p_u]$, where the lower bound $p_l \geq c$. This means that we never set a price that is lower than the cost, and it is very unlikely that we will make a sale if the selling price is above p_u . Our objective is to maximize the profit function as in (2). In our experiment, we choose $c = 300$, $p_l = 300$ and $p_u = 500$. We consider a finite number of possible bidding prices from 300 to 500 in increments of 10.

For the purposes of this example, we use a two-parameter model, that is, $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and $\mathbf{x}(p) = [1, p]^\top$. We begin with a prior mean $\boldsymbol{\theta}^0 = [500, -1]^\top$. The corresponding demand and profit curves, based on this prior, are shown as the solid green lines in Figure 1. The figure also shows three different realizations of $\boldsymbol{\mu}$ in which the maximum possible profits are “low,” “medium,” and “high.” Minor changes in the regression parameters can significantly alter the shape of the profit curve. From Figure 1a, we see that the prior is essentially telling us that any customer is highly likely to purchase the computer for prices between \$300 and \$500. Considering the upper and lower bounds we have chosen for our price, this type of prior can be understood as uninformative (since our belief suggests that a customer will buy the product for almost any price in the range).

The prior covariance matrix we choose for one specific setting is given by

$$\boldsymbol{\Sigma}^0 = \begin{bmatrix} 100 & 0 \\ 0 & 0.01 \end{bmatrix}.$$

Instead of interpreting this as our uncertainty about the parameters in the prior belief, it may be more meaningful to consider $\boldsymbol{\Sigma}^0$ as the uncertainty about possible prices that buyers will pay. Notice that the magnitudes of the two variances are quite different, again due to the extreme sensitivity of the profit curve to small changes in the regression parameters, particularly the price sensitivity μ_2 . If the variance of μ_2 is too large, this essentially means that μ_2 has a high probability of being positive, which is quite unlikely to occur in practice. Furthermore, in a practical application of the Bayes-greedy policy, we may have $\mathbf{x}(p) = [\tilde{\mathbf{x}}, p]^\top$, where $\tilde{\mathbf{x}}$ contains product and customer attributes unrelated to the price. For the purpose of myopically optimizing the price, this model is equivalent to a two-parameter model where multiple features are embedded into μ_1 , in which case the variance of μ_1 actually represents the variance of a sum of random variables, and should be much larger than the variance of μ_2 .

We compare the performances, including pricing decisions, single-period profit and cumulative profit, of the approach proposed in this paper (referred to as “IPA-Bayes”) and three alternative approaches. The results are reported for the first 20 iterations and averaged over 1000 sample paths, with the true value of $\boldsymbol{\mu}$ fixed according to the three scenarios shown in Figure 1. In all numerical experiments, the Bayes-greedy policy refers to the approximate policy in equation (18). We briefly discuss each of the three scenarios below.

Low-truth scenario. The parameters of the low truth setting are $\mu_1 = 40$ and $\mu_2 = -0.115$. With pre-specified values for μ_1 and μ_2 , the optimal bidding price is 340. As shown in Figure 2, all four methods start with the same bidding price initially and converge after 6 iterations, but the values they converge to are different. All methods converge to prices below the optimal selling price, but the prices from IPA-Bayes and the logistic regression method are closer to optimal, and produce similar profits. IPA-Bayes adjusts more quickly to new information than the other three methods, without the volatile behavior observed for frequentist logistic regression. Additionally, IPA-Bayes shows advantages in single-period profit during the first 5 iterations, resulting in higher cumulative profit.

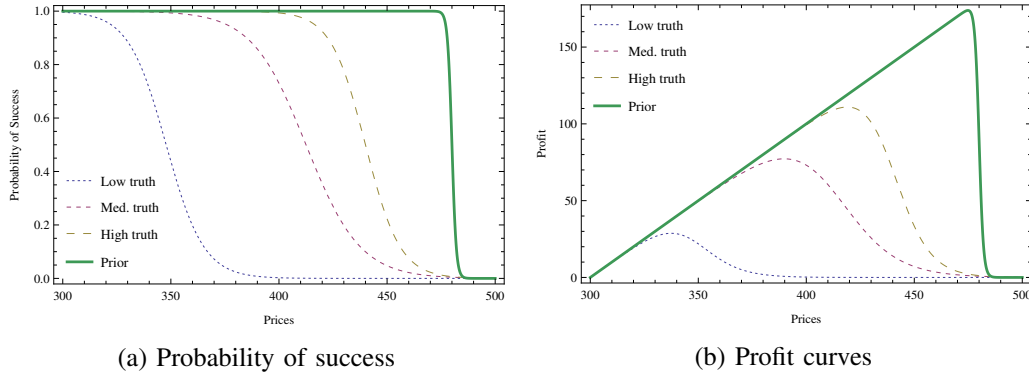


Figure 1: Probability of success and corresponding profit curve as a function of the price under three different scenarios

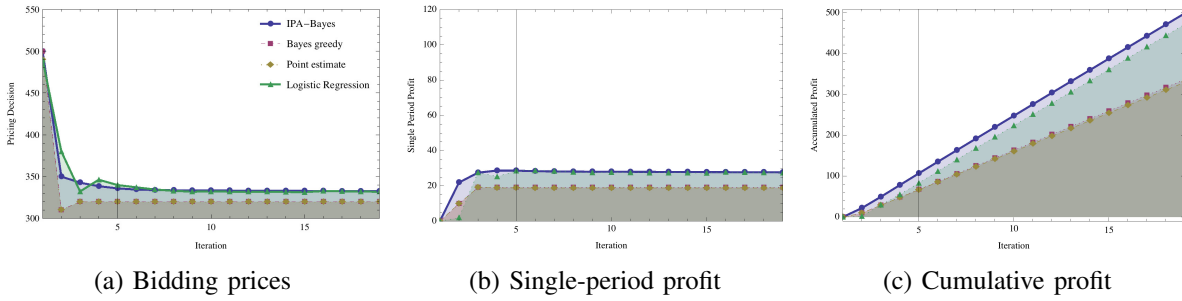


Figure 2: Plots of bidding prices, single-period profit and cumulative profits over time under the low-truth scenario

Medium-truth scenario. The true parameters are $\mu_1 = 32.5$ and $\mu_2 = -0.08$, and the optimal bidding price is 380. Figure 3 shows that all four methods converge to a price close to optimal. Both the single-period profit and the cumulative profit from the IPA-Bayes method dominate those from the other three methods. Note that, while the frequentist method explores higher prices than IPA-Bayes for some time, this behavior is actually too aggressive, and produces lower profits.

High-truth scenario. The parameters of the high-truth setting are $\mu_1 = 55$ and $\mu_2 = -0.125$, with the optimal price being \$420. As shown in Figure 4, both IPA-Bayes and Bayes-greedy start from a lower bidding price than the other methods, due to the effect of the uncertainty factor $\gamma^n(p)$. However, IPA-Bayes quickly adjusts and increases the bidding price, eventually getting close to optimal, and dominating the other methods in terms of single-period profit. After 10 iterations, frequentist logistic regression catches up and produces similar single-period profit, at the expense of volatile behavior and smaller profits in the early iterations.

Discussion. Frequentist logistic regression generally performs well after a few iterations. However, in the sequential setting, we have to refit a new logistic regression model after every observation, which becomes more time-consuming as the number of observations increases. The Bayesian learning model, while less accurate (due to the approximation of conjugacy), provides a quick and efficient way to update parameters, and generally produces a “smoother” sequence of prices; essentially, the uncertainty encoded in the covariance matrix smooths the pricing decision, compared to the volatile prices chosen by the frequentist method in the early iterations. Among the methods using approximate Bayesian inference, IPA-Bayes is consistently the best.

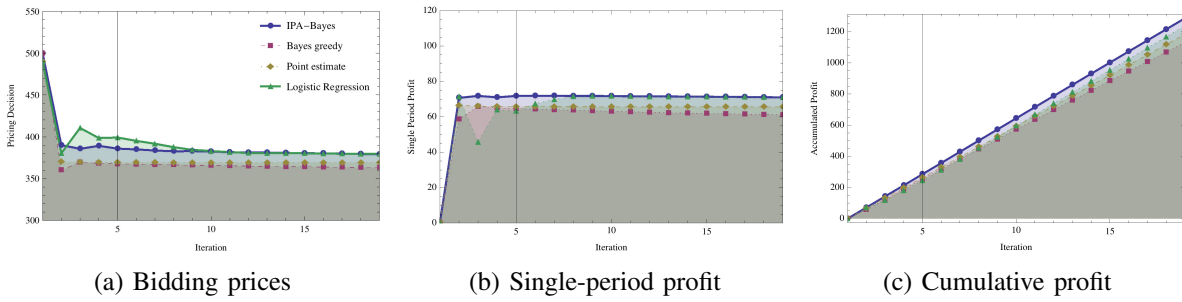


Figure 3: Plots of bidding prices, single-period profit and cumulative profits over time under the medium-truth scenario

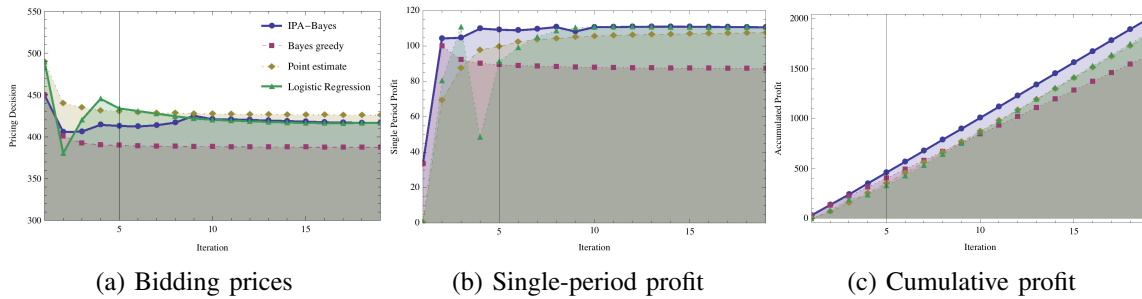


Figure 4: Plots of bidding prices, single-period profit and cumulative profits over time under the high-truth scenario

In these examples, some competing policies tend to perform very similarly to IPA-Bayes after about 10 iterations. However, in the specific context of B2B pricing, the early iterations are especially important because each individual contract tends to have much higher value, and the opportunity cost of pricing suboptimally is more severe. To give some perspective, if the value of every contract is on the order of hundreds of thousands, or millions, of dollars, the overall planning horizon will be shorter, and the first 5 – 10 iterations will become very significant.

5 CONCLUSION

We have presented an approximate Bayesian approach for learning the parameters in a logistic regression model, with specific application to learning revenue curves in B2B pricing problems. We use infinitesimal perturbation analysis and stochastic search to improve the quality of the approximation. We also consider a pricing policy that incorporates uncertainty about the parameters into the estimated expected revenue curve, and chooses a price that optimizes this aggregated function. The proposed model and pricing policy show encouraging results in our empirical experiments. Future work will test the proposed approach on real-world pricing data, where the underlying statistical model can be high-dimensional.

REFERENCES

- Besbes, O., and A. Zeevi. 2009. “Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms”. *Operations Research* 57 (6): 1407–1420.
- Blei, D. M., M. I. Jordan, and J. W. Paisley. 2012. “Variational Bayesian Inference with Stochastic Search”. In *Proceedings of the 29th International Conference on Machine Learning*, 1367–1374.

- Broder, J., and P. Rusmevichientong. 2012. "Dynamic pricing under a general parametric choice model". *Operations Research* 60 (4): 965–980.
- Carvalho, A. X., and M. L. Puterman. 2005. "Learning and pricing in an internet environment with binomial demands". *Journal of Revenue and Pricing Management* 3 (4): 320–336.
- Chhabra, M., and S. Das. 2011. "Learning the Demand Curve in Posted-Price Digital Goods Auctions". In *Proceedings of the 10th International Conference on Autonomous Agents and Multi-Agent Systems*, 63–70.
- Cope, E. 2007. "Bayesian strategies for dynamic pricing in e-commerce". *Naval Research Logistics* 54 (3): 265–281.
- Crooks, G. E. 2009. "Logistic approximation to the logistic-normal integral". Technical report, Lawrence Berkeley National Laboratory.
- Demidenko, E. 2005. *Mixed Models: Theory and Applications*. Wiley Series in Probability and Statistics. Wiley.
- den Boer, A. V., and B. Zwart. 2011. "Dynamic pricing and learning with finite inventories". *Submitted for publication*.
- Farias, V. F., and B. Van Roy. 2010. "Dynamic pricing with a prior on market response". *Operations Research* 58 (1): 16–29.
- Feng, Y., and G. Gallego. 1995. "Optimal starting times for end-of-season sales and optimal stopping times for promotional fares". *Management Science* 41 (8): 1371–1391.
- Fu, M. C. 2006. "Stochastic gradient estimation". In *Handbooks of Operations Research and Management Science, vol. 13: Simulation*, edited by S. G. Henderson and B. L. Nelson, 575–616. North-Holland Publishing, Amsterdam.
- Fu, M. C. 2008. "What you should know about simulation and derivatives". *Naval Research Logistics* 55 (8): 723–736.
- Gallego, G., and G. Van Ryzin. 1994. "Optimal dynamic pricing of inventories with stochastic demand over finite horizons". *Management Science* 40 (8): 999–1020.
- Harrison, J. M., N. B. Keskin, and A. Zeevi. 2012. "Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution". *Management Science* 58 (3): 570–586.
- Jaakkola, T. S., and M. I. Jordan. 2000. "Bayesian parameter estimation via variational methods". *Statistics and Computing* 10 (1): 25–37.
- Kim, S. 2006. "Gradient-based simulation optimization". In *Proceedings of the 2006 Winter Simulation Conference*, edited by L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, 159–167.
- Lazear, E. P. 1986. "Retail Pricing and Clearance Sales". *The American Economic Review* 76 (1): 14–32.
- L'Ecuyer, P. 1995. "On the interchange of derivative and expectation for likelihood ratio derivative estimators". *Management Science* 41 (4): 738–747.
- Minka, T. P. 2000. "Bayesian linear regression". Technical report, Microsoft Research.
- Powell, W. B., and I. O. Ryzhov. 2012. *Optimal Learning*. John Wiley and Sons.
- Qu, H., I. O. Ryzhov, and M. C. Fu. 2012. "Ranking and selection with unknown correlation structures". In *Proceedings of the 2012 Winter Simulation Conference*.
- Spall, J. C. 2005. *Introduction to stochastic search and optimization: estimation, simulation, and control*. Wiley-Interscience.
- Spiegelhalter, D. J., and S. L. Lauritzen. 1990. "Sequential updating of conditional probabilities on directed graphical structures". *Networks* 20 (5): 579–605.

AUTHOR BIOGRAPHIES

HUASHUAI QU is a Ph.D. candidate in Applied Mathematics, Statistics, and Scientific Computation at the University of Maryland. His research interests lie in the broad areas of optimal learning and simulation optimization. He was a recipient of the INFORMS Computing Society Best Student Paper Award and

WSC's Best Theoretical Paper Award in 2012. His email address is huashuai@math.umd.edu.

ILYA O. RYZHOV is an Assistant Professor in the Robert H. Smith School of Business at the University of Maryland. He received a Ph.D. in Operations Research and Financial Engineering from Princeton University in 2011. His research deals with optimal learning and the broader area of stochastic optimization, with applications in disaster relief, energy, and revenue management. He was a recipient of WSC's Best Theoretical Paper Award in 2012. His work has appeared in *Operations Research*, and he is a co-author of the book *Optimal Learning*, published in 2012 by John Wiley & Sons. His email address is iryzhov@rhsmith.umd.edu.

MICHAEL C. FU is Ralph J. Tyser Professor of Management Science in the Robert H. Smith School of Business, with a joint appointment in the Institute for Systems Research and affiliate faculty appointment in the Department of Electrical and Computer Engineering, all at the University of Maryland. His research interests include simulation optimization and applied probability, with applications in supply chain management and financial engineering. He has published four books: *Conditional Monte Carlo: Gradient Estimation and Optimization Applications* (1998 INFORMS Simulation Society Outstanding Publication Award); *Simulation-based Algorithms for Markov Decision Processes*; *Perspectives in Operations Research*; and *Advances in Mathematical Finance*. His email address is mfu@umd.edu.