Revenue maximization through dynamic pricing under unknown market behaviour

Sergio Morales-Enciso[1] and Jürgen Branke[2]

1 **Centre for Complexity Science, The University of Warwick**
Coventry, CV4 7AL, UK
S.Morales-Enciso@warwick.ac.uk

2 **Warwick Business School, The University of Warwick**
Coventry, CV4 7AL, UK
Juergen.Branke@wbs.ac.uk

Abstract

We consider the scenario of a multimodal memoryless market to sell one product, where a customer's probability to actually buy the product depends on the price. We would like to set the price for each customer in a way that maximizes our overall revenue. In this case, an exploration vs. exploitation problem arises. If we explore customer responses to different prices, we get a pretty good idea of what customers are willing to pay. On the other hand, this comes at the cost of losing a customer (when we set the price too high) or selling the product too cheap (when we set the price too low). The goal is to infer the true underlying probability curve as a function of the price (market behaviour) while maximizing the revenue at the same time. This paper focuses on learning the underlying market characteristics with as few data samples as possible by exploiting the knowledge gained from both exploring potentially profitable areas with high uncertainty and optimizing the trade-off between knowledge gained and revenue exploitation. The response variable being binary by nature, classification methods such as logistic regression and Gaussian processes are explored. Two new policies adapted to non parametric inference models are presented, one based on the efficient global optimization (EGO) algorithm and the second based on a dynamic programming approach. Series of simulations of the evolution of the proposed model are finally presented to summarize the achieved performance of the policies.

1998 ACM Subject Classification I.2.6 Learning

Keywords and phrases Dynamic pricing, revenue management, EGO, Gaussian processes for classification

Digital Object Identifier 10.4230/OASIcs.SCOR.2012.11

1 Introduction and literature review

Dynamic pricing is a strategy which aims to offer different prices for the exact same product to different customers. In general, this strategy is followed in order to maximize a firm's revenue by understanding how the market reacts to different prices.

Depending on the nature of the service or product offered, there are different possible scenarios. When designing a dynamic pricing policy, the distinction on whether finite or infinite inventories and time horizons are being considered is important. Another possible distinction is the nature of the market in terms of buying recurrence and the existence of a precedent reference price because recurrent customers for an already established product have been shown to develop a peak-end memory effect which influences their behaviour towards price changes [13]. In this paper, we consider the memoryless scenario with an infinite time horizon and infinite inventory. This is commonly the case when a new non seasonal product

3rd Student Conference on Operational Research (SCOR 2012).
Editors: Stefan Ravizza and Penny Holborn; pp. 11–20
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

is introduced to the market and both the expected life cycle of the product and the available inventory are sufficiently large.

Dynamic pricing strategies have been long studied in the context of physical distribution channels where advertised prices are targeted to the whole market and every price change carries a cost. Examples of this can be tracked back to revenue management research in the airline industry started in the 70s [10], or to [7] where optimal policies are studied for seasonal products with finite inventories sold through physical distribution channels. These studies try to understand the behaviour of the demand in terms of the arrival rates, which is affected after a period of time when prices are changed for a given epoch. Since all the market has access to the same information (advertised price), a change in the arrival rate of customers who actually buy is expected.

In the current market conditions, the internet offers a perfect scenario for dynamic pricing, since prices can be changed individually for each customer without incurring any cost. In fact, this is already a common practice among internet retailers as is shown by the controversial example of Amazon.com, which in September of 2000 ran a randomized pricing test across different customers [17], along with many other examples which are described in [9]. Many recent studies considering the internet as a distribution channel and accounting for more frequent price changes have been developed. An interesting characteristic of the recent studies is that they all consider the relation between the advertised price and the arrival rate of the buying customers as a descriptive measure of the market behaviour, like when dealing with physical distribution channels. Furthermore, most of the reviewed papers use a parametric regression model when inferring the demand curve, imposing a functional form to the unknown market behaviour (e.g. [3], [1], and [4]). Some of the studies justify the choice of a parametric model (like [6]), and only a few make use of non parametric models [2].

In order to take full advantage of the virtual markets, and keeping in mind that the firm aims to maximize the accumulated revenue and not to control the market behaviour, different prices can be quoted to each customer without disclosing the quoted price to the rest of the market. However, the fact that the information is not available to everyone removes the relationships between the quoted prices and the rate of arriving customers. For instance, reducing the price of a product will not necessarily increase the arrival rate of customers unless it is widely advertised, which is exactly what would be avoided in real cases if the market price sensitivity is to be studied. Because of this, we propose to estimate the overall market price sensitivity through the estimation of the probability of buying a product given a quoted price (see section 2), and we assume constant arrival rates, removing the need of a time index. The traditional approach to this problem and our proposed approach are equivalent in the sense that the arrival rate of customers with probability 1 of buying –as modelled in the former– and the probability of customers (buying and not buying) arriving at a constant rate –as proposed it the latter– are interchangeable. The difference mainly relies in the number of samples needed to understand the market behaviour and the way of performing the experiment, i.e. quoting undisclosed prices directly to the customer.

The goal of this paper is to provide insight on which is the best pricing policy to follow in order to maximize the accumulated revenue of a firm having to determine the price of a product in an unknown market under the described framework. We propose 2 policies (EGO and one step lookahead in revenue) and compare their performance with another 2 policies known in literature (random exploration and greedy) as well as with the optimal solution as a benchmark, since in a realistic scenario it would be unknown. EGO (Effective global optimization) policy is based on the methodology outlined in [8] and takes samples at the maximum expected improvement point. The one step lookahead policy is based on a dynamic programming approach and maximizes the overall revenue of the 2 next samples, which implicitly takes into account the information gained during the first sample, but does not include an explicit term to quantify information acquisition as opposed to the one step lookahead policy proposed in [6].

The next section addresses the problem of how to infer the probability of getting a positive answer from a customer given a quoted price, which is required by the policies to function. Section 3 details the derivation of the compared policies taking care of the mathematical details, and section 4 describes the implementation details and the results obtained for each of the compared policies. Finally, in section 5 our conclusions and future research paths are presented.

The main contributions of this paper are first, the use of Gaussian process for classification (GPC), a non parametric method of inference, as detailed in section 2. Second, the design of two new policies, which are well suited to work with non parametric models. Third, the consideration of multimodal markets as explained in 2, and –more importantly– fourth, the change of paradigm to approach the problem by proposing to use the probability of buying rather than the arrival rates as a description of the market behaviour.

2 Market behaviour inference

Every time a price is quoted to a customer, the customer has the choice to accept the product at the quoted price or to reject it. If the probability distribution for a customer accepting an offer at every given price were known it would be straightforward to determine the price to be quoted so that it would maximize the expected income. Nevertheless in our case, and often in real life, the probability distribution of obtaining a positive answer from the customer is not known for every possible price. This could be determined by making an extensive survey, but it would be suboptimal since many samples would be required at very low and very high prices which do not provide any profit. Besides, it is desirable to start maximizing the profit from the first quotes.

The aim in this section is to determine an accurate probability distribution $\mathbb{P}(y = 1|x;\mathcal{D})$ for each price $x \in [0, x_{max}]$ given the minimum possible number of observations $\mathcal{D} = \{(x_i, y_i)_{i=1}^n\} = (X, Y)$ so that the optimal price to be quoted to the customer can be found. This means that the response variable $y \in \{0, 1\}$ is binary by nature.

2.1 Logistic regression

One possibility is to use logistic regression, which assumes a functional form (1) for the market demand, and despite the name is a generalized linear model which works as a k-class classifier rather than as a regression. As shown in chapter 4 of [5], Bayesian logistic regression allows to express not only the expected mean probability, but also confidence in the estimate by using some approximations of which the interested reader can find further details in [12]. If the market is composed of only one type of customer, a logistic regression of first order should be used. But if for example the market is composed of more than one type of customer, each with different price sensitivity, the aggregated distribution would be multimodal and the inference process would require the inclusion of higher order transformations in the design matrix Φ in (1) in order to accurately capture this property. Multimodal markets are common and as a simple example we can consider a university, where staff members might be willing to pay more than students for a same product. A more realistic example is an online retailer selling products across different countries with different incomes and preferences.

$$\mathbb{P}(y=1|x;\mathcal{D}) = \frac{1}{1+e^{-\omega^T\Phi(X)}} \text{ , where } \omega^T \text{ is the transpose of the coefficient vector } \omega. \quad (1)$$

Knowing the number of modes beforehand will be a problem if the composition of the market is unknown. Failing to use the correct degree for the regression will result in a poor estimate, which limits in a considerable fashion the performance of any sampling policy using this inference method when the composition of the market is unknown. This is the main reason why logistic regression is not incorporated in any of the proposed sampling policies. Another drawback of the logistic regression for our goal is that it is difficult to incorporate prior information. In the Bayesian framework, the prior information relates to the coefficients ω to be inferred, and provides a way to express a posterior distribution on the coefficients inferred from the data. But for design matrices of order higher than 1, the functional form of the resulting curve is not trivial to control through ω.

2.2 Gaussian processes for classification (GPC)

In order to avoid the problem of determining the possibly multimodal composition of the market and to allow more flexibility to the model to adapt to the true shape of the market behaviour, a non parametric model is suggested. In particular, a GPC is used.

When used for regression, a Gaussian process (GP) is fully defined by a mean function which allows to introduce any prior information available into the model, and a covariance function which expresses the correlation between the data points [14]. As a result of applying GP for regression to a dataset, we obtain an estimate on the function generating the data, also called latent function f, along with the confidence of such estimate. This means that not only do we get the best fit of a function to the data, but also a distribution for each point of the function expressing how certain we are about the obtained estimate. Since the range of the latent function is $\mathbb{R}$, f is not suitable to be interpreted as a probability. So, in order to ensure the output falls in the interval [0,1], which is required for the classification case, a sigmoid function (λ) is applied to f. A complete and formal description on GP can be found in [14] and [11].

For our case, the GPC is defined by the following mean (m), covariance (k), and sigmoid (λ) functions:

$$m(x) = 0, \quad k(x,x') = \sigma_f exp\Bigg(-\frac{(x-x')^2}{2l^2}\Bigg), \quad \lambda(f) = \frac{1}{1+e^{-f}} \qquad (2)$$

Using a zero mean function means that no prior information is being introduced, or equivalently, the probability of a customer buying the product is $\frac{1}{2}$ a priori. Nevertheless, introducing any prior information on the shape of the resulting probability curve to be inferred would be as simple as changing the prior mean to the believed shape in order to improve the inference process. Furthermore, if the probability curve were suspected to follow a monotonously decreasing behaviour with respect to price, a fair assumption in many cases, it could be done by following the proposed methodology in [15]. The squared exponential covariance function (2) specifies how much a given data point influences the points in its vicinity and how far the vicinity extends to. The optimal parameters $\theta^* = (\sigma_f^*, l^*)$ are to be learnt from the available observations by maximizing the logarithm of the likelihood of the parameters given the data ($log(\mathcal{L}(\theta|\mathcal{D})) = -\frac{1}{2}Y^TK^{-1}Y - log|K| - \frac{n}{2}log(2\pi)$) with respect to θ. K is the $n \times n$ covariance matrix containing the resulting value of the kernel function of all possible combinations of the observations. $\lambda(f)$ (2) is the logistic function which is used to shrink

f so that the output can be interpreted as a probability. Once the optimal parameters are known, the expected value of f evaluated at a given price is estimated using (3) and the variance of the estimate is given by (4).

$$\mathbb{E}[f(x_{new})|\mathcal{D}, \theta^*] = k(x_{new}, X)K^{-1}Y \tag{3}$$

$$Var[f(x_{new})|\mathcal{D}, \theta^*] = k(x_{new}, x_{new}) - k(x_{new}, X)K^{-1}k(X, x_{new}) \tag{4}$$

Where $k(x_{new}, X)$ is the $1 \times n$ row vector resulting from applying the kernel function from the new data point to all the data points in $\mathcal{D}$ and $k(X, x_{new}) = k(x_{new}, X)^T$.

Finally, the probability of a customer accepting the quote given the price x_{new} is given by (5) and the confidence of this prediction is given by (6).

$$\mu(x) := \mathbb{P}(y_{new} = 1|x_{new}; \mathcal{D}) = \lambda\Big(\mathbb{E}[f(x_{new})|\mathcal{D}, \theta^*]\Big) \tag{5}$$

$$\sigma^2(x) := Var[\mathbb{P}(y_{new} = 1|x_{new}; \mathcal{D})] = \lambda\Big(\mathbb{E}[f(x_{new})|\mathcal{D}, \theta^*] + Var[f(x_{new})|\mathcal{D}, \theta^*]\Big) - \lambda\Big(\mathbb{E}[f(x_{new})|\mathcal{D}, \theta^*]\Big) \tag{6}$$

Figure 1 shows the resulting probability of buying $\mathbb{P}(y = 1)$ for all the prices as inferred from $n = 5$ observations using a GPC. The observations were obtained by sampling from a bimodal market, which is represented by the weighted sum of the probabilities of buying which are Bernoulli distributed with parameter $q(x)$ as shown in plot (a) in Figure 1.

3 Policies to compare

The purpose of a policy is to have a rule which dictates the best action to take given our current knowledge at any given point in time, providing a systematic way of taking decisions. Under stochastic conditions an optimal policy can only guarantee up to some probability that the recommended action will be the best. In general, when a process can be simulated, repeated realizations of the same process are performed in order to learn the best course of action. When trying to learn the market behaviour, there are two main challenges. First, the goal is to learn the transition probabilities governing the process, i.e. the distribution of the possible outcomes given the current state and action taken. This makes the process impossible to simulate, since the probability distribution underlying the samples is unknown. And second, the value of taking an action, i.e. the obtained revenue for a given quote in this case, depends not only on the current state but on all the history of actions taken on which the current belief is based, making the process non-Markovian. This makes the problem to be a partially observable non Markov decision process (POnMDP), which can be treated as a POMDP where the state space grows exponentially each time an action is taken [16].

This paper considers five policies. The first, called optimal policy or Π_{true}, is provided only as a benchmark and upper bound for the others since it requires the true probability curve to be known. It consists simply of repeatedly quoting the price x^* that maximizes the expected revenue given complete information. The next two are standard policies that have been proposed earlier (c.f. [6] for example) out of which one is the random policy Π_r and the other is the greedy policy Π_g. Π_r takes samples at the price x resulting from a uniform distribution over the whole interval of possible prices $x^* \sim U[x|0, x_{max}]$. This means

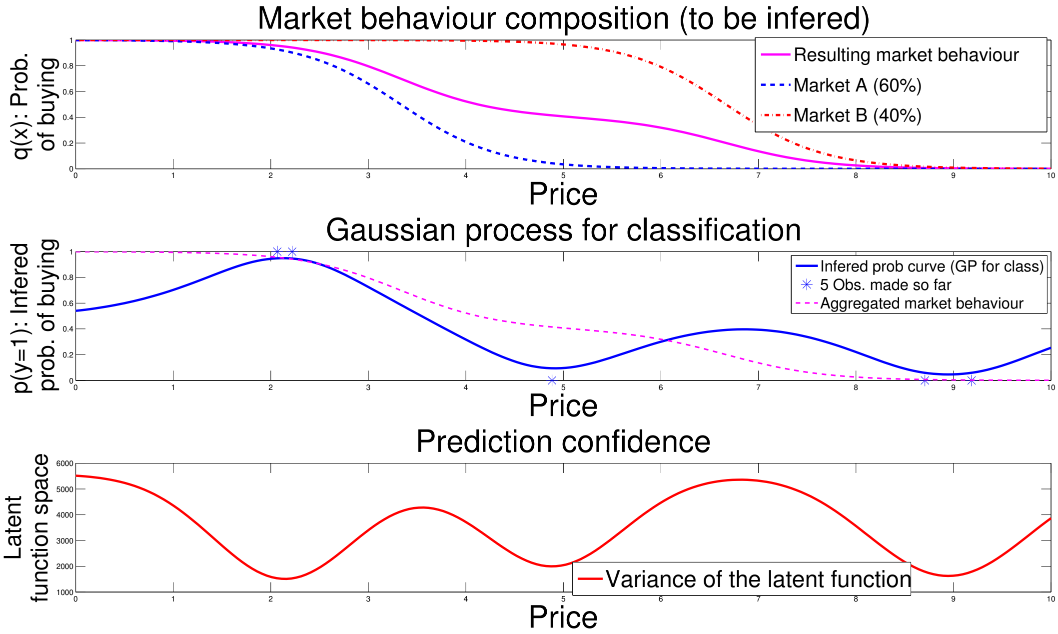

Figure 1 Illustration of market behaviour inference using GPC. In (a), a bimodal aggregated market behaviour resulting from the composition of two types of customers with different proportions is shown. The estimated probability of a customer accepting a quote resulting from using a GPC with only 5 observations is illustrated in (b), and (c) shows the variance of our estimations made in (b) which decreases (i.e. confidence increases) where there are samples to support the belief.

that it only focuses on exploration in a completely uninformed way and does not focus on exploitation at all. Finally two new policies are proposed and explained in detail: EGO and one step lookahead in revenue.

3.1 Greedy policy Π_g

The greedy policy seeks to maximize the immediate revenue, so it samples where the expected revenue at the current time is maximized given the most recent state of knowledge $\mathcal{D}_n$. The expected revenue $\mathbb{E}_{x_n}[R(x_n)|\mathcal{D}_n]$ is calculated directly by multiplying the probability of a customer accepting a quote given the quote times the quoted price. The next best sample is chosen according to (7).

$$x_n^* = \underset{x_n}{\operatorname{argmax}}(\underset{x_n}{\mathbb{E}}[R(x_n)|\mathcal{D}_n]) = \underset{x_n}{\operatorname{argmax}}\left(x_n\mathbb{P}(y_n = 1|x_n, \mathcal{D}_n)\right) \tag{7}$$

3.2 EGO Policy Π_{ego}

EGO stands for efficient global optimization, and is a methodology proposed in [8], illustrated in particular in the engineering context where the need of fitting response surfaces from data samples often arises. The algorithm is based on sampling where the expected improvement is maximized. The improvement $I(x)$ is defined as the difference between the current maximum of the expected revenue known so far r^* and any other possible revenue at the given x provided it is larger than r^* ($I(x) = 0$ otherwise). To calculate the expected improvement $\mathbb{E}[I(x)]$, the probability distribution of the possible revenues at x must be known. This can be calculated since the GP provides the distribution of the belief across the possible

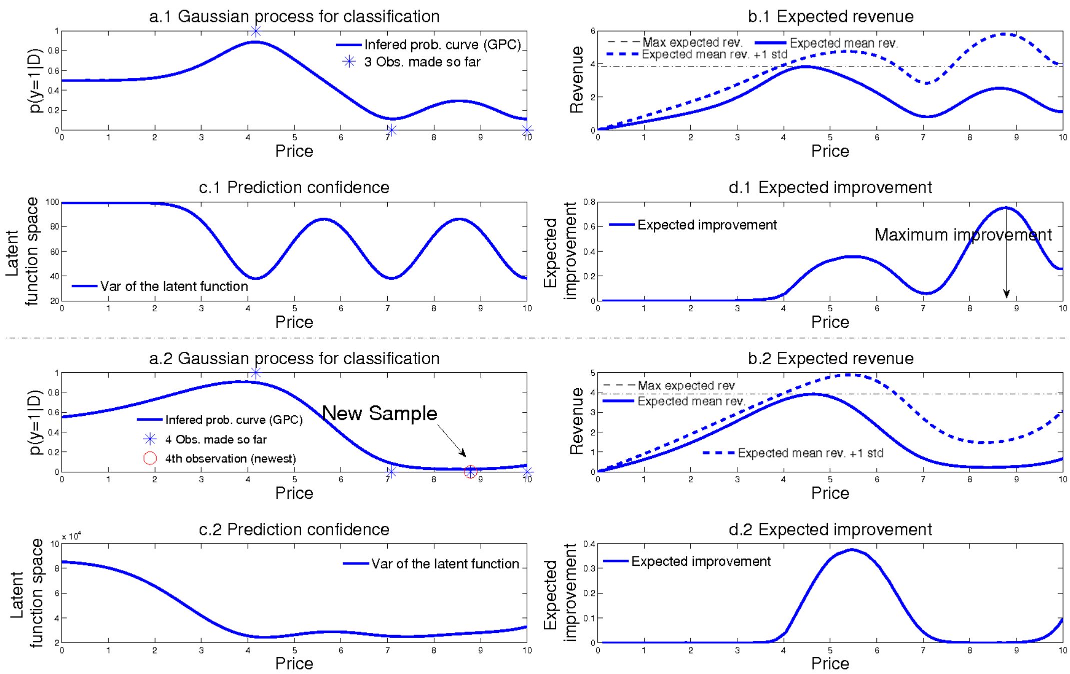

Figure 2 Two steps of the EGO policy in action are shown. Starting with (a.1), which shows the estimated probability of buying given the 3 shown data points, and (c.1) showing the confidence of the predictions, the expected revenue along with its confidence interval can be calculated as (b.1) illustrates. Then, using (9), the expected improvement can be calculated (d.1) and the best action to take determined by using (10).

values, which is specified by the mean (5) and variance (6). By multiplying these two values times the price x, we obtain the distribution on the expected revenue at x, which is used to compute the expected improvement (9).

$$I(x) = \max(r^* - \mathbb{E}[R(x)], 0) \quad \text{where} \quad r^* = \max_x(\mathbb{E}[R(x)]) \tag{8}$$

$$\mathbb{E}[I(x)] = \int_{r^*}^{\infty} I(x)\mathbb{P}(\mathbb{E}[R(x)])dx \quad \text{where} \quad \mathbb{P}(\mathbb{E}[R(x)]) \sim \mathcal{N}(\mathbb{E}[R(x)]|\mu(x), \sigma^2(x)) \tag{9}$$

Once the expected improvement $\mathbb{E}[I(x)]$ is known, the next best sample should be taken where $\mathbb{E}[I(x)]$ is maximized (10).

$$x^* = \underset{x}{\operatorname{argmax}}(\mathbb{E}[I(x)]) \tag{10}$$

The EGO policy explicitly takes into account both information acquisition and exploitation of what is believed to be the action with the highest reward. Besides, as the number of samples increases, the confidence intervals narrow, placing each time less weight in the exploration part. An illustration of how this policy works is provided in Figure 2.

3.3 One step lookahead in revenue Π_{dp1}

The one step lookahead in revenue policy proposes to sample at the maximum of the sum of the immediate expected revenue given the current observations $\mathbb{E}[R(x_n)|\mathcal{D}_n]$ plus the expected revenue at the next step given the current data together with the outcome of the action taken in the first step $\mathbb{E}[R(x_{n+1})|\mathcal{D}_n \cup \{(x_n, y_n)\}]$ appropriately weighted by the current belief (11). The first part of the sum corresponds to the greedy policy Πg. Since the action taken in step n influences our belief on the market behaviour, in order to calculate the second part of (11), all the possible outcomes of action x_n along with their possible responses y_n and the corresponding belief update, which follows from having a new sample, should be taken into account.

$$\mathbb{E}_{x_n}[R(x_n)+R(x_{n+1})] = \max_{x_n \in \mathcal{X}} \left\{ \mathbb{E}_{x_n}\left[R(x_n)|\mathcal{D}_n\right] + \mathbb{E}_{x_n}\left[\max_{x_{n+1}}(\mathbb{E}_{x_{n+1}}[R(x_{n+1})|\mathcal{D}_n \cup \{(x_n, y_n)\}])\right] \right\} \tag{11}$$

Let $\mathbb{P}_n := \mathbb{P}_n(y_n = 1|\mathcal{D}_n)$ denote the estimated probability of a customer buying a product considering the observations available at step n, let $\mathbb{P}^+_{n+1} := \mathbb{P}_{n+1}(y_{n+1} = 1|\mathcal{D}_n \cup \{(x_n, y_n = 1)\})$ denote how the estimation of the probability would update at step $n + 1$ had the action x_n been taken and had its outcome been a positive answer $y_n = 1$, and let $\mathbb{P}^-_{n+1} := \mathbb{P}_{n+1}(y_{n+1} = 1|\mathcal{D}_n \cup \{(x_n, y_n = 0)\})$ denote the case where the outcome for action x_n had been a negative answer $y_n = 0$.

Since there are only two possible outcomes for y_n, and the revenue obtained when $y_n = 0$ is 0, only the cases with positive answers contribute to the expected revenue. However, the two possible outcomes must be taken into account after taking action x_n, since in both cases the belief would be updated in a different manner. Calculating the expectations given the available data, obtained by multiplying the price times the probability of getting a yes, we obtain:

$$\mathbb{E}_{x_n}[R(x_n)+R(x_{n+1})] = \max_{x_n} \left\{ x_n\mathbb{P}_n + \mathbb{P}_n \max_{x_{n+1}} \left(x_{n+1}\mathbb{P}^+_{n+1}\right) + (1-\mathbb{P}_n) \max_{x_{n+1}} \left(x_{n+1}\mathbb{P}^-_{n+1}\right) \right\} \tag{12}$$

Finally, the next best sample according to Π_{dp1} is given by finding the price to be quoted such that it maximizes (12), which is expressed in (13)

$$x^*_n = \underset{x_n}{\operatorname{argmax}} \ \mathbb{E}_{x_n}[R(x_n) + R(x_{n+1})] \tag{13}$$

4 Implementation and results

The accumulated revenue achieved throughout a given number of quotes is considered in order to compare the performance of the policies described above. The accumulated revenue is given by the sum of the products of the quotes made times the response obtained: $R_{acc} = X \cdot Y$.

To see how each of the policies perform, a simulation of the process was implemented and the accumulated revenue was tracked for the first 50 samples. The market is assumed to follow a bimodal probability curve $q(x)$ as shown in plot (a) in Figure 1. So, sampling the market is simulated by drawing a response from a Bernoulli distribution with parameter $q(x)$.

For each of the proposed policies, the simulation starts with 2 samples $\{x_0, x_1\} \in [0, 10]$ and their response $\{y_0, y_1\} \in \{0, 1\}$. If there is at least one positive and one negative response, then the policies start to run. Otherwise, new samples are taken until there is at least one of each possible responses. This is done to ensure the inference process is not mislead from the beginning. The new samples (before the policies start to run) are taken at $min(X)/2$ if Y is only composed of zeros, or at $max(X) + (10 - max(X))/2$ if Y is composed only of ones.

For Π_{true} and Π_r there is no need to perform any inference process. For the rest of the policies, each time a new sample is taken, the belief of the probability of buying is updated

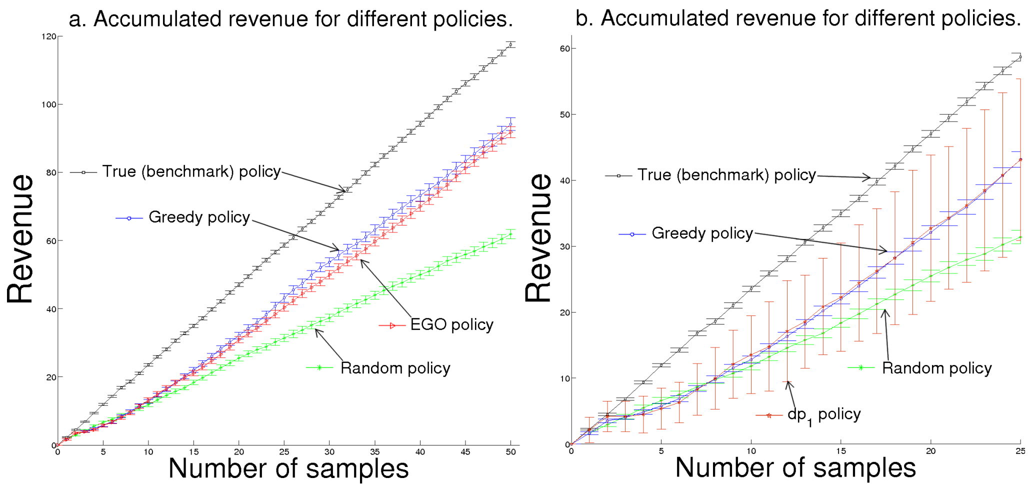

Figure 3 Performance of the five policies described in section 3. No statistically significant difference was found in the performance for policies Π_g, Π_{ego}, and Π_{dp1}, although they all shown better performance compared to Π_r.

by running the GPC and the price at which the next best sample is taken (x^*) is determined by applying each policy.

For all the policies except Π_{dp1} the simulation was run for 100 replications, each with different random seeds, but common random seeds were kept across different policies. The accumulated revenue for each policy and each replication was tracked along 50 samples, allowing to provide the results with confidence intervals. This is shown in Figure 3(a). Π_{dp1} was only ran for 50 replications and up to 25 samples because of its computational requirements. For clarity, the obtained results are presented in a separate plot in Figure 3(b).

After running the numerical simulations, it was found that the random policy Π_r performed the worst. This is due to the focus on constant uninformed exploration and the lack of exploitation. Nonetheless, the simulations show no statistical difference between the other 3 policies compared (Π_g, Π_{ego}, and Π_{dp1}) which seems counter intuitive. One possible explanation could be that no additional information is used across these 3 policies, even if the information available is being treated differently.

5 Conclusion and future research

In this paper, the utility of dynamic pricing used to maximize the accumulated revenue of a firm was reviewed. In particular, a memoryless market with an infinite time horizon and infinite inventory scenario was considered. A new approach to measure the market price sensitivity better adapted to virtual market characteristics was proposed, and two sampling policies (Π_{ego} and Π_{dp1}) were adapted to work with non parametric inference models and compared to 3 other policies commonly found in literature. It has been shown that there is no significant difference between the proposed policies and the greedy policy. This motivates the authors to explore the design of new policies and the adaptation of known policies to the described paradigm. In particular, understanding the strengths of policies known to outperform the greedy policy in the traditional framework can provide insight on how to create better performing policies in the proposed setting. Future lines of research shall include the dynamic case where the market sensitivity changes with time –hinting to constantly consider exploration– and a delayed reward measurement case where the response of a sample is not known until after certain time.

References

1 Victor F. Araman and René Caldentey. Dynamic pricing for nonperishable products with demand learning. *Operations Research*, 57(5):1169–1188, 2009.

2 Victor F. Araman and René Caldentey. Revenue management with incomplete demand information. *Forthcoming in Encyclopedia of Operations Research*, 2011.

3 Yossi Aviv and Amit Pazgal. Pricing of short life-cycle products through active learning. *Working paper.*

4 Yossi Aviv and Amit Pazgal. A partially observed markov decision process for dynamic pricing. *Management Science*, 51(9):1400–1416, 2005.

5 Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics).* Springer-Verlag New York, Inc., 1st edition, 2006.

6 Alexandre X. Carvalho and Martin L. Puterman. Learning and pricing in an internet environment with binomial demands. *Journal of Revenue and Pricing Management*, 3(4):320 – 336, 2005.

7 Guillermo Gallego and Garrett van Ryzin. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science*, 40(8):999–1020, 1994.

8 Donald R. Jones, Matthias Schonlau, and William J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492, December 1998.

9 P.K. Kannan and Praveen K. Kopalle. Dynamic pricing on the internet: Importance and implications for consumer behavior. *International Journal of Electronic Commerce*, 5(3):63 – 83, 2001.

10 Ken Littlewood. *Forecasting and control of passenger bookings.* American Airlines Inc, 1972.

11 David J. C. Mackay. *Information Theory, Inference and Learning Algorithms.* Cambridge University Press, 1st edition, 2003.

12 David J.C. MacKay. The evidence framework applied to classification networks. *Neural Computation*, 4(5):720–736, 1992.

13 Javad Nasiry and Ioana Popescu. Dynamic pricing with loss-averse consumers and peak-end anchoring. *Operations Research*, 59(6):1361–1368, 2011.

14 Carl E. Rasmussen and Christopher Williams. *Gaussian Processes for Machine Learning.* MIT Press, 2006.

15 Jaakko Riihimäki and Aki Vehtari. Gaussian processes with monotonicity information. *Journal of Machine Learning Research: Workshop and Conference Proceedings*, AISTATS2010 special issue, 2010.

16 Stuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach.* Pearson Education, 2nd edition, 2003.

17 Troy Wolverton. Amazon backs away from test prices, 2000. http://news.cnet.com/2100-1017-245631.html (Accessed 01/04/2012).