Probabilistic Pricebots

Amy R. Greenwald
Department of Computer Science
Brown University, Box 1910
Providence, RI 02912 USA
amygreen@cs.brown.edu

Jeffrey O. Kephart
IBM Institute for Advanced Commerce
IBM Thomas J. Watson Research Center
Yorktown Heights, NY 10598 USA
kephart@us.ibm.com

ABSTRACT

Past research has been concerned with the potential of embedding deterministic pricing algorithms into *pricebots*: software agents used by on-line sellers to automatically price Internet goods. In this work, probabilistic pricing algorithms based on *no-regret* learning are explored, in both high-information and low-information settings. It is shown via simulations that the long-run empirical frequencies of prices in a market of no-regret pricebots can converge to equilibria arbitrarily close to an asymmetric Nash equilibrium; however, instantaneous price distributions need not converge.

Keywords

Shopbots, Pricebots, Economic software agents

1. INTRODUCTION

Pricebots, agents that employ automated pricing algorithms, are beginning to appear on the Internet. An early example resides at `buy.com`. This agent monitors its primary competitors' prices and then automatically undercuts the lowest. Driven by the ever-growing use of shopbots, which enhance buyer price sensitivity, we anticipate a proliferation of Internet pricebots, potentially generating complex price dynamics. This paper is concerned with the dynamics of interaction among pricebot algorithms. Ultimately, this line of research aims to identify those pricebot algorithms that are most likely to be profitable, from both an individual and a collective standpoint.

Recently, a simple model of shopbots and pricebots [7] was introduced, and a variety of (mostly deterministic) pricing algorithms were simulated [8]. Motivated in part by a game-theoretic analysis of this model which yields only mixed-strategy Nash equilibria, this paper explores the use of probabilistic pricing based on *no-regret* learning [4, 5], in various informational settings. Among the deterministic algorithms studied previously, one requires complete information about buyer demand and competitors' prices, while a second depends only on the individual pricebot's previous history of prices and profits. One of the early conclusions was that knowledge is power—pricebots that have access to and make use of more information earn greater profits. In the present work it is demonstrated that probabilistic pricebots with low-informational requirements can earn profits comparable to probabilistic pricebots with high-informational requirements, and moreover, the profit earned correspond to the game-theoretic equilibrium values.

This paper is organized as follows. In the next section, we present our model of an economy consisting of shopbots and pricebots. This model is analyzed from a game-theoretic point of view in Section 3. In Section 4, we discuss both the high-information and low-information variants of the price-setting strategies of interest: no internal regret learning and no external regret learning. Section 5 describes simulations of collections of pricebots that implement these strategies, in both informed and naive settings. We find that no-regret pricebots can converge to an equilibrium that is arbitrarily close to Nash equilibrium, although instantaneous price distributions need not converge. Finally, Section 6 is the concluding section, in which we discuss the profitability of deterministic pricing algorithms in comparison with no-regret learning in both high- and low-information settings.

2. MODEL

In this section, we present a summary of the model of shopbots and pricebots; for details, see [7]. Consider an economy in which there is a single homogeneous good that is offered for sale by S sellers and of interest to B buyers, with $B \gg S$. Each buyer b generates purchase orders at random times, with rate ρ_b, while each seller s reconsiders (and potentially resets) its price p_s at random times, with rate ρ_s. The value of the good to buyer b is v_b, and the cost of production for seller s is c_s.

A buyer b's utility for the good is a function of price. In particular, $u_b(p) = v_b - p$, if $p \leq v_b$, and $u_b(p) = 0$, otherwise. In other words, a buyer obtains positive utility if and only if the seller's price is less than the buyer's valuation of the good; otherwise, the buyer's utility is zero. We assume buyers consider the prices offered by sellers using one of the following strategies: (i) *Bargain Hunter:* buyer checks the offer price of all sellers, determines the seller with the lowest price, and purchases the good if that lowest price is less than the buyer's valuation. This type of buyer corresponds to those who utilize shopbots. (ii) *Any Seller:* buyer selects a seller at random, and purchases the good if the price

AGENTS'01, May 28-June 1, 2001, Montréal, Quebec, Canada.

charged by that seller is less than the buyer's valuation. This type of buyer represents those who are either uninformed, or interested in product characteristics other than price: *e.g.*, quality, convenience. Fraction w_A of buyers employs the any seller strategy, while fraction w_B behaves as bargain hunters, with $w_A + w_B = 1$.

A seller s's expected profit per unit time π_s is a function of the price vector $\vec{p}$, as follows: $\pi_s(\vec{p}) = (p_s - c_s)D_s(\vec{p})$, where $D_s(\vec{p})$ is the rate of demand for the good produced by seller s. This rate of demand is determined by the overall buyer rate of demand, the likelihood of the buyers selecting seller s as their potential seller, and the likelihood that seller s's price p_s does not exceed the buyer's valuation v_b. If $\rho = \sum_b \rho_b$, and if $h_s(\vec{p})$ denotes the probability that seller s is selected, while $g(p_s)$ denotes the fraction of buyers whose valuations satisfy $v_b \geq p_s$, then $D_s(\vec{p}) = \rho B h_s(\vec{p}) g(p_s)$. Without loss of generality, define the time scale *s.t.* $\rho B = 1$. Now $\pi_s(\vec{p})$ is interpreted as the expected profit for seller s per unit sold systemwide. Moreover, seller s's profit is such that $\pi_s(\vec{p}) = (p_s - c_s)h_s(\vec{p})g(p_s)$.

3. ANALYSIS

We now present a game-theoretic analysis of the prescribed model viewed as a one-shot game[1], assuming $c_s = c$, for all sellers s, and $v_b = v$, for all buyers b.[2] Recall that $B \gg S$; in particular, the number of buyers is assumed to be very large, while the number of sellers is a good deal smaller. In accordance with this assumption, it is reasonable to study the strategic decision-making of the sellers alone, since their relatively small number suggests that the behavior of individual sellers indeed influences market dynamics, whereas the large number of buyers renders the effects of individual buyers' actions negligible.[3] Assuming the distribution of buyer behavior is exogenously determined and fixed, and that sellers are profit maximizers, we derive the symmetric mixed strategy Nash equilibrium of the sellers' (and later, pricebots') game, and then we derive an asymmetric variant.

A Nash equilibrium is a vector of prices at which sellers maximize individual profits and from which they have no incentive to deviate. [11]. There are no pure strategy (*i.e.*, deterministic) Nash equilibrium in the prescribed model whenever $0 < w_A < 1$ and the price quantum is sufficiently small [7]. There do exist mixed strategy Nash equilibria, however, the symmetric variety of which we derive presently.

[1]The analysis presented in this section applies to the one-shot version of our model, while the simulation results reported in Section 5 focus on repeated (asynchronous) settings. We consider the Nash equilibrium of the one-shot game, rather than its iterated counterpart, for at least two reasons: (i) the Nash equilibrium of the stage game played repeatedly is in fact a Nash equilibrium of the repeated game; (ii) the Folk Theorem of repeated game theory states that virtually all payoffs in a repeated game correspond to a Nash equilibrium, for sufficiently large values of the discount parameter. We isolate the stage game Nash equilibrium as an equilibrium of particular interest.

[2]Assuming uniform valuations is tantamount to assuming $g(p) = \Theta(v - p)$, the step function defined as follows:

$$\Theta(v-p) = \begin{cases} 1 & \text{if } p \leq v \\ 0 & \text{otherwise.} \end{cases} \qquad (1)$$

[3]Nonetheless, in a related study we consider endogenous buyer, as well as seller, decisions.

Let $f(p)$ denote the probability density function according to which all sellers set their equilibrium prices, and let $F(p)$ be the corresponding cumulative distribution function (CDF). In this stochastic setting, the event that any seller is the low-priced seller occurs with probability $[1 - F(p)]^{S-1}$. We now obtain an expression for expected seller demand: $h(p) = w_A(1/S) + w_B[1 - F(p)]^{S-1}$, for $p \leq v$. Note that $h(p)$ is expressed as a function of scalar price p, given that probability distribution $F(p)$ describes the other sellers' expected prices. Profits $\pi(p) = (p-c)h(p)$, for all prices $p \leq v$.

A Nash equilibrium in mixed strategies requires that (i) sellers maximize individual profits, given other sellers' pricing profiles, so as there is no incentive to deviate, and (ii) all prices assigned positive probability yield equal profits otherwise it would not be optimal to randomize. To evaluate those profits, let $p = v$. Buyers are willing to pay as much as v, but no more; thus, $F(v) = 1$. It follows that $h(v) = w_A(1/S)$, and moreover that $\pi(v) = w_A(1/S)(v - c)$ (this implies, incidentally, that total profits are $w_A(v - c)$). Setting $\pi(p) = \pi(v)$ and solving for $F(p)$ yields:

$$F(p) = 1 - \left[\left(\frac{w_A}{w_B S}\right)\left(\frac{v-p}{p-c}\right)\right]^{\frac{1}{S-1}} \qquad (2)$$

which implicitly defines p and $F(p)$ in terms of one another. $F(p)$ is only valid in the range $[0, 1]$. As noted previously, the upper boundary of $F(p)$ occurs at $p = v$; the lower boundary is computed by setting $F(p) = 0$ in Eq. 2:

$$p^* \stackrel{\text{def}}{=} p = c + \frac{w_A(v-c)}{w_A + w_B S} \qquad (3)$$

Thus, Eq. 2 is valid in the range $p^* \leq p \leq v$. A similar derivation of the symmetric mixed strategy equilibrium appears in Varian [14]. Greenwald, *et al.* [8] present various generalizations.

In addition to the symmetric Nash equilibrium, asymmetric Nash equilibria exist, prescribed by the following structure.[4] For $2 \leq \sigma \leq S$, $S - \sigma$ sellers deterministically set their prices to the monopolistic price v, while the remaining σ sellers employ a mixed strategy described by the cumulative distribution function $F_\sigma(p)$, derived as follows. In an asymmetric setting, the event that one of the nondeterministic sellers is lowest-priced occurs with probability $[1 - F_\sigma(p)]^{\sigma-1}$. Now expected demand for the σ sellers pricing according to mixed strategies is given by: $h(p) = w_A(1/S) + w_B[1 - F_\sigma(p)]^{\sigma-1}$. Following the argument given in the symmetric case, we set $\pi(p) = (p - c)h(p)$ equal to equilibrium profits $\pi(v)$ and solve for $F_\sigma(p)$, which yields the following mixed strategy distribution (see Fig. 1):

$$F_\sigma(p) = 1 - \left[\left(\frac{w_A}{w_B S}\right)\left(\frac{v-p}{p-c}\right)\right]^{\frac{1}{\sigma-1}} \qquad (4)$$

4. LEARNING

When the widespread adoption of shopbots by buyers forces sellers to become more competitive, sellers may well respond by creating *pricebots* that automatically set prices so as to maximize profitability. It seems unrealistic, however, to expect that pricebots will simply compute a Nash equilibrium and fix prices accordingly. The real business

[4]The symmetric equilibrium is a special case of the asymmetric one in which $S = \sigma$.

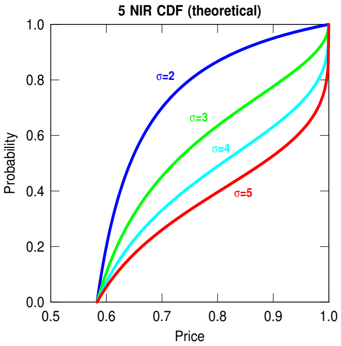

Figure 1: CDFs for mixed-strategy components of asymmetric Nash equilibrium price distribution, for values of $\sigma = 2, 3, 4, 5$; $\sigma = 5$ is the symmetric solution.

world is fraught with uncertainties that undermine the validity of game-theoretic analyses: sellers lack perfect knowledge of buyer demands, and they have an incomplete understanding of their competitors' strategies. In order to be deemed profitable, pricebots will need to learn from and adapt to changing market conditions.

In this paper, we study adaptive pricebot algorithms based on variants of no-regret learning—specifically, no external [5] and no internal regret [4]—emphasizing the differing levels of information on which the algorithms depend. An agent algorithm that requires as input the relevant profits at all its possible price points (including the expected profits that would have been obtained by prices that are not set) are referred to as *informed* algorithms. Those algorithms which operate in the absence of any information other than that which pertains to the actual set price are referred to as *naive* algorithms. We also consider *responsive* variants of (both the informed and naive) no-regret learning algorithms, which learn based on exponentially decayed histories, and are therefore more apt to respond quickly to changes in the environment.

4.1 Definitions

Before describing our simulation results, we define no external and no internal regret, and describe the no-regret algorithms of interest. This description is presented in generic game-theoretic terms, from the point of view of an individual player, as if that player were playing a repeated game Γ^t against nature.[5] From this perspective, let π_i^t denote the payoffs obtained by the player of interest at time t via strategy i. (Let j, as well as i, range over the set of pure strategies M.) Mixed strategy weights at time t are given by the probability vector $q^t = (q_i^t)_{1 \le i \le m}$, where $m = |M|$. The expected payoffs of mixed strategy q^t are denoted $\mathbb{E}[\pi_q^t]$.

Now let h^t be the subset of the history of repeated game Γ^t that is known to the agent at time t. Let H^t denote the set of all such histories of length t, and let $H = \bigcup_0^\infty H^t$. A learning algorithm A is a map $A : H \to Q$, where Q is the agent's set of mixed strategies. The agent's mixed strategy at time $t+1$ is contingent on the elements of the history known through time t: *i.e.*, $q^{t+1} = A(h^t)$. The no-regret learning algorithms of interest in this study depend only on historic information regarding the payoffs obtained through time t, unlike for example fictitious play, which explicitly depends on the strategies of all players as well as all payoffs. For an informed player, a history h^t has the form $\langle i^1, (\pi_i^1)\rangle, \ldots, \langle i^t, (\pi_i^t)\rangle$, where (π_i^τ) is the vector of informed payoffs at time $1 \le \tau \le t$ for all strategies i. For a naive player, a history h^t has the form $\langle j^1, \pi_{j^1}^1\rangle, \ldots, \langle j^t, \pi_{j^t}^t\rangle$, which records only the payoffs of the strategy j^τ that is played at time $1 \le \tau \le t$.

[5]Nature is taken to be a conglomeration of all opponents.

The regret ρ the player feels for playing strategy i rather than strategy j is simply the difference in payoffs obtained by these strategies at time t: $\rho(j, i) = \pi_j^t - \pi_i^t$. Suppose a player's learning algorithm prescribes that he play mixed strategy q^t at time t. Then the regret the player feels toward strategy j is the difference between the expected payoffs of strategy q^t and the payoffs of strategy j, namely: $\rho(j, q^t) = \pi_j^t - \mathbb{E}[\pi_q^t]$. A learning algorithm A is said to exhibit ϵ-*no external regret* iff for all histories h^t, for all strategies j,

$$\lim_{T \to \infty} \sup \frac{1}{T} \sum_{t=1}^{T} \rho(j, q^t) < \epsilon \tag{5}$$

where $q^t = A(h^t)$ for all $1 \le t \le T$. In words, the limit of the sequence of average regrets between the player's sequence of mixed strategies and all possible fixed alternatives is less than ϵ. If the algorithm exhibits ϵ-no external regret for all $\epsilon > 0$, then it is said to exhibit no external regret.

No internal regret can be understood in terms of conditional regrets. Given an algorithm A that generates sequence of plays $\{i^t\}$, the conditional regret $R^T(j, i)$ the player feels toward strategy j conditioned on strategy i is the sum of the regrets *at all times t that the player plays strategy i*:

$$R_A^T(j, i) = \sum_{\{1 \le t \le T | i^t = i\}} \rho(j, i) \tag{6}$$

A learning algorithm exhibits no internal regret iff in the limit it yields no conditional regrets on average. Expressed in terms of expectation, algorithm A satisfies *no internal regret* iff for all histories h^t, strategies i, j, $\epsilon > 0$,

$$\lim_{T \to \infty} \sup \frac{1}{T} \sum_{t=1}^{T} q_i^t \rho(j, i) < \epsilon \tag{7}$$

where $q^t = A(h^t)$ for all $1 \le t \le T$. It is well-known that an algorithm satisfies the no internal regret property iff its empirical distribution of play converges to correlated equilibrium (see, for example, Foster and Vohra [4] and Hart and Mas-Colell [10]). Moreover, no internal regret implies no external regret, and these two properties are equivalent in two strategy games.

4.2 No External Regret Learning

Freund and Schapire [5] propose an algorithm (NER) that achieves no external regret via a multiplicative updating scheme. Their algorithm is dependent on the cumulative payoffs achieved by all strategies, including the surmised payoffs of strategies which are not in fact played. Let P_i^t denote the cumulative payoffs obtained through time t via strategy i: *i.e.*, $\mathrm{P}_i^t = \sum_{x=1}^t \pi_i^x$. Now the weight assigned to strategy i at time $t+1$, for $\alpha > 0$, is given by:

$$q_i^{t+1} = \frac{(1+\alpha)^{\mathrm{P}_i^t}}{\sum_j (1+\alpha)^{\mathrm{P}_j^t}} \tag{8}$$

The naive variant of NER, namely NER_ϵ is obtained by (i) imposing an artificial lower bound on the probability with which strategies are played, in order to ensure that the space of strategies is adequately explored, and (ii) utilizing estimates of cumulative payoffs that depend only on payoffs obtained by strategies that are actually employed. For $\epsilon \in (0,1]$, let $\hat{q}_i^t = (1-\epsilon)q_i^t + \epsilon/m$ be the weight assigned by NER_ϵ to strategy i, and let $\hat{\pi}_i^t = \mathbf{1}_i^t \pi_i^t / \hat{q}_i^t$.[6] Estimated cumulative payoffs (notation $\hat{\text{P}}_i^t$) are given by $\hat{\text{P}}_i^t = \sum_{x=1}^t \hat{\pi}_i^x$. NER_ϵ updates weights according to the update rule given in Eq. 8, but $\hat{\text{P}}_i^t$ is used in place of P_i^t. NER_ϵ is due to Auer, *et al.* [1]. NER can be also made responsive via exponential smoothing. Given $\gamma \in (0,1]$, NER_γ is defined by substituting $\tilde{\text{P}}_i^t$ into Eq. 8, where in informed settings and naive settings, respectively, $\tilde{\text{P}}_i^{t+1} = (1-\gamma)\, \tilde{\text{P}}_i^t + \pi_i^{t+1}$ and $\tilde{\text{P}}_i^{t+1} = (1-\gamma)\tilde{\text{P}}_i^t + \hat{\pi}_i^{t+1}$.

4.3 No Internal Regret Learning

We now discuss no internal regret learning (NIR), due to Foster and Vohra [4], and a simple implementation due to Hart and Mas-Colell [10]. The regret felt by a player at time t is formulated as the difference between the payoffs obtained by utilizing the player's strategy of choice, say i, and the payoffs that could have been achieved had strategy j been played instead: $r_{i\to j}^t = q_i^t(\pi_j^t - \pi_i^t)$. The cumulative regret $\text{R}_{i\to j}^t$ is the summation of regrets from i to j through time t: $\text{R}_{i\to j}^t = \sum_{x=1}^t r_{i\to j}^x$. Now internal regret is defined as follows: $\text{IR}_{i\to j}^t = (\text{R}_{i\to j}^t)^+$, where $X^+ = \max\{X, 0\}$. If the regret for having played strategy j rather than strategy i is significant, then the NIR procedure for updating weights increases the probability of playing strategy i. According to Hart and Mas-Colell, if strategy i is played at time t,

$$q_j^{t+1} = \frac{1}{\mu}\text{IR}_{i\to j}^t \quad \text{and} \quad q_i^{t+1} = 1 - \sum_{j\neq i} q_j^{t+1} \tag{9}$$

where $\mu \geq (m-1)\max_{j\in M} \text{IR}_{i\to j}^t$ is a normalizing term.

Like NER, NIR depends on complete payoff information at all times t, including information that pertains to strategies that are not employed at time t. NIR_ϵ, which is applicable in naive settings, depends on an estimate of internal regret that is based only on the payoffs obtained by the strategies that are actually played, and the approximate weights associated with those strategies. An estimated measure of expected regret $\hat{r}_{i\to j}^t$ is given by $\hat{r}_{i\to j}^t = \hat{q}_i^t(\hat{\pi}_j^t - \hat{\pi}_i^t) = (\hat{q}_i^t/\hat{q}_j^t)\mathbf{1}_j^t\pi_j^t - \mathbf{1}_i^t\pi_i^t$, where $\hat{q}_i^t$ and $\hat{q}_j^t$ are defined as in NER_ϵ. NIR_ϵ updates weights using Eq. 9, with estimated cumulative internal regret $\hat{\text{IR}}_{i\to j}^t$, based on $\hat{\text{R}}_{i\to j}^t$, in place of $\text{IR}_{i\to j}^t$.

No internal regret learning can also be made responsive (NIR_γ) in both informed and naive cases via an exponential smoothing of regret. Given $\gamma \in (0,1]$, exponentially smoothed cumulative regret, denoted $\tilde{\text{R}}_{i\to j}^t$, is computed in terms of either $r_{i\to j}^t$ or $\hat{r}_{i\to j}^t$, depending on whether the setting is informed or naive: *i.e.*, $\tilde{\text{R}}_{i\to j}^{t+1} = (1-\gamma)\tilde{\text{R}}_{i\to j}^t + r_{i\to j}^t$, or $\tilde{\text{R}}_{i\to j}^{t+1} = (1-\gamma)\tilde{\text{R}}_{i\to j}^t + \hat{r}_{i\to j}^t$. NIR_γ then uses $\tilde{\text{IR}}_{i\to j}^t = (\tilde{\text{R}}_{i\to j}^t)^+$ as its measure of internal regret.

5. SIMULATIONS

This section describes simulations of markets in which anywhere from 2 to 5 adaptive pricebots employ various mixtures of no regret pricing strategies. At each time step, a pricebot, say s, is randomly selected and given the opportunity to set its price, which it does by generating a price according its current price distribution. Profits are then computed for all pricebots. The profits for pricebot s are taken to be *expected* profits,[7] given current price vector $\vec{p}$:

$$\pi_s(\vec{p}) = \left[\frac{w_A}{S} + \delta_{\lambda_s(\vec{p}),0}\frac{w_B}{\tau_s(\vec{p})+1}\right](p_s - c) \tag{10}$$

where $\lambda_s(\vec{p})$ is the number of competitors' prices that are less than p_s and $\tau_s(\vec{p})$ denotes the number that are exactly equal to p_s. Given its profits, pricebot s uses its respective learning algorithm to update its price distribution. At this point in our simulations, we measure the Kolmogorov-Smirnov (K-S) distance e between the *symmetric* Nash equilibrium CDF and the empirical CDF, computed as the average of the absolute differences between these CDFs over all m prices p_i:

$$e = \frac{1}{m}\sum_{i=1}^m |F_{\text{symNash}}(p_i) - F_{\text{emp}}(p_i)| \tag{11}$$

Strictly speaking, Eq. 10 holds only if p_s does not exceed the buyers' valuation v; otherwise the seller's profit is zero. For simulation purposes, this property was ensured by constraining the pricebots to set their prices among a discrete set of cardinality $m = 51$, spaced equally in the interval $[c, v]$, where $c = 0.5$ and $v = 1$. The mixture of buyer types was set at $w_A = w_B = 0.5$. Simulations were iterated for 100 million time steps for NIR pricebots and 10 million time steps for NER pricebots.

5.1 NIR Pricebots

We now present simulation results for no internal regret learning. Our main observation is that NIR pricebots, both informed and naive, converge to Nash equilibrium. This is not entirely surprising, as NIR is known to converge within the set of correlated equilibria [4], of which Nash equilibria form a (proper) subset. Furthermore, NIR has previously been observed to converge to Nash, rather than correlated, equilibria in games with small numbers of strategies [6]. In the present model, where the number of strategies varies between 51 and 501, we again find that NIR converges to Nash equilibrium. The detailed nature of the convergence, however, is quite different between NIR and NIR_ϵ.

First, we consider informed NIR pricebots. In simulations of 2 to 5 NIR pricebots, the empirical price distributions have been observed to evolve to a mixed strategy Nash equilibrium—usually an asymmetric one. A typical example involving 3 informed NIR_γ pricebots is shown in Fig. 2a. In this experiment, the responsiveness parameter was set to a relatively small value, $\gamma = 10^{-6}$; the results are qualitatively similar for the ordinary non-responsive form of the learning algorithm: *i.e.*, $\gamma = 0$. The long-run cumulative empirical probability distributions coincide almost perfectly with the theoretical asymmetric Nash equilibrium for $\sigma = 2$: one pricebot always sets its price to 1, while the other 2 play the mixed-strategy equilibrium computed in Eq. 4 with $\sigma = 2$.

Figs. 2b and 2c reveal that the convergence path to the asymmetric Nash equilibrium is not as regular as one might suppose: the system experiences a punctuated equilibrium.

[6] $\mathbf{1}_i^t$ is the indicator function, which has value 1 if strategy i is employed at time t, and 0 otherwise.

[7] We have also experimented with profits based on simulated buyer purchases, which introduced noise into the profit function. While the amplitude of the noise decreases as the size of the buyer population increases, such increases also increase simulation times. Expected profits enabled us to explore the behavior in the limit of infinitely large buyer populations without suffering inordinately long running times.

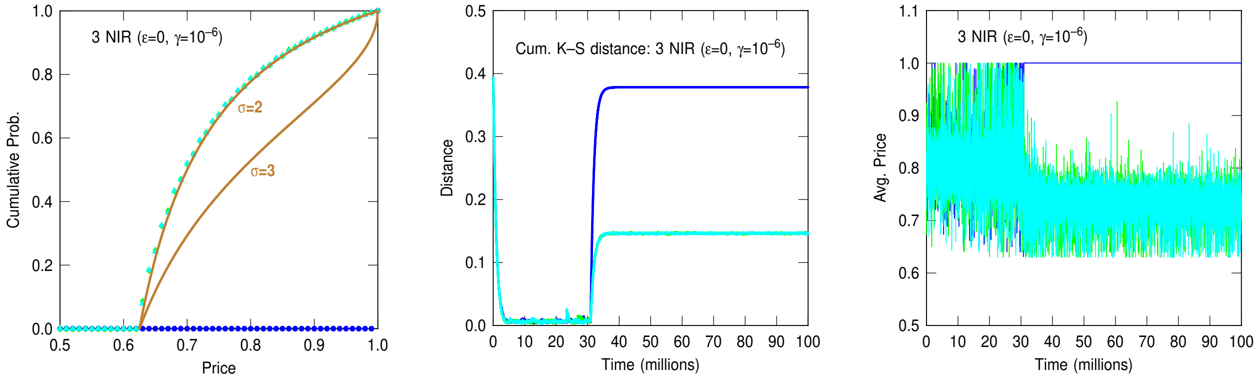

Figure 2: 3 informed NIR_γ pricebots, $\gamma = 10^{-6}$. a) Empirical CDF at time 100 million, with symmetric ($\sigma = 3$) and asymmetric ($\sigma = 2$) Nash CDFs superimposed. b) K-S distance between empirical and symmetric Nash CDFs over time. c) Instantaneous average prices over time.

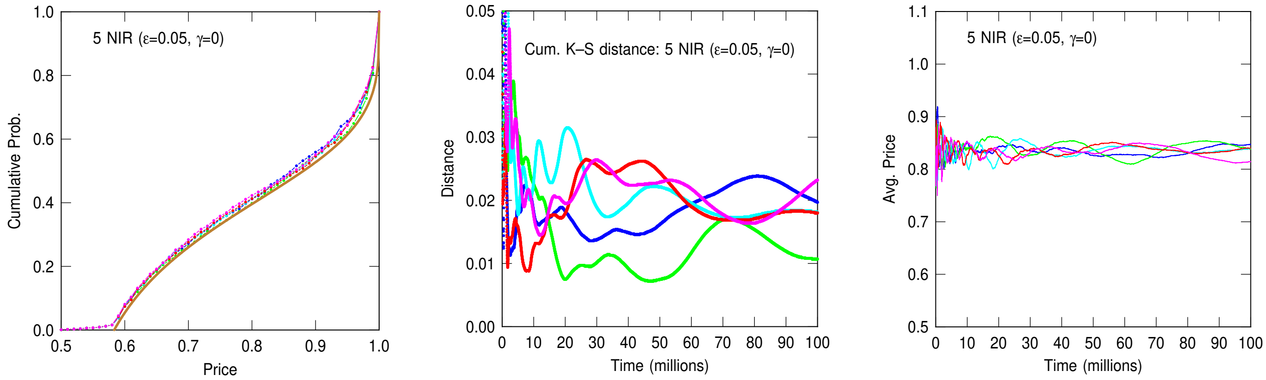

Figure 3: 5 naive NIR_ϵ pricebots, $\epsilon = 0.05$. a) Empirical CDF at time 100 million, with symmetric Nash CDF superimposed. b) K-S distance between empirical and symmetric Nash CDFs over time. c) Instantaneous average prices over time.

Fig. 2b displays the K-S distance between the *symmetric* Nash equilibrium CDF and the empirical CDF. By about time 3 million, the empirical price distributions for the pricebots begin to closely approximate the *symmetric* Nash equilibrium ($\sigma = 3$). This behavior continues until about time 30.92 million, with the K-S distance hovering close to 0.005 for each pricebot. Quite abruptly, however, at time 30.93 million, the K-S distance for one pricebot quickly rises toward 0.378, while that of the other 2 rises to 0.146. These are precisely the K-S distances between the symmetric Nash distribution and the pure and mixed components of the asymmetric Nash distribution with $\sigma = 2$. The conclusion is clear: at time 30.93 million, there is a spontaneous and abrupt transition from the symmetric ($\sigma = 3$) to the asymmetric ($\sigma = 2$) Nash equilibrium.

Fig. 2c provides some additional insight into the learning dynamics. At a given moment in time, each pricebot maintains an instantaneous model price distribution from which it randomly draws its price. Fig. 2c displays the mean of this model distribution as a function of time. The average prices are seen to be highly volatile. (The *actual* prices set by the pricebots vary even more wildly with time!) The sudden shift between Nash equilibria is again evident from this viewpoint: at time roughly 30 million, one of the average prices is pinned at 1, while the other prices start fluctuating wildly around 0.736, consistent with $\sigma = 2$. In numerous experiments (of 2 to 5 NIR pricebots), we consistently observed equilibrium shifts, always toward lower values of σ. Fig. 2c, which portrays instantaneous average prices over time, suggests that volatility decreases with σ, which partly explains why equilibria shift in this manner. Intuitively, the volatility of an equilibrium consisting of entirely mixed strategies exceeds that of an alternative equilibrium consisting of some pure and some mixed strategies.

The volatility of the average prices suggests that the symmetric mixed-strategy Nash equilibrium is in some sense unstable. At any one moment, the various pricebots are likely to be generating prices according to distributions that diverge substantially from the Nash distribution. Moreover, the instantaneous model distributions drift very quickly over time, with little temporal correlation even on a time scale as short as 10000 time steps. Even before the shift between equilibria is made, it is evident that the pricebots have often experimented with the deterministic strategy (always set price to 1). Remarkably, the pricebots' learning algorithms exert the right pressure on one another to ensure that, aver-

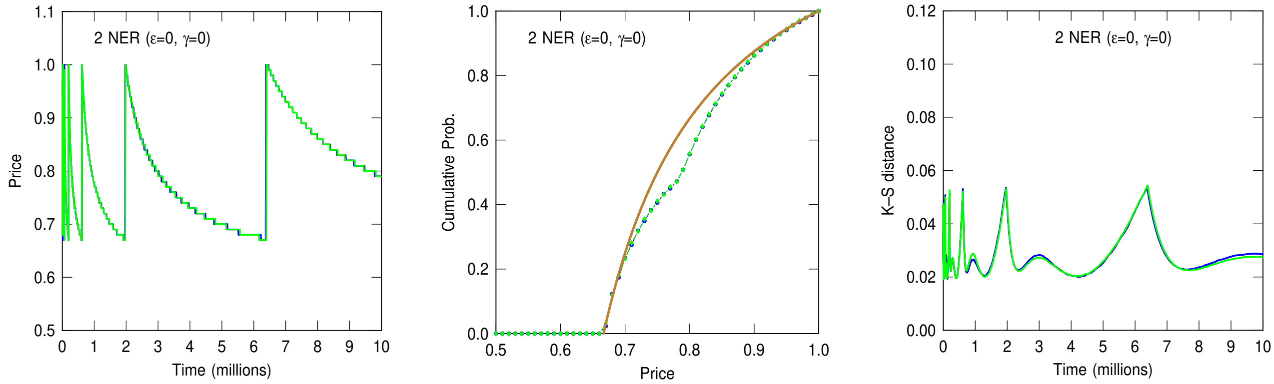

Figure 4: 2 informed NER pricebots. a) Actual prices over time. a) Empirical CDF at time 10 million, with symmetric Nash CDF superimposed. c) K-S distance between empirical and symmetric Nash CDFs over time.

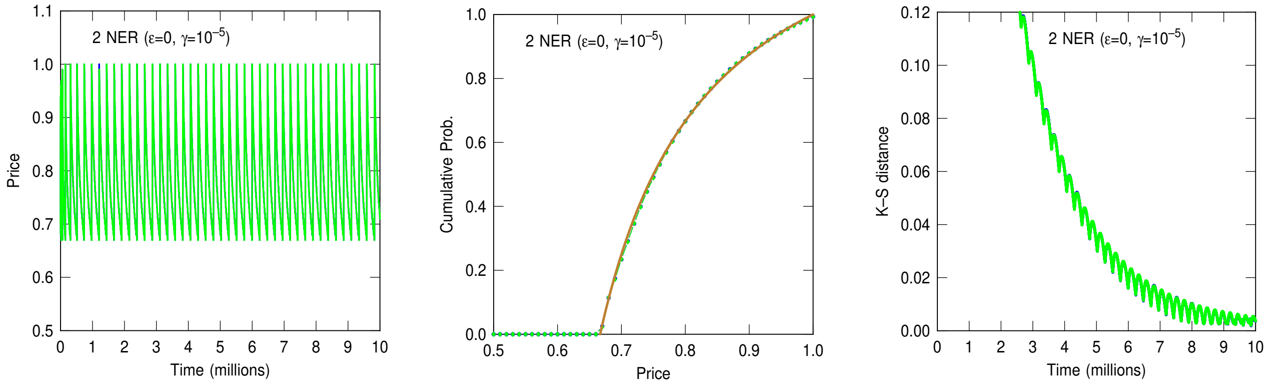

Figure 5: 2 informed, responsive NER pricebots; $\gamma = 10^{-5}$. a) Actual prices over time. b) Empirical CDF at time 10 million, with exponential smoothing over time scale of 2 million. c) K-S distance between empirical and symmetric Nash CDFs.

aged over a long period of time, the distribution of prices actually set by each pricebot is a Nash distribution—typically symmetric at first, becoming more and more asymmetric as time progresses. The time-average profits also reflect this: they are approximately 0.0838 for each pricebot, which is quite close to the theoretical value of 0.0833 for both the symmetric and asymmetric Nash equilibria.

We have also conducted numerous simulations of naive NIR$_\epsilon$ pricebots, finding that the empirical distributions always converge to distributions that become arbitrarily close to the *symmetric* Nash equilibrium, as the exploration parameter $\epsilon \to 0$. Fig. 3a shows the CDF for 5 NIR$_\epsilon$ pricebots with $\epsilon = 0.05$, all of which overlay the symmetric Nash CDF very closely. The K-S distances, depicted in Fig. 3b, settle in the range of roughly 0.01 and 0.025. The instantaneous average prices computed from the pricebots' model distributions are plotted in Fig. 3c. The volatile behavior of the informed NIR pricebots is somewhat diminished; average prices for NIR$_\epsilon$ pricebots oscillate near the mean price of the symmetric Nash equilibrium, namely 0.81. For $\epsilon = 0.02$, the average NIR$_\epsilon$ pricebots' profits were 0.0498, negligibly lower than the theoretical value of 0.05. In contrast, for $\epsilon = 0.1$, the CDFs for all pricebots are all extremely close to one another, but there is a consistent bias away from the symmetric Nash price distribution—one that leads to somewhat lower prices on average. Moreover, the instantaneous average prices (and therefore the underlying price distributions) are considerably less volatile. Increasing the amount of exploration leads to greater consistency and stability, but also leads to greater deviations from the Nash distribution.

5.2 NER Pricebots

In this section, we present simulation results for no external regret pricing (NER, with $\alpha = 0.5$). There are a number of learning algorithms that satisfy the no external regret optimality criterion (the earliest are due to Hannan [9] and Blackwell [2]). Since no internal regret and no external regret are equivalent in two-strategy games, no external regret algorithms converge (in empirical frequencies) to correlated equilibrium in two-strategy games. Greenwald *et al.* [6] report that many no external regret algorithms converge to Nash equilibrium in games of two strategies.

As illustrated in Fig. 4, the learning dynamics are quite different in the present case, where there are $m = 51$ different strategies. In Fig. 4a, 2 informed NER pricebots (with $\alpha = 0.5$) never settle at a deterministic equilibrium. Moreover, the empirical CDF depicted in Fig. 4b deviates significantly from the Nash CDF; and the K-S distance shown in Fig. 4c oscillates indefinitely with an exponentially increasing period, and is apparently bounded away from zero. This behavior indicates that, unlike NIR pricebots, the long-run empirical behavior of 2 NER pricebots will never reach the symmetric Nash CDF, even after infinite time.

Instead, the 2 NER pricebots engage in cyclical price wars, with their prices highly correlated, behavior not unlike myopic (MY) best-response pricebots [7]. NER price war cycles differ from MY price war cycles, however, in that the length of NER cycles grows exponentially, whereas the length of MY cycles is constant. This outcome results because NER (non-responsive) pricebots learn from the ever-growing history dating back to time 0, while myopic learning at time t is based only on time $t-1$. The play between 2 NER pricebots in the present setting is reminiscent of fictitious play in the Shapley game, a 2-player game of 3 strategies for which there is no pure strategy Nash equilibrium [12].

In order to eliminate exponential cycles, we now turn to the responsive algorithm NER$_\gamma$, with responsiveness parameter $\gamma = 10^{-5}$ that smooths the pricebots' observed history. This smoothing effectively limiting the previous history to a finite time scale on the order of $1/\gamma$. Price war cycles are still observed, but they quickly converge to a constant period of roughly S/γ (see Fig. 5a). In order to compute an empirical CDF, we again used exponential weighting to smooth the empirical play, but rather than using a time scale of 100,000 (which would only remember a portion of the price-war cycle) we lengthened the time scale to 2 million. The smoothed empirical CDF that results (see Fig. 5b) is extremely close to the Nash CDF, with a final K-S distance of only 0.0036.

Fig. 5c depicts the K-S distance between the smoothed empirical distribution and the symmetric Nash equilibrium over time. Both pricebots' errors are plotted, but the values are so highly correlated that only one error function is apparent. The errors diminish in an oscillatory fashion over time, reaching 0.0036 after 10 million time steps. The long-run empirical distribution of play of responsive NER$_\gamma$ pricebots is bounded away from Nash by a small function of γ, but approaches Nash as $\gamma \to 0$.

Finally, we report on results for naive NER$_\epsilon$ pricebots (not shown). For non-responsive NER$_\epsilon$ pricebots, price-war cycles with exponentially increasing period can still be discerned despite being obscured somewhat by a uniform peppering of exploratory prices. The empirical CDF is somewhat closer to the Nash equilibrium than for non-responsive informed NER pricebots, with the cumulative K-S distance dropping to a minimum of 0.021 in the course of its oscillatory trajectory. However, it appears that it is not destined to converge to Nash. For NER pricebots that are both naive and responsive, the price correlations are greatly diminished—prices over time appear random to the naked eye. Nonetheless, the empirical CDF (computed just as for the informed NER pricebots) is again quite close to the Nash CDF. The final K-S distances (after 10 million time steps) for the two pricebots are 0.0136 and 0.0182.

Apparently, responsiveness makes an important difference in NER pricing. For non-responsive NER, the price dynamics never approach Nash behavior, but for responsive NER$_\gamma$, the time-averaged play approaches Nash as $\gamma \to 0$. For responsive NER$_\epsilon$, finite values of γ, ϵ lead to near-Nash time-averaged play, which approaches Nash as $\gamma, \epsilon \to 0$.

6. CONCLUSION

This paper investigated probabilistic no-regret learning in the context of dynamic on-line pricing. Specifically, simulations were conducted in an economic model of shopbots and pricebots. It was determined that both no internal regret learning and (responsive) no external regret learning converge to Nash equilibrium, in the sense that the long-term empirical frequency of play coincides with Nash. Neither algorithm generated probability distributions which themselves converged to the Nash equilibrium, however.

It remains to simulate heterogeneous mixtures of pricebots, both deterministic and non-deterministic. Preliminary studies to this effect suggest that MY (*a.k.a.* Cournot best-reply dynamics [3]), a reasonable performer in high-information settings, outperforms informed no-regret learning algorithms (both NIR and NER). MY has no obvious naive analogue, however, and thus far the naive versions of no-regret learning have outperformed any naive implementation of MY. Thus it seems that no-regret learning would be more feasible than classic best-reply dynamics in pricing domains like the Internet, where only limited payoff information is available.

The scope of the results presented in this paper is not limited to e-commerce models such as shopbots and pricebots. Bots could also be used to make networking decisions, for example, such as along which of a number of routes to send a packet or at what rate to transmit. No-regret learning is equally applicable in this scenario (as noted by the inventors of the no external regret learning algorithm [5]). In future work, it would be of interest to analyze and simulate a game-theoretic model of networking via no-regret learning in attempt to validate the popular assumption that Nash equilibrium dictates a network's operating point [13].

7. REFERENCES

[1] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. Schapire. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, pages 322–331. ACM Press, November 1995.

[2] D. Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6:1–8, 1956.

[3] A. Cournot. *Recherches sur les Principes Mathematics de la Theorie de la Richesse*. Hachette, 1838.

[4] D. Foster and R. Vohra. Regret in the on-line decision problem. *Games and Economic Behavior*, 21:40–55, 1997.

[5] Y. Freund and R. Schapire. Game theory, on-line prediction, and boosting. In *Proceedings of the 9th Annual Conference on Computational Learning Theory*, pages 325–332. ACM Press, May 1996.

[6] A. Greenwald, E. Friedman, and S. Shenker. Learning in network contexts: Results from experimental simulations. *Games and Economic Behavior: Special Issue on Economics and Artificial Intelligence*, In Press 2000.

[7] A. Greenwald and J. Kephart. Shopbots and pricebots. In *Proceedings of Sixteenth International Joint Conference on Artificial Intelligence*, volume 1, pages 506–511, August 1999.

[8] A. Greenwald, J. Kephart, and G. Tesauro. Strategic pricebot dynamics. In *Proceedings of First ACM Conference on E-Commerce*, pages 58–67, November 1999.

[9] J. Hannan. Approximation to Bayes risk in repeated plays. In M. Dresher, A. Tucker, and P. Wolfe, editors, *Contributions to the Theory of Games*, volume 3, pages 97–139. Princeton University Press, 1957.

[10] S. Hart and A. M. Colell. A simple adaptive procedure leading to correlated equilibrium. Technical report, Center for Rationality and Interactive Decision Theory, 1997.

[11] J. Nash. Non-cooperative games. *Annals of Mathematics*, 54:286–295, 1951.

[12] L. Shapley. A value for n-person games. In H. Kuhn and A. Tucker, editors, *Contributions to the Theory of Games*, volume II, pages 307–317. Princeton University Press, 1953.

[13] S. Shenker. Making greed work in networks: A game-theoretic analysis of switch service disciplines. *IEEE/ACM Transactions on Networking*, 3:819–831, 1995.

[14] H. Varian. A model of sales. *American Economic Review, Papers and Proceedings*, 70(4):651–659, September 1980.