Maintaining Equilibria During Exploration in Sponsored Search Auctions*

Jennifer Wortman[1], Yevgeniy Vorobeychik[2], Lihong Li[3], and John Langford[4]

[1] Department of Computer and Information Science, University of Pennsylvania
[2] Department of Electrical Engineering and Computer Science, University of Michigan
[3] Department of Computer Science, Rutgers University
[4] Yahoo! Research, New York, NY

Abstract. We introduce an exploration scheme aimed at learning advertiser click-through rates in sponsored search auctions with minimal effect on advertiser incentives. The scheme preserves both the current ranking and pricing policies of the search engine and only introduces one set of parameters which control the rate of exploration. These parameters can be set so as to allow enough exploration to learn advertiser click-through rates over time, but also eliminate incentives for advertisers to alter their currently submitted bids. When advertisers have much more information than the search engine, we show that although this goal is not achievable, incentives to deviate can be made arbitrarily small by appropriately setting the exploration rate. Given that advertisers do not alter their bids, we bound revenue loss due to exploration.

1 Introduction

Recent years have seen an explosion of interest in sponsored search auctions, due in large part to the unique opportunity for targeted advertising and the resulting billions of dollars in revenue. Most sponsored search auctions display a list of advertisements on the sidebar or other sections of a search engine's results page, ranked by some function of advertisers' revealed willingness-to-pay for every click on their ad. The advertisers in turn pay the search engine for every click their ad receives. While several pricing schemes have been circulated in the literature [8], by far the most popular is a generalization of second-price auctions, under which each advertiser pays the lowest bid that is sufficient to ensure that the ad remain in its current slot. Typically the number of available slots for advertisements on the first search page is fixed, and thus only high ranking advertisements are displayed.

An essential part of both designing sponsored search auction mechanisms and bidding in them is the knowledge of the probability that a given ad is clicked each time it is displayed in a particular slot for a particular search query or keyword. This probability is known as the *click-through rate* or *CTR* of the ad. Knowledge of these click-through rates helps advertisers determine optimal bidding behavior. CTRs can also be an integral part of the ad ranking policy. For example, it is common for policies to rank bidders by the product of their bid and some function of their *relevance*, a slot-independent measure of CTR. Throughout the paper, we assume that CTRs do not change over time.

Most of the existing literature on sponsored search auctions treats CTRs as known. When advertisers first enter the system, however, their CTRs are not yet known either by the search engine or even by the advertisers themselves, and can only be estimated over time based on the observed clicks. Observations are inherently limited to slots in which ads appear, and estimates are generally poor for advertisers with low rank that do not usually appear at all. Furthermore, without the assumption of factorable CTRs, little can be said about CTRs of an ad in slots in which it has not previously appeared (or has appeared only a small number of times). Thus there is a need for an exploration policy that periodically perturbs the current slate of displayed ads, showing some in alternate slots and occasionally displaying those ads that are ranked below the last slot. Ideally, this exploration policy should not be difficult to incorporate into the current sponsored

* A preliminary version of this work appeared in the *Third International Workshop on Internet and Network Economics* [13]. Part of this work was done while J. Wortman, Y. Vorobeychik, and L. Li were visiting Yahoo! Research.

search mechanisms. Additionally, if the advertisers' bids have reached an equilibrium, the exploration policy should, when possible, eliminate the incentives for bidders to change their bids, thereby destabilizing the auction. Such destabilization can result in negative user and advertiser experience, as well as unnecessary loss in revenue to the search engine, and can make exploration harder to control.

There are, of course, trivial ways of exploring while maintaining equilibrium bids. For example, the search engine could rank, display, and price the ads as usual with probability α, and rank the ads randomly (say, choosing uniformly among all possible rankings) with probability $1 - \alpha$. The problem with this approach is that in order to maintain equilibrium bids, the search engine must not charge for any clicks received during exploration, which could clearly lead to a significant loss of revenue.

In this paper, we address the problem of learning the click-through rates for each ad in every slot. Our primary goal is to maintain an equilibrium bid configuration if the bidders did indeed play according to an equilibrium prior to exploration, while allowing the search engine to charge for all clicks received during exploration. When advertisers have significantly more knowledge than the search engine, we provide bounds on the amount that any advertiser could gain by deviating. This incentive to deviate can be minimized by reducing exploration, at the cost of slowing down the process of learning the CTRs. Additionally, we bound the revenue loss that the search engine incurs due to exploration, as compared to maintaining a policy based on current estimates of CTRs.

A similar problem has been addressed by Pandey and Olston [10] and Gonen and Pavlov [5]. The former work addresses the learning problem without considering advertiser incentives. The latter addresses both. Our model differs from existing ones in three primary ways:

1. We avoid imposing a particular ranking policy or introducing a new pricing scheme so that changes to existing systems are minimal.
2. The data gathered by our approach can be incorporated into general learning algorithms using sample selection debiasing techniques [6].
3. We avoid the standard but unrealistic assumption that click-through rates can be factored into advertiser- and slot-specific components.

2 Notation and Definitions

We consider an auction for a particular keyword in which there are N advertisers (alternately called bidders or players) placing bids. Since our analysis can be repeated for each keyword, the restriction to a single keyword is without loss of generality. (Indeed, the analysis can even be generalized to incorporate arbitrary context information, as long as the number of contexts is finite and advertisers may submit separate bids for each [3]. In Section 9 we generalize some of our results to the continuous context setting.) We assume that the search engine has K slots with non-negligible CTRs. Throughout the discussion on incentives, we assume that the CTRs depend only on the ad being displayed and the slot in which it is shown. Thus, we use c_i^s to denote the true CTR of player i in slot s. We assume that for each player i, $c_i^s > c_i^t$ whenever $1 \leq s < t \leq K$. For convenience, we define $c_i^s = 0$ for $s > K$ and $s < 1$. In most of our analysis we deal explicitly with estimated click-through rates; the search engine estimates are denoted by $\hat{c}_i^s$, whereas the advertiser i's estimates are denoted by $\tilde{c}_i^s$. Finally, we let v_i denote the value of a click to player i. In other words, v_i is the expected amount that player i will gain from a click before considering payments.

For now we assume that prior to the exploration process, advertisers are ranked according to their bid b_i multiplied by a weight w_i which is an increasing function of their estimated relevance scores for the particular keyword. Setting this weight equal to the advertiser's relevance recovers the standard rank-by-revenue model; setting it equal to 1 recovers rank-by-bid [9]. Each advertiser pays a price per click equal to the lowest bid that maintains his current position; thus the price paid by bidder i in rank s is $p_i^s = w_{s+1}b_{s+1}/w_i$. Without loss of generality, assume that advertisers are indexed in the order in which they are ranked when playing equilibrium, i.e. advertiser i is in slot i in the ranking.

The relevance score of an advertiser, which we denote by e_i, can be thought of as an average CTR over all slots for the given keyword. We might choose to define this relevance as $\sum_{s=1}^{K} c_i^s$ or alternately as

$\sum_{s=1}^{K} c_i^s / c_s$ where c_s is the "average" CTR that any ad might expect to receive on slot s. Observe that when c_i^s is factorable into the product $e_i c_s$, both of these relevance scores are proportional to e_i. We can fix the weights for each advertiser prior to (each phase of) exploration and reveal the new estimates of CTRs at the end of the exploration period only, allowing greater control of exploration.

We assume that prior to exploration the advertisers converge to a symmetric Nash equilibrium, a variant of Nash equilibrium introduced simultaneously by Varian [12] and Edelman et al. [2]. We slightly alter the standard definition to take into account CTR estimates as follows.

Definition 1. *A* symmetric Nash equilibrium (SNE) *is an ordering and a set of bids such that for every player i and for every slot s, $\tilde{c}_i^i \left(v_i - p_i^i\right) \geq \tilde{c}_i^s \left(v_i - p_i^s\right)$, where $\tilde{c}_i^s$ denotes player i's CTR estimate at slot s.*

Existence of at least one symmetric Nash Equilibrium was proved in a slightly different setting than ours by Börgers et al. [1]. Their proof applies essentially without change to our setting. More details can be found in Appendix A.

Theorem 1. *In the generalized second-price auction where the bidders are ranked by $w_i b_i$ and payments are p_i^i, if the bidder i's valuations per click are given by v_i and CTRs for all bidders are strictly decreasing, there exists a symmetric Nash equilibrium.*

3 An Algorithm for Exploration

We begin by describing a simple algorithm for learning click-through rates. Below (in Section 4) we show that we can set parameters of this algorithm in such a way as to minimize or entirely eliminate incentives for advertisers to deviate from a pre-exploration SNE. Our key condition will be that throughout the entire run of the algorithm the prices which the advertisers pay are fixed to their pre-exploration equilibrium prices.

The algorithm, which we call `k-swap` (Algorithm 1), starts by ranking ads by the product of bid and weight as usual, and repeatedly chooses pairs of ads to swap in order to explore. In particular, each time the given keyword receives an *impression* (i.e. each time a query is made on the keyword), a swapping distance $k \in \{1, \cdots, K\}$ is chosen from some distribution (e.g. uniformly at random). The algorithm calculates or looks up a swapping probability for each pair of slots s and $s+k$ that are a distance k apart. (The method for choosing these probabilities will be discussed in Section 4.) Finally, the algorithm uses this set of swapping probabilities to decide which (if any) pair of ads to swap.

We must be careful about how pairs of ads are chosen to be swapped so we can avoid swapping the same ad more than once on a single query. Let S_i denote the event that the ads in slots i and $i+k$ are swapped and let $r_i^k = \Pr(S_i)$ be the probability that this event occurs. We have

$$\Pr(S_i) = \Pr(S_i | S_{i-k}) \Pr(S_{i-k}) + \Pr(S_i | \neg S_{i-k}) \Pr(\neg S_{i-k}).$$

To avoid conflicting swaps, we can set $\Pr(S_i | S_{i-k}) = 0$, which implies that $\Pr(S_i | \neg S_{i-k}) = \Pr(S_i) / \Pr(\neg S_{i-k}) = r_i^k / (1 - r_{i-k}^k)$, which is no greater than one as long as we enforce that $r_{i-1}^k + r_i^k \leq 1$.

For the sake of this algorithm, all ads with rank $K+1, \cdots, N$ can be thought of as sharing slot $K+1$. Thus whenever an ad in slot $s \leq K$ is chosen to swap with slot $K+1$, any ad with rank $K+1, \cdots, N$ could be displayed in slot s. We do not discuss how the algorithm might decide which losing ad to display, but one could imagine giving preference to ads that have not often been displayed in the past.

4 Maintaining Equilibrium During Pairwise Swapping

In this section, we consider the effect on advertiser incentives of implementing an exploration policy that occasionally chooses pairs of ads that are k slots apart to swap or moves an undisplayed ad into slot $K-k+1$ for some *fixed* value of k. By ensuring that advertisers do not have incentives to deviate from equilibrium bids for any fixed k, we ensure that the advertisers do not deviate throughout the entire run of `k-swap`.

Algorithm 1 The `k-swap` algorithm.

```
Input: r_i^k for all 1 ≤ i, k ≤ K
Calculate all swapping probabilities r_i^k
for all queries on the given keyword do
  Randomly select a k ∈ {1, ··· , K}
  for i = 1 to min{k, K − k + 1} do
    Set S_i ← 1 with probability r_i^k, S_i ← 0 otherwise
  end for
  for i = k + 1 to K − k + 1 do {Note that this statement is null if 2k > K}
    if S_{i−k} = 1 then
      Set S_i ← 0
    else
      Set S_i ← 1 with probability r_i^k/(1 − r_{i−k}^k), S_i ← 0 otherwise
    end if
  end for
  for i = 1 to K − k do
    Swap the ads in slots i and i + k if S_i = 1
  end for
  if S_{K−k+1} = 1 then
    Choose an i ∈ {K + 1, ··· , N} to display in slot K − k + 1
  end if
end for
```

We assume that the search engine bases the weights w_i on the CTR estimates $\hat{c}_i^s$, and fix the prices paid by the advertisers through the entire run of `k-swap`. The updated CTR estimates obtained during exploration are only reported to advertisers after the algorithm completes. In practice, the algorithm may need to be run in multiple phases, interleaving exploration with updates of CTR estimates, and allowing sufficient time for advertisers to reach a new equilibrium after each phase.

Our assumptions raise a conceptual question: if the advertisers care about the *real* CTRs, how can we maintain incentives given only estimates? We posit that often advertisers do not know the CTRs any better than the search engine and formulate their own optimization problem (at least approximately) in terms of the estimates provided by the search engine; that is, we assume that $\tilde{c}_i^s = \hat{c}_i^s \; \forall i, s$. We consider the case in which advertisers have additional information about their CTRs and provide bounds on their incentives to deviate as a result of exploration in Section 6.

The results that follow depend on each advertiser's value per click. In practice, it is not necessary for the search engine to know these values in order to benefit from our results. In particular, we do not actually advocate setting the swapping probabilities separately for each individual auction, but rather fixing probabilities in such a way that the guarantees will hold for most "typical" auctions.

Since all analysis in this section is for a fixed value of k, we drop the superscript and use r_i in place of r_i^k to denote be the probability that ads i and $i+k$ are swapped. These probabilities can be represented as multiples of r_1, i.e. $r_i = \alpha_i r_1$. Then, if α_i are set exogenously (for example, $\alpha_i = 1$ for all $1 \leq i \leq K$), `k-swap` has only one tunable parameter, r_1, for a fixed value of k. For convenience of notation, we define $\alpha_i = 0$ for all $i < 1$ and $i > K-k+1$. In order to allow exploration of CTRs of all bidders, we let r_{K-k+1} designate the total probability that *any* losing bidder is swapped into slot $K-k+1$. Let q_s denote the probability that a losing bidder with rank $K+1 \leq s \leq N$ is displayed *conditional* on *some* losing ad being displayed; thus, the probability that a particular losing bidder s gets selected is $q_s r_{K-k+1}$. We have that $\sum_{s=K+1}^{N} q_s = 1$. Finally, define $q_{max} = \max_{K+1 \leq s \leq N} q_s$.

Once we add exploration, the *effective* estimate of CTR for advertiser i in slot s is no longer $\hat{c}_i^s$. Rather, now with some probability r_{s-k} the ad in slot s is moved to slot $s-k$, and with some probability r_s the ad is moved to slot $s+k$. Then the new effective estimate of CTR of player i for rank s is

$$\hat{c}_i'^s = (1 - r_{s-k} - r_s)\hat{c}_i^s + r_{s-k}\hat{c}_i^{s-k} + r_s\hat{c}_i^{s+k}.$$

(Recall that $r_s = 0$ and $\hat{c}_i^s = 0$ for $s < 1$ and $s > K - k + 1$. We can replace CTR with effective CTR because the prices paid by all advertisers remain fixed for the duration of exploration.)

Let $D_{i,s} = \alpha_s(\hat{c}_i^s - \hat{c}_i^{s+k}) - \alpha_{s-k}(\hat{c}_i^{s-k} - \hat{c}_i^s)$. Observe that $r_1 D_{i,s}$ is the marginal CTR loss of advertiser i in slot s when exploration is allowed. We now define the quantities $J_{i,j}$ and Z_i which are used in Theorem 2:

$$J_{i,j} = (v_i - p_i^i)D_{i,i} - (v_i - p_i^j)D_{i,j} \quad (1)$$

$$Z_i = (v_i - p_i^i)D_{i,i} + \alpha_{K-k+1} q_{max} \hat{c}_i^{K-k+1} v_i. \quad (2)$$

To get some intuition about what these mean, note that $r_1 J_{i,j}$ is the difference between the marginal loss in expected payoff due to exploration that the advertiser i receives in slot j and the marginal loss in expected payoff due to exploration in slot i. Similarly, $r_1 Z_i$ is the difference between the marginal loss in payoff due to exploration that the advertiser i receives by switching to rank above $K + 1$ (and thereby not occupying any slot) and the marginal loss due to exploration in slot i.

The following result gives the conditions under which exploration does not incent advertisers to change their bids and characterizes the settings in which this is not possible.

Theorem 2. *Assume that each advertiser $i \in \{1, \cdots, K\}$ strictly prefers his current slot to all others in equilibrium, i.e. the condition $(v_i - p_i^i)\hat{c}_i^i > (v_i - p_i^j)\hat{c}_i^j$ holds for all $1 \le i, j \le K, i \ne j$ whenever $J_{i,j} > 0$ and $v_i - p_i^i > 0$ $\forall i$ whenever $Z_i > 0$. Then for generic valuations and relevances there exists an $r_1 > 0$ such that no advertiser has incentive to deviate from the pre-exploration SNE bids once exploration is added. In particular, any r_1 satisfying the following set of conditions is sufficient:*

$$r_1 \le \min \left\{ \min_{2 \le i \le K} \frac{1}{\alpha_i + \alpha_{i-k}}, \qquad \min_{1 \le i \le K; Z_i > 0} \frac{1}{Z_i}(v_i - p_i^i)\hat{c}_i^i, \right.$$
$$\left. \min_{1 \le i,j \le K; i \ne j; J_{i,j} > 0} \frac{1}{J_{i,j}} \left((v_i - p_i^i)\hat{c}_i^i - (v_i - p_i^j)\hat{c}_i^j \right) \right\}.$$

Before moving on, let us examine the three terms inside the min in this bound. The first term simply enforces that for each ad, the sum of the probability that the ad is swapped to higher slot and the probability that the ad is swapped to a lower slot is no more than 1. The second term ensures that no ad that is displayed at equilibrium has incentive to provide a lower bid and no longer be displayed after swapping is introduced. The final term guarantees that no ad prefers to switch to another slot when swapping is introduced. Intuitively, if these conditions are satisfied, then the advertisers will remain in equilibrium when swapping is introduced.

The proof of Theorem 2, which can be found in Appendix B, is broken into three cases. First it is necessary to show that if the conditions in the theorem statement hold, advertisers $1, \cdots, K$ have no incentive to alter their bids in order to switch to other ranks within $1, \cdots, K$. Next we must show that these advertisers have no incentives to alter their bids in order to move to slots lower than K. Finally, we show that advertisers ranked below slot K before exploration have no incentive to deviate.

To get some intuition about how the theorem can be applied and about the magnitude of r_1, consider the following example.

Example 1. Suppose that there are 3 advertisers bidding on 2 slots. Let $\hat{c}_i^1 = 0.2$ and $\hat{c}_i^2 = 0.1$ for all players $i \in \{1, 2, 3\}$; all players thus have the same weights no matter how the weights are defined. Let $v_1 = 10$, $v_2 = 5$, and $v_3 = 2$. Suppose that prior to exploration each advertiser bids his value per click and pays the next highest bid. One can easily verify that this configuration constitutes a symmetric Nash equilibrium in which player 1 gets slot 1 and player 2 gets slot 2.

Let us fix $\alpha_2 = 1$, so $r_1 = r_2$. Now we can determine the setting of r_1 that allows us to swap neighboring ads ($k = 1$) without introducing incentives to deviate during exploration. Applying the first constraint, we find the condition that $r_1 \le 0.5$ must hold. This is clearly necessary; ad 2 cannot be swapped up and down each with probability more than 0.5.

By the second constraint, since $Z_1 = 1.5$ and $Z_2 = 0.5$, we must have that $r_1 \leq 0.6$. Finally, since $J_{1,2} = 0.5$, we get the condition that $r_1 \leq 0.4$. Combining the effects of these constraints, we see that we can set both swapping probabilities as high as 0.4 without giving any of the advertisers incentive to deviate during exploration.

Suppose we wished to increase the swapping probabilities to $0.4 + \epsilon$ for some small positive ϵ to allow slightly more exploration. Then the new effective click-through rate for advertiser 1 at the top rank during exploration would be $\hat{c}_1^1(0.6 - \epsilon) + \hat{c}_1^2(0.4 + \epsilon) = 0.16 - 0.1\epsilon$, while his effective click-through rate in the second rank would be $\hat{c}_1^1(0.4 + \epsilon) + \hat{c}_1^2(0.2 - 2\epsilon) = 0.1$. Thus the expected payoff to advertiser 1 in rank 1 is $(10 - 5)(0.16 - 0.1\epsilon) = 0.8 - 0.5\epsilon$, while his expected payoff in rank 2 is $(10 - 2)(0.1) = 0.8$, giving him incentive to deviate from his pre-exploration equilibrium bid. □

As the example suggests, the bounds in Theorem 2 are close to tight. In fact, the bounds can be made tight simply by replacing q_{max} with the conditional probability with which ad i would be selected if it were not in one of the top K ranks.

5 Learning Bounds

In this section, we bound the error of our estimated click-through rates for each advertiser in each slot after Q queries have been made on the given keyword. Let $n_{i,s}$ denote the number of times we have observed advertiser i in slot s, and let $z_{i,s,j}$ be the indicator random variable which is 1 if ad i is clicked the jth time it appears in slot s, and 0 otherwise. Finally, let $\pi_{i,s}^k$ be the probability that ad i is displayed at slot s when we are swapping ads that are k slots apart, as discussed in Section 4.

To simplify the presentation of results, we assume that the swapping distance k is drawn uniformly at random from $\{1, \cdots, K\}$ for each query, but the extension to arbitrary distributions is straight-forward.

Theorem 3. *Suppose the* `k-swap` *algorithm has been run for Q queries with a fixed set of broadcasted CTR estimates. Let $\hat{c}_i^s$ be our* new *estimate of CTR, defined as $\hat{c}_i^s = (1/n_{i,s}) \sum_{j=1}^{n_{i,s}} z_{i,s,j}$ for all advertisers i and slots s such that $n_{i,s} \geq 1$. Then for any $\delta \in (0, 1)$, with probability at least $1 - \delta$, the following holds for all i and s for which we have made at least one observation:*

$$|\hat{c}_i^s - c_i^s| \leq \sqrt{\frac{\ln(2KN/\delta)}{2n_{i,s}}}.$$

Furthermore, with probability at least $1 - \delta$, for all i and s, we have that $n_{i,s} \geq \max\{(Q/K) \sum_{k=1}^K \pi_{i,s}^k - \sqrt{Q \ln(2KN/\delta)/2}, 0\}$.

Proof. Let $x_1, \cdots, x_n$ be independent random variables, with $x_z = 1$ if a click is observed the zth time that ad i is displayed in slot s, and $x_z = 0$ otherwise. Clearly $E[(1/n) \sum_{i=1}^n x_i] = c_i^s$. Thus we can apply Hoeffding's inequality [7] to these random variables to bound the error of our estimation of the click-through rate of ad i in slot s given that we have observed the ad in this slot $n_{i,s}$ times. Specifically, for any $\delta' \in (0, 1)$, with probability $1 - \delta'$,

$$|\hat{c}_i^s - c_i^s| \leq \sqrt{\frac{\ln(2/\delta')}{2n_{i,s}}} \ .$$

We can apply Hoeffding's inequality once again to bound the deviation of $n_{i,s}$ from its expectation. We find that with probably $1 - \delta'$,

$$|n_{i,s} - E[n_{i,s}]| \leq \sqrt{Q \ln(2/\delta')/2} \ .$$

Setting $\delta = \delta'/(NK)$ and applying the union bound completes the proof. □

Thus as the number of queries Q grows, our estimates of the CTR vectors for each advertiser converge to the true CTR vectors.

6 Bounds on the Incentives of "Omniscient" Advertisers

If players have much more information about the actual click-through rates than the search engine, it is unlikely that we can entirely eliminate incentives of advertisers to change their bids during exploration. However, if we can bound the error in our estimates of the click-through rates, we can also bound how much advertisers can gain by deviating. When incentives to deviate are small, we may reasonably expect advertisers to maintain their equilibrium bids, since computing the new optimal bids can be costly. The search engine may further dull benefits from deviation by charging a small fee to advertisers when they change their bids.

From this point on, we assume that the error in search engine estimates of the CTRs is uniformly bounded by ϵ; that is, $|c_i^s - \hat{c}_i^s| \leq \epsilon$ for every i and s.

Assume that r_1^k were set such that the bidders have no incentive to change their bids if they use $\hat{c}_i^s$ as their CTR estimates. We now establish how much incentive they have to deviate if they know their *actual* CTR c_i^s, that is, $\tilde{c}_i^s = c_i^s$; we call such advertisers "omniscient".

Theorem 4. *The most that any omniscient advertiser can gain by deviating in expectation per impression is* $\max_{1 \leq i \leq K} 2\epsilon(v_i - p_i^K)$.

Proof. For all i and s,

$$\begin{aligned}\left|c_i^{'s} - \hat{c}_i^{'s}\right| &= \left|(1 - r_{s-k} - r_s)(c_i^s - \hat{c}_i^s) + r_{s-k}(c_i^{s-k} - \hat{c}_i^{s-k}) + r_s(c_i^{s+k} - \hat{c}_i^{s+k})\right| \\ &\leq (1 - r_{s-k} - r_s)\epsilon + r_{s-k}\epsilon + r_s\epsilon \ = \ \epsilon,\end{aligned}$$

and so $c_i^{'i} \geq \hat{c}_i^{'i} - \epsilon$. Now consider a player i deviating to slot (or rank) j:

$$\begin{aligned}c_i^{'i}(v_i - p_i^i) &\geq \hat{c}_i^{'i}(v_i - p_i^i) - \epsilon(v_i - p_i^i) \\ &\geq \hat{c}_i^{'j}(v_i - p_i^j) - \epsilon(v_i - p_i^i) \geq c_i^{'j}(v_i - p_i^j) - \epsilon(v_i - p_i^i) - \epsilon(v_i - p_i^j) \\ &\geq c_i^{'j}(v_i - p_i^j) - 2\epsilon(v_i - p_i^K).\end{aligned}$$

The theorem statement follows. □

This bound has the intuitive property that as our CTR estimates improve, the bound on incentives to deviate from equilibrium bids improves as well. (Note that given r_1^k the actual payoffs to deviation are not affected as we learn unless we also publicize the learned information.) It is also intuitive, however, that incentives diminish if the exploration probabilities fall. This motivates the following alternate bound which shows that we can make the incentives to deviate arbitrarily small even for omniscient advertisers by appropriately setting r_1^k.

Theorem 5. *The most that any omniscient advertiser can gain by deviating in expectation per impression is*

$$\max_{1 \leq i,j,k \leq K} \left\{ r_1^k \left(\alpha_i(\hat{c}_i^i - \hat{c}_i^{i+k}) + \alpha_{j-k}(\hat{c}_i^{j-k} - \hat{c}_i^j) + 2\epsilon(\alpha_i + \alpha_{j-k}) \right) (v_i - p_i^K) \right\}.$$

Proof. Consider incentives for some player i for deviating to slot j and let

$$\mu = r_1(\alpha_i(\hat{c}_i^i - \hat{c}_i^{i+k}) + \alpha_{j-k}(\hat{c}_i^{j-k} - \hat{c}_i^j) + 2\epsilon(\alpha_i + \alpha_{j-k}))(v_i - p_i^K).$$

Then,

$$
\begin{aligned}
&[(1-r_{i-k}-r_i)c_i^i + r_{i-k}c_i^{i-k} + r_i c_i^{i+k}](v_i - p_i^i) + \mu \\
&\geq [(1-r_{i-k}-r_i)c_i^i + r_{i-k}c_i^{i-k} + r_i c_i^{i+k}](v_i - p_i^i) + r_i(c_i^i - c_i^{i+k})(v_i - p_i^K) \\
&\quad + r_{j-k}(c_i^{j-k} - c_i^j)(v_i - p_i^K) \\
&\geq [(1-r_{i-k}-r_i)c_i^i(v_i - p_i^i) + r_{i-k}c_i^{i-k}(v_i - p_i^i) + r_i c_i^i(v_i - p_i^i) + r_{j-k}(c_i^{j-k} - c_i^j)(v_i - p_i^j) \\
&= c_i^i(v_i - p_i^i) + r_{i-k}(c_i^{i-k} - c_i^i)(v_i - p_i^i) + r_{j-k}(c_i^{j-k} - c_i^j)(v_i - p_i^j) \\
&\geq c_i^j(v_i - p_i^j) + r_{j-k}(c_i^{j-k} - c_i^j)(v_i - p_i^j) \\
&\geq c_i^j(v_i - p_i^j) + r_{j-k}(c_i^{j-k} - c_i^j)(v_i - p_i^j) - r_j(c_i^j - c_i^{j+k})(v_i - p_i^j) \\
&= [(1-r_{j-k}-r_j)c_i^j + r_{j-k}c_i^{j-k} + r_j c_i^{j+k}](v_i - p_i^j),
\end{aligned}
$$

where the inequalities follow from the assumption that $c_i^i(v_i - p_i^i) \geq c_i^j(v_i - p_i^j)$ and the fact that $p_i^s \geq p_i^K$ for any slot s. □

7 Bounds on Revenue Loss Due to Exploration

We now assume that the advertisers play according to the symmetric Nash equilibrium that was played prior to exploration and, as in the previous section, assume that the errors of the search engine's estimates of CTRs are uniformly bounded by ϵ with high probability. Given these assumptions, the theorem that follows bounds the loss in revenue due entirely to exploration.

Theorem 6. *The maximum expected loss to the search engine revenue per impression due to exploration is bounded by*

$$
\max_{1\leq k\leq K}\left\{ r_1^k \sum_{i=2}^{K} p_i^i \left(\alpha_i(\hat{c}_i^i - \hat{c}_i^{i+k}) - \alpha_{i-k}(\hat{c}_i^{i-k} - \hat{c}_i^i) + 2\epsilon\right)\right\}.
$$

Proof. Fix the swapping constant k. For any player i, the change in revenue can be bounded as

$$
\begin{aligned}
\sum_{i=1}^{K} c_i^i p_i^i - \sum_{i=1}^{K} {c'}_i^i p_i^i &= \sum_{i=1}^{K} p_i^i(c_i^i - {c'}_i^i) = \sum_{i=1}^{K} p_i^i(c_i^i - (1-r_{i-k}-r_i)c_i^i - r_{i-k}c_i^{i-k} - r_i c_i^{i+k}) \\
&= \sum_{i=1}^{K} p_i^i((r_{i-k}+r_i)c_i^i - r_{i-k}c_i^{i-k} - r_i c_i^{i+k}) \\
&= \sum_{i=1}^{K} p_i^i(r_i(c_i^i - c_i^{i+k}) - r_{i-k}(c_i^{i-k} - c_i^i)) \\
&\leq \sum_{i=1}^{K} p_i^i(r_i(\hat{c}_i^i - \hat{c}_i^{i+k}) - r_{i-k}(\hat{c}_i^{i-k} - \hat{c}_i^i) + 2\epsilon) \\
&= r_1 \sum_{i=1}^{K} p_i^i \left(\alpha_i(\hat{c}_i^i - \hat{c}_i^{i+k}) - \alpha_{i-k}(\hat{c}_i^{i-k} - \hat{c}_i^i) + 2\epsilon\right).
\end{aligned}
$$

The inequality in the penultimate step above follows by the bound on the error in CTRs, as well as by the fact that $r_i + r_{i-k} \leq 1$. Summing over all players and taking the max over all swapping distances yields the bound. □

8 Special Cases

In this section we study the problem of exploration while maintaining a pre-exploration symmetric Nash equilibrium in two special cases. In both cases, it is only necessary to swap adjacent pairs of ads in order to learn reasonable estimates of advertiser CTRs.

8.1 Factorable Click-Through Rates

The first special case we consider is the commonly studied setting where $c_i^s = e_i c_s$; that is, CTRs are factored into a product of advertiser relevance and slot-specific factors. To simplify presentation, we assume c_s is known and e_i is to be learned for all advertisers, since there are far more data for estimating c_s than e_i.

Maintaining Equilibrium Under these assumptions, using `k-swap` may seem strange; after all, we can learn e_i for all advertisers $i \leq K$ just as well by leaving them in their current slots! The only problem to be addressed then is to learn CTRs of losing bidders. Consequently, if we truly believe that CTRs are factorable, we need only do adjacent-ad swapping ($k = 1$) and can set $r_1 = \cdots = r_{K-1} = 0$ and only allow $r_K > 0$. In this case, we need not worry about deviations by advertisers in slots $1, \ldots, K-1$ to alternative slots $1, \ldots, K-1$, since the effective CTRs for these deviations are unchanged. Additionally, no advertiser wants to deviate to slot K, since the CTR in this slot is strictly lower than it was before exploration, and no advertiser ranked $K+1, \ldots, N$ wants a higher slot, since their effective CTRs increase. Thus we need only consider the incentives of the advertiser in slot K. It is not difficult to verify that the condition under which exploration does not affect advertiser K's incentives is

$$r_K \leq \min\left\{ \min_{1\leq j\leq K-1} \frac{c_K\left(v_K - p_K^K\right) - c_j\left(v_K - p_K^j\right)}{c_K(v_K - p_K^K)}, \frac{v_K - p_K^K}{v_K(q_{max}+1) - p_K^K} \right\},$$

and we can find an $r_K > 0$ when $c_K(v_K - p_K^K) > c_j(v_K - p_K^j)$ for $j < K$.

There is, however, another possible scenario in which exploration might be useful under the factorable CTR assumption. Suppose that we initially posit the factorable CTR model, but want to verify whether this is really the case. To do so, we can use adjacent-ad swapping to form multiple estimates of e_i using data from multiple adjacent slots. By comparing these estimates, we can vet our current model while also improving our CTR estimates for losing bidders.

Since CTR is factorable, our analysis need only consider the effective slot-specific CTRs, which we assume are known, $c'_s = (1 - r_{s-1} - r_s)c_s + r_{s-1}c_{s-1} + r_s c_{s+1}$. For all i, let

$$\alpha_i = \prod_{j=2}^{i} \frac{c_{j-1} - c_j}{c_j - c_{j+1}} \ .$$

By setting the swapping probabilities in this manner, the effective CTRs in slots $2, \cdots, K-1$ are unchanged when exploration is added. We can now simplify the bounds and characterization of Theorem 2. In particular, the precondition of the theorem and the second bound on r_1 need only to hold for $i = 1$. Furthermore, it can be shown that in the factorable setting, the precondition $(v_1 - p_1^1)c_1 > (v_1 - p_1^j)c_j$ always holds in the minimum revenue SNE [12] for generic valuations and relevances (see Appendix C for a formal statement and proof). Putting everything together, we can prove the following result in the factorable setting.

Theorem 7. *Consider the setting in which CTRs are factorable into the product of advertiser relevance and a slot-specific CTR factor. Let r_i be defined as in Equation 5 for all $i \in \{2, \cdots, K\}$. Assume that advertiser 1 strictly prefers his current slot to all others in equilibrium, i.e. the condition $\left(v_1 - p_1^1\right)c_1 > \left(v_1 - p_1^j\right)c_j$ holds for all $2 \leq j \leq K$ whenever $J_{1,j} > 0$. Then for generic valuations and relevances there exists an $r_1 > 0$*

such that no advertiser has an incentive to deviate from the pre-exploration SNE bids once exploration is added. Any r_1 for which the following conditions hold is sufficient:

$$r_1 \le \min \left\{ \min_{2 \le i \le K} \frac{1}{\alpha_i + \alpha_{i-1}}, \qquad \min_{1 \le i \le K; Z_i > 0} \frac{1}{Z_i}(v_i - p_i^i) c_i^i \right.$$
$$\left. \min_{2 \le j \le K; J_{1,j} > 0} \frac{1}{J_{1,j}} ((v_1 - p_1^1) c_1 - (v_1 - p_1^j) c_j). \right\}$$

where

$$J_{1,j} = \left(v_1 - p_1^1\right)(c_1 - c_2) + \left(v_1 - p_1^j\right)(\alpha_{j-1}(c_{j-1} - c_j) - \alpha_j(c_j - c_{j+1}))$$

and Z_i is defined according to Equation 2.

The proof of Theorem 7 is provided in Appendix D.

Learning Bounds As in the general setting, it is possible to derive learning bounds that show that as the number of observed queries grow, our estimates of the advertiser CTR vectors grow arbitrarily close to the true CTRs with high probability.

As before, let $n_{i,s}$ denote the number of times we have observed advertiser i in slot s at the current fixed CTR estimates, and let $z_{i,s,j}$ be a random binary variable indicating whether or not the ad i was clicked on the jth time it appeared in slot s.

Theorem 8. *Suppose that CTRs can be factored into advertiser-dependent and slot-dependent components. In other words, for all i and s, $c_i^s = e_i c_s$ where c_s is known. Suppose we have observed $n_{i,s}$ instances of ad i at slot s with a fixed set of broadcasted CTR estimates. Let $\hat{c}_i^s$ be our* new *estimate of CTR, defined as*

$$\hat{c}_i^s = \frac{c_s}{c_{s_i} n_{i,s_i}} \sum_{j=1}^{n_{i,s_i}} z_{i,s_i,j}$$

for all advertisers i and slots s, where $s_i = \text{argmax}_s\, c_s \sqrt{n_{i,s}}$. Then for any $\delta \in (0,1)$, with probability $1 - \delta$, for all i and s,

$$|\hat{c}_i^s - c_i^s| \le \frac{c_s}{c_{s_i}} \sqrt{\frac{\ln(2N/\delta)}{2n_{i,s_i}}} .$$

Proof. For each advertiser i, we base our estimate of click-through rate on data from the slot s_i maximizing $c_s \sqrt{n_s}$. Let

$$\hat{c}_i^{s_i} = \frac{1}{n_{i,s_i}} \sum_{j=1}^{n_{i,s_i}} z_{i,s_i,j}$$

be the estimate of click-through rate in this slot. By Hoeffding's inequality, for any $\delta' \in (0,1)$, with probability $1 - \delta'$,

$$|\hat{c}_i^{s_i} - c_i^{s_i}| \le \sqrt{\frac{\ln(2/\delta')}{2n_{i,s_i}}}.$$

Since for any s, $c_i^s = c_i^{s-i}(c_s/c_{s_i})$ and $\hat{c}_i^s = \hat{c}_i^{s-i}(c_s/c_{s_i})$, we thus have for all s

$$|\hat{c}_i^s - c_i^s| \le \frac{c_s}{c_{s_i}} \sqrt{\frac{\ln(2/\delta')}{2n_{i,s_i}}}.$$

We want this claim to hold for all N advertisers. Setting $\delta' = \delta/N$ and applying the union bound completes the proof. ◻

8.2 Click-through Rates with Constant Slot Ratios

Next we consider adjacent-ad swapping ($k = 1$) for the case in which for each player i, the click-through rates have constant ratios for adjacent slots. That is, for all i and all $1 \leq s \leq K-1$, we assume that $c_i^{s+1}/c_i^s = \gamma_i \leq 1$ where γ_i is advertiser-dependent and unknown. Let $\hat{\gamma}_i$ denote the search engine estimate of γ_i and suppose as before that advertisers use these as their own estimates. Let $\alpha_j = 1$ for every $j \in \{2, \ldots, K-1\}$, so $r_1 = r_2 = \cdots = r_{K-1}$. Additionally, let $\alpha_K = \min\{(\hat{\gamma}_i - 1)^2/q_{max}, 1\}$.

Maintaining Equilibrium As in the previous section, we can considerably simplify the bounds and characterization of Theorem 2 in this special case. In particular, the first and second bounds on r_1 must hold, but the third bound on r_1 and the precondition need only to hold for $i = 1$ and $i = K$.

Theorem 9. *Suppose that CTRs are of the form $c_i^s = e_i(\hat{\gamma}_i)^{s-1}$ for all i and s. Assume that advertisers $i = 1, \cdots, K$ strictly prefer their current slots to all others in equilibrium, i.e. the following condition holds for $i = 1, K$ and all $j \in \{1, \cdots, K\}, j \neq i$ whenever $J_{i,j} > 0$:*

$$(v_i - p_i^i)\hat{c}_i^i > (v_i - p_i^j)\hat{c}_i^j.$$

Furthermore, assume that for all i such that $Z_i > 0$, $v_i - p_i^i > 0$. Then for generic valuations and relevances there exists $r_1 > 0$ such that no advertiser has an incentive to deviate from the pre-exploration symmetric Nash equilibrium bids once exploration is added. Any r_1 satisfying the following set of conditions is sufficient:

$$r_1 \leq \min \left\{ \frac{1}{2}, \quad \min_{i=1,K; Z_i > 0} \frac{\hat{c}_i^1}{Z_i}(v_i - p_i^i)(\hat{\gamma}_i)^{i-1}, \right.$$
$$\left. \min_{i=1,K; 1 \leq j \leq K; J_{i,j} > 0} \frac{\hat{c}_i^1}{J_{i,j}} \left((v_i - p_i^i)(\hat{\gamma}_i)^{i-1} - (v_i - p_i^j)(\hat{\gamma}_i)^{j-1} \right) \right\}$$

Appendix E contains the proof of this theorem.

Learning Bounds for Ad-Dependent Constant Slot Ratios We can also prove analogous learning bounds in this setting and show that it is only necessary to explore via adjacent-ad swapping in order to obtain CTR estimates for all advertisers at all slots. Again, let $n_{i,s}$ denote the number of times we have observed advertiser i in slot s at the current fixed CTR estimates, and let $z_{i,s,j}$ be a random variable indicating whether or not the ad i was clicked on the jth time it appeared in slot s.

Theorem 10. *Suppose that CTRs are of the form $c_i^s = e_i(\gamma_i)^{s-1}$ for all i and s, where γ_i is unknown and $e_i = c_i^1$. Suppose we have observed $n_{i,s}$ instances of ad i at slot s with a fixed set of broadcasted CTR estimates. Let $\hat{c}_i^s$ be our* new *estimate of CTR, defined as:*

$$\hat{c}_i^s = \frac{(\hat{\gamma}_i)^{s-s_i}}{n_{i,s_i}} \sum_{j=1}^{n_{i,s_i}} z_{i,s_i,j}$$

where

$$\hat{\gamma}_i = \frac{(1/n_{i,s_i+1}) \sum_{j=1}^{n_{i,s_i+1}} z_{i,s_i+1,j}}{(1/n_{i,s_i}) \sum_{j=1}^{n_{i,s_i}} z_{i,s_i,j}} .$$

Then for any $\delta \in (0,1)$, with probability $1 - \delta$, for all advertisers i and slots s,

$$|\hat{c}_i^s - c_i^s| \leq \left(s(s+1) + s_i^2 + 1\right) \sqrt{\frac{\ln(4N/\delta)}{n_i}} + \frac{(s(s+1) + s_i^2)\ln(4N/\delta)}{\hat{c}_i^{s_i} n_i} ,$$

where $n_i = \min\{n_{i,s_i}, n_{i,s_i+1}\}$.

The proof is somewhat more involved than the proofs of the analogous Theorems 3 and 8, and appears in Appendix F.

9 Dealing with Continuous Context Information

A keyword entered by a user is just one of many characteristics of the broad *context* in which advertisements are displayed. In general, the CTR of an ad could depend heavily on other contextual information, such as the searcher's gender or location or the time of day [4, 11]. Our approach is directly applicable in settings in which the user may submit separate bids for each context [3]. However, it is not clear whether these results apply in general when the context contains features with very large or even continuous domains, or when bids cannot be conditioned on context. Since the presence of such features is quite common, we now briefly explain how parts of our analysis can be generalized to this setting.

For the remainder of the discussion, we will focus on the scenario in which the subspace of contextual features is continuous, and assume that advertisers are not able to condition their bids on context. We denote the specific *realized* context by x and let X be the set of all possible contexts that can arise. We assume that the only effect that a particular context x has is on the click-through rates and, consequently, make these functions of context for all slots and players. Thus, $c_i^s(x)$ now denotes the click-through rate of player i in slot s when context is realized to x. As before, we assume that for each player i for each x, $c_i^s(x) > c_i^t(x)$ whenever $1 \le s < t \le K$, and define $c_i^s(x) = 0$ for $s > K$. For the moment, we do not commit to any specific ranking policy that the search engine will follow, but note rather that the effect of the ranking policy will be localized to the expected CTRs and prices that the bidders will actually face. Given the particular ranking of advertisers, an advertiser i will pay some price, $p_i^s(x)$ for appearing in (or switching to) slot s, and the expected price (with respect to the distribution over context) of bidder i induced by the ranking policy is denoted by $p_i(x, b)$.

Let b denote the entire profile of advertiser bids and b_{-i} the profile of bids of all players other than i. The expected utility of the advertiser i can then be written as

$$\begin{aligned} U_i(b) &= E_{x\sim D}[u_i(x,b)] = E_{x\sim D}[c_i(x,b)(v_i - p_i(x,b))] \\ &= v_i E_{x\sim D}[c_i(x,b)] - E_{x\sim D}[c_i(x,b)p_i(x,b)], \end{aligned}$$

where D is the distribution of context, $c_i(x, b)$ is the CTR of player i under context x given the ranking induced by $w_i(x)$ and the configuration of all bids b, and, similarly, $p_i(x, b)$ is the price per click paid by i under x given bids b and the induced ranking.

To guide exploration in the setting with context, we introduce the exploration function, $r_1(x)$, which is the probability that the bidder in slot 1 will be swapped down, analogous to the scheme we proposed above. To guide the exploration parameters of the remaining slots, we can use the fixed multipliers α_s exactly in the same manner as we have done until now: that is, we let $r_s(x) = \alpha_s r_1(x)$. Thus, we have only one parameter to control throughout exploration, albeit this "parameter" can now be a function of context. As before, we use $c_i^{'s}(x)$ to denote the effective CTR during exploration, which can be expressed as

$$c_i^{'s}(x) = (1 - r_{s-k}(x) - r_s(x))c_i^s(x) + r_{s-k}(x)c_i^{s-k} + r_s(x)c_i^{s+k}.$$

It is easy to see that context changes nothing in terms of maintaining equilibria during exploration: our scheme above can be applied directly.[5] What it complicates is the analysis of bounds on the incentives to deviate of omniscient advertisers as well as the analysis of revenue loss due to exploration. We now address these questions to a limited degree.

9.1 Bound on Incentives of "Omniscient" Advertisers

First, we present a bound on the incentives that advertisers may have to deviate if they know their CTRs exactly. Fix b_{-i} (e.g., to equilibrium b_{-i}^*) and suppose that the bidders were in equilibrium before exploration.

[5] One qualification that must be added is that it is as yet unknown whether an equilibrium of a generalized second-price auction always exists in the continuous context setting if bids cannot be conditioned on context.

Theorem 11. *Suppose* $|E_{x\sim D}[\hat{c}_i^s(x)] - E_{x\sim D}[c_i^s(x)]| \leq \epsilon$ *for all* i, s. *Furthermore, suppose that there is* $\bar{p} < \infty$ *such that* $E_{x\sim D}[p_i(x,b)] \leq \bar{p}$ *for all* $b_i \in \mathbb{R}$ *and the players are omniscient. Then any player will gain at most* $4\bar{r}\bar{\alpha}(v_i E_{x\sim D}[\hat{c}_i^1(x)] + \bar{p} + \epsilon)$ *by deviating from a pre-exploration equilibrium, where* $\bar{r} = \sup_{x\in X} r_1(x)$ *and* $\bar{\alpha} = \max_s \alpha_s$.

The proof of this theorem can be found in Appendix G. To gain some intuition about the bound, note that $v_i E_{x\sim D}[\hat{c}_i^1(x)]$ is the expected value that advertiser would get by appearing in the first slot, providing the strongest incentive for advertisers to deviate. The other reason for advertisers to deviate is to reduce the payments; hence, the payment upper bound, $\bar{p}$, in the expression. Finally, the more accurate the advertisers' information about CTRs (as quantified by ϵ in the expression), the less effective our scheme of maintaining equilibria will be.

As we can see, the conditions in Theorem 11 are very strong. A natural question then arises: Are there reasonable settings in which the bound would be of value? In the next section, we demonstrate that under relatively mild assumptions, the key condition that $E_{x\sim D}[p_i(x,b)] \leq \bar{p}$ is satisfied in generalized second-price auctions under the ranking rules of the kind we considered throughout this paper.

Weighted Ranking Rules in Generalized Second-Price Auctions Suppose that for any context $x \in X$ the bidders are ranked by the product of their weight, $w_i(x)$, and bid b_i. If the bidder i is ranked in slot s, then the generalized second-price auction rules will charge the price of

$$p_i(x,b) = \frac{b_{s-1}w_{s-1}(x)}{w_i(x)}.$$

We are interested in determining whether the expectation of $p_i(x,b)$ is finite for every bidder. That there is indeed such an upper bound should be rather intuitive: as long as $p_i(x,b)$ attains a maximum on the context domain X, this maximum can serve as the desired upper bound. A set of sufficient conditions for the existence of this upper bound is the subject of the following theorem.

Theorem 12. *Suppose that the following conditions are satisfied:*

1. $w_i(x)$ *are continuous on* X *for every advertiser* i,
2. $w_i(x) > 0$ *for all* $x \in X$ *and for every* i,
3. X *is compact.*

Furthermore, suppose that b_{-i} *is fixed. Then there exists* $\bar{p} < \infty$ *such that* $E_x[p_i(x,b)] \leq \bar{p}$ *for all* b_i *and for every advertiser* i.

Proof. To begin, observe that

$$p_i(x,b) = \frac{b_{s-1}w_{s-1}(x)}{w_i(x)} \leq \frac{\max_{t\neq i} b_t w_t(x)}{w_i(x)} \leq \max_{t\neq i} b_t \frac{\max_{t\neq i} w_t(x)}{w_i(x)}.$$

Consequently,

$$E_x[p_i(x,b)] \leq \max_{t\neq i} b_t \left[\frac{\max_{t\neq i} w_t(x)}{w_i(x)}\right].$$

Since $w_t(x)$ is continuous for all t, so is $w_{max}(x) = \max_{t\neq i} w_t(x)$. Furthermore, since $w_i(x) > 0$ for all $x \in X$, the ratio $w_{ratio}(x) = w_{max}(x)/w_i(x)$ is also continuous on X. Now, since X is compact and $w_{ratio}(x)$ is continuous on X, there exists $W = \max_{x\in X} w_{ratio}(x) < \infty$ by the Weierstrass theorem. Thus, we can bound the expected payment of bidder i:

$$E_x[p_i(x,b)] \leq \max_{t\neq i} b_t E_x[W] = W \max_{t\neq i} b_t.$$

By setting $\bar{p} = W \max_{t\neq i} b_t$ and observing that it does not depend on b_i, we obtain the desired result. ◻

9.2 Bound on Expected Revenue Loss Due to Exploration

A final question we address in the continuous context setting is the bound on how much revenue will be lost by the search engine while the exploration is taking place. For this bound, we assume as above that the bidders maintain their equilibrium bids b^* during exploration. If not, revenue loss can be quite unpredictable, since it is unclear in what ways the bidders will change their bids once out of equilibrium (and a new equilibrium adjustment process, even if convergent, may be quite long).

Recall that N is the total number of advertisers bidding on the keyword.

Theorem 13. *Suppose that the players maintain their equilibrium bids b^* during exploration. Then revenue loss due to exploration is bounded by*

$$2\bar{r}\bar{\alpha}\sum_{i=1}^{N} E_{x\sim D}[p_i(x,b^*)] \ ,$$

where $\bar{r} = \max_{x\in X} r_1(x)$ and $\bar{\alpha} = \max_s \alpha_s$.

Proof. Let $\phi(s)(x,b^*)$ denote the player that receives slot s when context is x and let us partition the context space as $X = X_1^s \cup \cdots \cup X_N^s$, where X_i^s is the set of realizations of x such that advertiser i gets slot s. Then

$$\begin{aligned}
\Delta_{Rev} &= \sum_{s=1}^{K} E_x\left[p^s_{\phi(s)(x,b^*)}(x)(c^s_{\phi(s)(x,b^*)}(x) - c'^{s}_{\phi(s)(x,b^*)}(x))\right] \\
&= \sum_{s=1}^{K} E_x\Big[p^s_{\phi(s)(x,b^*)}(x) r_s(x)(c^s_{\phi(s)(x,b^*)}(x) - c^{s+k}_{\phi(s)(x,b^*)}(x)) \\
&\qquad\qquad -r_{s-k}(x)(c^{s-k}_{\phi(s)(x,b^*)}(x) - c^s_{\phi(s)(x,b^*)}(x))\Big] \\
&\le \sum_{s=1}^{K} E_x\left[p^s_{\phi(s)(x,b^*)}(x)(r_s(x)c^s_{\phi(s)(x,b^*)}(x) + r_{s-k}(x)c^s_{\phi(s)(x,b^*)}(x))\right] \\
&\le 2\bar{r}\bar{\alpha}\sum_{s=1}^{K} E_x\left[p^s_{\phi(s)(x,b^*)}(x)c^s_{\phi(s)(x,b^*)}(x)\right] \\
&\le 2\bar{r}\bar{\alpha}\sum_{s=1}^{K} E_x\left[p^s_{\phi(s)(x,b^*)}(x)\right] \\
&= 2\bar{r}\bar{\alpha}\sum_{s=1}^{K}\sum_{i=1}^{N}\int_{X_i^s} p_i^s(x)dF(x) = 2\bar{r}\bar{\alpha}\sum_{i=1}^{N}\sum_{s=1}^{K}\int_{X_i^s} p_i^s(x)dF(x) \\
&= 2\bar{r}\bar{\alpha}\sum_{i=1}^{N} E_x[p_i(x,b^*)].
\end{aligned}$$

The third line above yields an upper bound by simply removing the negative terms from the expression. The fourth uses the fact that $\bar{r}\bar{\alpha}$ is the upper bound on the exploration probability. The fifth step uses the fact that CTR is just a probability of a click and is therefore at most one. □

10 Conclusion

We introduced an exploration scheme that allows search engines to learn click-through rates for advertisements. We showed how, when possible, to set the exploration parameters in order to eliminate the incentives for advertisers to deviate from a pre-exploration symmetric Nash equilibrium. In situations in which we cannot entirely eliminate incentives to change bids, we can make the gain for changing bids small enough to ensure that bid manipulation is hardly worth advertisers' efforts. Finally, we derived a bound on worst-case expected per-impression revenue loss due to exploration. Since this loss is zero in the limit of no exploration, we can set exploration parameters in order to make it arbitrarily small, while still ensuring that we eventually learn click-through rates.

The analysis in this paper leaves open a number of interesting questions. For example, we do not address methods of setting the `k-swap` parameters in order to optimize various trade-offs between revenue loss and information gain. Additionally, the analysis of the continuous-context setting presented in Section 9 is far from complete. As it is becoming more and more clear that contextual information can significantly effect click-through rate, a complete analysis of exploration in this setting would be of great value.

Acknowledgments

The authors are grateful to Yiling Chen and Alexander Strehl for insightful comments on early drafts of this paper, and to David Pennock for useful discussions.

References

1. Tilman Börgers, Ingemar Cox, Martin Pesendorfer, and Vaclav Petricek. Equilibrium bids in auctions of sponsored links: theory and evidence. Technical report, University of Michigan, 2007.
2. Benjamin Edelman, Michael Ostrovsky, and Michael Schwarz. Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords. *American Economic Review*, 9(1):242–259, March 2007.
3. Eyal Even-Dar, Michael Kearns, and Jennifer Wortman. Sponsored search with contexts. In *the 3rd International Workshop on Internet and Network Economics*, 2007.
4. Arpita Ghosh, Hamid Nazerzadeh, and Mukund Sundararajan. Computing optimal bundles for sponsored search. In *the 3rd International Workshop on Internet and Network Economics*, 2007.
5. Rica Gonen and Elan Pavlov. An incentive-compatible multi-armed bandit mechanism. In *the 26th Annual ACM Symposium on Principles of Distributed Computing*, 2007.
6. James Heckman. Sample selection bias as a specification error. *Econometrica*, 47:153–161, 1979.
7. Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.
8. Sebastien Lahaie. An analysis of alternative slot auction designs for sponsored search. In *ACM Conference on Electronic Commerce*, 2006.
9. Sebastien Lahaie and David M. Pennock. Revenue analysis of a family of ranking rules for keyword auctions. In *ACM Conference on Electronic Commerce*, 2007.
10. Sandeep Pandey and Christopher Olston. Handling advertisements of unknown quality in search advertising. In *Advances in Neural Information and Processing Systems 19*, pages 1065–1072, 2007.
11. David Parkes and Tuomas Sandholm. Optimize-and-dispatch architecture for expressive ad auctions. Paper presented at the First Workshop on Sponsored Search Auctions, Vancouver, Canada, June, 2005.
12. Hal Varian. Position auctions. *International Journal of Industrial Organization*, 25(6):1163–1178, 2007.
13. Jennifer Wortman, Yevgeniy Vorobeychik, Lihong Li, and John Langford. Maintaining equilibria during exploration in sponsored search auctions. In *the 3rd International Workshop on Internet and Network Economics*, 2007.

Appendix

A Proof of Theorem 1

Define modified valuations of all bidders to be $v'_i = w_i v_i$ and the corresponding bids of all players to be $b'_i = w_i b_i$. With these modified bids and valuations, there is a symmetric Nash equilibrium in the resulting sponsored search auction game, which, by definition, has the property that

$$\tilde{c}_i^i(v'_i - p'_i) \geq \tilde{c}_i^s(v'_i - p'_s) \ \forall s \neq i,$$

where $p'_s = b'_{s+1}$. To reverse our initial transformation, divide both sides by w_i to obtain

$$\tilde{c}_i^i(v_i - p_i^i) \geq \tilde{c}_i^s(v_i - p_i^s) \; \forall s \neq i,$$

which is the definition of SNE in our context.

B Proof of Theorem 2

In order to prove Theorem 2, we must make sure that when the conditions stated in the theorem hold, no advertiser is happier changing his bid and moving to a different position in the ranking. We break this proof into multiple parts, summarized by the following series of lemmas. Combining the pieces yields a set of conditions that show that no advertiser has incentive to deviate when exploration is added.

The first condition,

$$r_1 \leq \min_{2 \leq i \leq K} \frac{1}{\alpha_i + \alpha_{i-k}},$$

ensures that $r_i + r_{i-k} \leq 1$ for all i. The other two conditions are the subject of Lemmas 1 and 2 below.

Incentives of $1, \cdots, K$ to Switch to Alternate Ranks in $1, \cdots, K$

We first verify that players $i \in \{1, \cdots, K\}$ do not want to switch to alternate ranks $j \in \{1, \cdots, K\}$. The following lemma gives conditions to guarantee this. Recall that we define $c_1^0 = 0$ and $\alpha_0 = 0$.

Lemma 1. *Assume that the condition*

$$(v_i - p_i^i)\hat{c}_i^i > (v_i - p_i^j)\hat{c}_i^j,$$

holds for every $i, j \in \{1, \cdots, K\}$ in equilibrium. Suppose that prior to exploration advertisers' bids are in a symmetric Nash equilibrium. Then there is $r_1 > 0$ such that players in slots $1, \cdots, K$ do not wish to switch to other slots in $1, \cdots, K$ as long as

$$r_1 \leq \min_{1 \leq i,j \leq K; J_{i,j} > 0} \frac{1}{J_{i,j}} \left((v_i - p_i^i)\hat{c}_i^i - (v_i - p_i^j)\hat{c}_i^j \right)$$

where $J_{i,j}$ is as defined in Equation 1.

Proof. To ensure that players in slots $1, \cdots, K$ do not have incentive to switch to other slots within this range, we need the condition

$$\begin{aligned} &\left((1 - r_{i-k} - r_i)\hat{c}_i^i + r_{i-k}\hat{c}_i^{i-k} + r_i\hat{c}_i^{i+k}\right)(v_i - p_i^i) \\ &\geq \left((1 - r_{j-k} - r_j)\hat{c}_i^j + r_{j-k}\hat{c}_i^{j-k} + r_j\hat{c}_i^{j+k}\right)(v_i - p_i^j) \end{aligned}$$

to be satisfied for any $i, j \in \{1, \cdots, K\}$. Using the definition $r_i = \alpha_i r_1$ and rearranging terms, we obtain the equivalent condition

$$r_1 J_{i,j} \leq (v_i - p_i^i)\hat{c}_i^i - (v_i - p_i^j)\hat{c}_i^j. \tag{3}$$

Under our assumption, the right hand side of Equation 3 is always strictly positive. When $J_{i,j} \leq 0$, any $r_1 \in [0, 1]$ satisfies the condition we need. When $J_{i,j} > 0$, we need to guarantee

$$r_1 \leq \frac{(v_i - p_i^i)\hat{c}_i^i - (v_i - p_i^j)\hat{c}_i^j}{J_{i,j}}.$$

Therefore, we must have for all $i, j \in \{1, \cdots, K\}$ such that $J_{i,j} > 0$

$$r_1 \leq \frac{1}{J_{i,j}} \left((v_i - p_i^i)\hat{c}_i^i - (v_i - p_i^j)\hat{c}_i^j \right).$$

□

Incentive of Players $1, \ldots, K$ to Move to Ranks $K+1, \ldots, N$

The previous lemma showed that the K highest ranked advertisers do not have incentive to deviate to another rank between 1 and K when exploration is added under certain conditions. We now consider the conditions that guarantee that these advertisers do not want to move to a rank greater than K.

Lemma 2. *If the players are in a symmetric Nash equilibrium with $v_i > p_i^i$ whenever $Z_i > 0$ before exploration and if*

$$r_1 \leq \min_{1 \leq i \leq K; Z_i > 0} \frac{1}{Z_i} \left(\hat{c}_i^i (v_i - p_i^i) \right)$$

where Z_i is defined as in Equation 2, then the K highest ranking advertisers have no incentive to switch to any slot below K for generic valuations and relevances. Furthermore, it is always possible to set r_1 in such a way such that the above holds and $r_1 > 0$.

Proof. We show that any player $i \in \{1, \cdots, K\}$ has no incentive to deviate to any slot $j \in \{K+1, \cdots, N\}$. For any $j \in \{K+1, \cdots, N\}$, in order to guarantee that player i does not want to switch to slot j, the following condition must hold:

$$\left((1 - r_{i-k} - r_i) \hat{c}_i^i + r_{i-k} \hat{c}_i^{i-k} + r_i \hat{c}_i^{i+k} \right) (v_i - p_i^i) \geq r_K q_j \hat{c}_i^{K-k+1} v_i.$$

As we require that this condition hold simultaneously for all j, it becomes equivalent to

$$r_1 Z_i \leq \hat{c}_i^i (v_i - p_i^i) \tag{4}$$

Since $v_i - p_i^i > 0$, the right hand side of Equation 4 is strictly positive. When $Z_i \leq 0$, the constraint is satisfied trivially. Thus we need only that when $Z_i > 0$,

$$r_1 \leq \frac{1}{Z_i} \left(\hat{c}_i^i (v_i - p_i^i) \right)$$

□

Incentives of Players $K+1, \ldots, N$ to Move to Slots $1, \ldots, K$

Lemma 3. *No advertiser in slots $K+1, \cdots, N$ prefers to deviate and move to any slot in $1, \cdots, K$ when exploration is added.*

Proof. Since the set of advertisers $K+1, \cdots, N$ received no clicks in the pre-exploration symmetric Nash equilibrium, we know that it must be the case that for all i and j such that $K+1 \leq i \leq N$ and $1 \leq j \leq K$,

$$v_i - p_i^j \leq 0.$$

In other words, these advertisers have a value per click that is lower than the equilibrium price per click of any slot between 1 and K. When exploration is added, the CTRs of these slots might change, but the price is still higher than these advertisers' values. Thus the advertisers still are not interested in these slots when exploration is added. □

Note that we have not addressed the incentives for players $K+1, \cdots, N$ to deviate to other ranks $K+1$ or higher. This case will depend on the way in which the distribution q_i is chosen. There are many possible choices for which players $K+1, \cdots, N$ will not have incentive to swap ranks among themselves.

C Proof of Player 1's Strict Preference at SNE

The following theorem is briefly mentioned in Section 8.1. Here we state the result formally and provide a very simple proof.

Theorem 14. *Suppose that the players are playing a minimum symmetric Nash equilibrium. Then for generic valuations and relevances* $(v_1 - p_1^1)c_1 > (v_1 - p_1^j)c_j$.

Proof.

$$\begin{aligned} c_1 w_2 b_2 - c_j w_{j+1} b_{j+1} &= \sum_{t=1}^{K}(c_t - c_{t+1})w_{t+1}v_{t+1} - \sum_{t=j}^{K}(c_t - c_{t+1})w_{t+1}v_{t+1} \\ &= \sum_{t=1}^{j-1}(c_t - c_{t+1})w_{t+1}v_{t+1} \\ &\leq w_2 v_2 \sum_{t=1}^{j-1}(c_t - c_{t+1}) = w_2 v_2 (c_1 - c_j). \end{aligned}$$

For generic valuations and relevances, $w_2v_2(c_1 - c_s) < w_1v_1(c_1 - c_s)$ and, consequently, $c_1w_2b_2 - c_jw_{j+1}b_{j+1} < w_1v_1c_1 - w_1v_1c_j$ for every $2 \leq j \leq K$. Rewriting, we get $c_1(w_1v_1 - w_2b_2) > c_j(w_1v_1 - w_{j+1}b_{j+1})$ and we recover the desired strict inequality. □

D Proof of Theorem 7

To begin, note that by defining α_i to be

$$\alpha_i = \prod_{j=2}^{i}\left(\frac{c_{j-1} - c_j}{c_j - c_{j+1}}\right),$$

we obtain the following recursive expression for r_i:

$$r_i = r_{i-1}\left(\frac{c_{i-1} - c_i}{c_i - c_{k+1}}\right) \tag{5}$$

for all $i \in \{2, \cdots, K\}$. This setting of probabilities is convenient because the CTRs of an ad in slots $2, \cdots, K$ do not change when exploration is added if bids do not change, as shown in the following useful lemma.

Lemma 4. *When r_i are computed recursively by Equation 5, then $c'_i = c_i$ for all $i \in \{2, \cdots, K\}$.*

Proof. For any $i \in \{2, \cdots, K\}$,

$$\begin{aligned} c'_i &= (1 - r_{i-1} - r_i)c_i + r_{i-1}c_{i-1} + r_i c_{i+1} \\ &= c_i + r_{i-1}(c_{i-1} - c_i) - r_i(c_i - c_{i+1}) \\ &= c_i + r_{i-1}(c_{i-1} - c_i) - r_{i-1}\left(\frac{c_{i-1} - c_i}{c_i - c_{i+1}}\right)(c_i - c_{i+1}) = c_i. \end{aligned}$$

□

The next lemma provides a simple and intuitive result for generic valuations and relevances.

Lemma 5. *In a symmetric Nash equilibrium with $w_iv_i \geq w_{i+1}v_{i+1}$, $w_iv_i > w_{i+1}b_{i+1}$ for $1 \leq i \leq K$ for generic valuations and relevances.*

Proof. Suppose $w_i v_i \leq w_{i+1} b_{i+1}$ for some $i \in \{1, 2, \cdots, K\}$. For $j > 1$, the following upper bound on $w_j b_j$ holds in a symmetric Nash equilibrium (a simple extension of Varian [12, 9]):

$$w_j b_j \leq w_{j-1} v_{j-1} (1 - \beta_j) + w_{j+1} b_{j+1} \beta_j \tag{6}$$

where $\beta_j = c_j / c_{j-1} < 1$. From the above bound with $j = i + 1$ and the assumption that $w_i v_i \leq w_{i+1} b_{i+1}$, it follows that

$$w_i v_i \leq w_i v_i (1 - \beta_{i+1}) + w_{i+2} b_{i+2} \beta_{i+1}$$

and, therefore $w_i v_i \leq w_{i+2} b_{i+2}$. Applying the bound in Equation 6 again with $j = i + 2$, we see that

$$w_{i+2} b_{i+2} \leq w_{i+1} v_{i+1} (1 - \beta_{i+2}) + w_{i+3} b_{i+3} \beta_{i+2} \ .$$

When CTRs are factorable, $w_i v_i \geq w_{i+1} v_{i+1}$ [1, 2, 12]. Since for generic values and relevances, $w_{i+1} v_{i+1} < w_i v_i$, we have $w_i v_i \beta_{i+2} < w_{i+3} b_{i+3} \beta_{i+2}$ and, consequently, $w_i v_i < w_{i+3} b_{i+3} \leq w_{i+1} b_{i+1}$. Thus, $w_i v_i < w_{i+1} b_{i+1}$ and

$$v_i - \frac{w_{i+1} b_{i+1}}{w_i} < 0 \ .$$

But this is a contradiction, since player i would then want to switch to slot $K + 1$, whereas we assumed that all bidders were in a Nash equilibrium. ☐

Thus, $v_i > p_i^i$ for all players i when CTRs are factorable.

Given Lemma 4, note that if none of the bidders in slots $2, \cdots, N$ wanted to move up to slot 1 in equilibrium before exploration, they have even less incentive to do so once exploration is added since the effective CTR for slot 1 is now lower and the effective CTR of their own slots is the same by our definition of α_i. Furthermore, none of the bidders in slots $2, \cdots, N$ want to switch to alternate slots in $2, \cdots, K$ since the effective CTRs are now the same for all of these slots and do not depend on the bidder's identity due to the factorization assumption. Consequently, we need only examine whether or not the top bidder wants to move down, or whether any bidder might like to move into a slot below K. The analysis of these cases is directly analogous to the analysis in the proof of Theorem 2, and the sufficient conditions on r_1 are derived in the same manner.

E Proof of Theorem 9

This proof is very similar to the proofs of Theorem 2 and 7. In fact, the incentives for player 1 and players $K, K+1, \cdots, N$ can be analyzed exactly as in the proof of Theorem 2, and the conditions on r_1 follow from this analysis as before. Thus here we focus only on the incentives of players $2, \cdots, K - 1$.

Incentives of Players $2, \ldots, K-1$ to Move to Slots $1, \ldots, K$

Lemma 6. *In the constant ratio setting, when $r_i = r_1$ for $i \in 2, \cdots, K$, players $2, \ldots, K-1$ do not have incentive to deviate to other slots $1, \ldots, K$ during adjacent-ad swapping exploration for any r_1.*

Proof. To ensure that players in slots $2, \cdots, K-1$ do not have incentive to switch to other slots within this range, we need the condition

$$\begin{aligned} &\left((1 - r_{i-1} - r_i)\hat{c}_i^i + r_{i-1}\hat{c}_i^{i-1} + r_i \hat{c}_i^{i+1}\right)(v_i - p_i^i) \\ &\geq \left((1 - r_{j-1} - r_j)\hat{c}_i^j + r_{j-1}\hat{c}_i^{j-1} + r_j \hat{c}_i^{j+1}\right)(v_i - p_i^j) \end{aligned}$$

to be satisfied for any $i, j \in \{1, \cdots, K\}$. Setting $r_i = r_1$ for all i and plugging in $\hat{c}_i^{s-1} = \hat{c}_i^s / \hat{\gamma}_i$ and $\hat{c}_i^{s+1} = \hat{c}_i^s \hat{\gamma}_i$, we get

$$\begin{aligned} (1 - r_{s-1} - r_s)\hat{c}_i^s + r_{s-1}\hat{c}_i^{s-1} + r_s \hat{c}_i^{s+1} &= \hat{c}_i^s + \left(\hat{\gamma}_i - 2 + \frac{1}{\hat{\gamma}_i}\right) r_1 \hat{c}_i^s \\ &= \left(1 + \frac{r_1 (\hat{\gamma}_i - 1)^2}{\hat{\gamma}_i}\right) \hat{c}_i^s \end{aligned}$$

for every player i and for all $s = 2, \ldots, K-1$. Thus, the condition that needs to be satisfied is

$$\hat{c}_i^i \left(1 + \frac{r_1(\hat{\gamma}_i - 1)^2}{\hat{\gamma}_i}\right)(v_i - p_i^i) \geq \hat{c}_i^j \left(1 + \frac{r_1(\hat{\gamma}_i - 1)^2}{\hat{\gamma}_i}\right)(v_i - p_i^j)$$

which is equivalent to the symmetric equilibrium condition before exploration and thus holds by assumption. For deviation to slot 1, note that

$$1 + \frac{r_1(\hat{\gamma}_i - 1)^2}{\hat{\gamma}_i} \geq 1,$$

and thus $\hat{c}_i^{'i} \geq \hat{c}_i^i$ for every $i = 2, \ldots, K-1$, whereas $\hat{c}_1^{'1} \leq \hat{c}_1^1$. Consequently, if deviations to 1 were unprofitable prior to exploration, they are certainly still unprofitable with exploration. Finally, for deviations to slot K, note that the effective CTR for slot K is $\hat{c}_i^K(1 + r_1(\hat{\gamma}_i - 2)) \leq \hat{c}_i^K(1 + r_1(\hat{\gamma}_i + 1/\hat{\gamma}_i - 2))$. Thus,

$$\begin{aligned}\hat{c}_i^i \left(1 + \frac{r_1(\hat{\gamma}_i - 1)^2}{\hat{\gamma}_i}\right)(v_i - p_i^i) &\geq \hat{c}_i^K \left(1 + \frac{r_1(\hat{\gamma}_i - 1)^2}{\hat{\gamma}_i}\right)(v_i - p_i^j)\\ &\geq \hat{c}_i^K \left(1 + r_1(\hat{\gamma}_i - 2)\right)(v_i - p_i^j) \ .\end{aligned}$$

□

Incentives of Players $2, \ldots, K-1$ to Move to Ranks $K+1, \ldots, N$

Lemma 7. *In the constant ratio setting, when $r_i = r_1$ for $i \in 2, \cdots, K$, players in slots $2, \ldots, K-1$ have no incentive to move to ranks $K+1, \ldots, N$ during adjacent-ad swapping exploration for any r_1.*

Proof. Since $\alpha_K = \min\{\frac{(\hat{\gamma}_i - 1)^2}{q_{max}}, 1\}$, we have that $\alpha_K q_{max} \leq (\hat{\gamma}_i - 1)^2$ or $\frac{(\hat{\gamma}_i - 1)^2}{\hat{\gamma}_i} \geq \alpha_K q_{max} (\frac{1}{\hat{\gamma}_i})^{K-i}$.

In order to eliminate incentives to deviate to slots $K+1, \ldots, N$, we need to satisfy

$$\left(1 + \frac{r_1(\hat{\gamma}_i - 1)^2}{\hat{\gamma}_i}\right)(v_i - p_i^i) \geq r_1 \alpha_K q_{max} (\frac{1}{\hat{\gamma}_i})^{K-i} v_i,$$

or, alternatively,

$$(v_i - p_i^i) \geq r_1 \left(v_i \left(\alpha_K q_{max} \left(\frac{1}{\hat{\gamma}_i}\right)^{K-i} - \frac{(\hat{\gamma}_i - 1)^2}{\hat{\gamma}_i}\right) - \frac{p_i^i(\hat{\gamma}_i - 1)^2}{\hat{\gamma}_i}\right).$$

But since $(\hat{\gamma}_i - 1)^2/\hat{\gamma}_i \geq \alpha_K q_{max}(1/\hat{\gamma}_i)^{K-i}$, we know that $\alpha_K q_{max}(1/\hat{\gamma}_i)^{K-i} - (\hat{\gamma}_i - 1)^2/\hat{\gamma}_i \leq 0$, and the right-hand side is at most 0. Since the left-hand side is at least 0 (otherwise our assumption of equilibrium prior to exploration does not hold), any r_1 satisfies the condition. □

F Proof of Theorem 10

The proof is divided into a sequence of lemmas. The first one, stated below, follows from a direct application of Hoeffding's inequality and the union bound. Recall that s_i is a chosen slot to explore for ad i.

Lemma 8. *For each i, with probability at least $1 - \delta$, we have*

$$|\hat{c}_i^{s_i} - c_i^{s_i}| \leq \Delta_i, \quad \left|\hat{c}_i^{s_i+1} - c_i^{s_i+1}\right| \leq \Delta_i, \tag{7}$$

where

$$\Delta_i = \max\left\{\sqrt{\frac{\ln(4/\delta)}{2n_{i,s_i}}}, \sqrt{\frac{\ln(4/\delta)}{2n_{i,s_i+1}}}\right\}.$$

Lemma 9. *If Equation 7 holds, then for each i we have*

$$|\hat{\gamma}_i^s - \gamma_i^s| \le \frac{s(s+1)\Delta_i}{\hat{c}_i^{s_i}}.$$

Proof. First we can see that for any i,

$$\begin{aligned}
|\hat{\gamma}_i - \gamma_i| &= \left| \frac{\hat{c}_i^{s_i+1}}{\hat{c}_i^{s_i}} - \frac{c_i^{s_i+1}}{c_i^{s_i}} \right| = \frac{\left|\hat{c}_i^{s_i+1} c_i^{s_i} - \hat{c}_i^{s_i} c_i^{s_i+1}\right|}{\hat{c}_i^{s_i} c_i^{s_i}} \\
&= \frac{\left|(\hat{c}_i^{s_i+1} - c_i^{s_i+1}) c_i^{s_i} - (\hat{c}_i^{s_i} - c_i^{s_i}) c_i^{s_i+1}\right|}{\hat{c}_i^{s_i} c_i^{s_i}} \\
&\le \frac{\left|\hat{c}_i^{s_i+1} - c_i^{s_i+1}\right| c_i^{s_i} + \left|\hat{c}_i^{s_i} - c_i^{s_i}\right| c_i^{s_i+1}}{\hat{c}_i^{s_i} c_i^{s_i}} \\
&\le \frac{\Delta_i (c_i^{s_i} + c_i^{s_i+1})}{\hat{c}_i^{s_i} c_i^{s_i}} \le \frac{2\Delta_i}{\hat{c}_i^{s_i}}.
\end{aligned}$$

Now, by Taylor's theorem, we have

$$(\hat{\gamma}_i)^s = (\gamma_i)^s + s(\hat{\gamma}_i - \gamma_i)\left((\hat{\gamma}_i)^{s-1} - (\gamma_i)^{s-1}\right) + \frac{1}{2}s(s-1)(\hat{\gamma}_i - \gamma_i)^2(\tilde{\gamma})^{s-2},$$

for some $\tilde{\gamma}$ between γ_i and $\hat{\gamma}_i$. Since $\left|(\hat{\gamma}_i)^{s-1} - (\gamma_i)^{s-1}\right| \le 1$ and $0 \le \gamma_i, \hat{\gamma}_i, \tilde{\gamma} \le 1$,[6] we have

$$\begin{aligned}
|(\hat{\gamma}_i)^s - (\gamma_i)^s| &= \left| s(\hat{\gamma}_i - \gamma_i)\left((\hat{\gamma}_i)^{s-1} - (\gamma_i)^{s-1}\right) + \frac{1}{2}s(s-1)(\hat{\gamma}_i - \gamma_i)^2(\tilde{\gamma})^{s-2} \right| \\
&\le |\hat{\gamma}_i - \gamma_i| \left(s + \frac{s(s-1)}{2} \right) \le \frac{s(s+1)\,|\hat{\gamma}_i - \gamma_i|}{2} \le \frac{s(s+1)\Delta_i}{\hat{c}_i^{s_i}}.
\end{aligned}$$

□

With these lemmas, we can prove Theorem 10. For each i and s,

$$\begin{aligned}
|\hat{c}_i^s - c_i^s| &= \left|(\hat{\gamma}_i)^{s-s_i} \hat{c}_i^{s_i} - (\gamma_i)^{s-s_i} c_i^{s_i}\right| \\
&= \left|\left((\hat{\gamma}_i)^{s-s_i} - (\gamma_i)^{s-s_i}\right) \hat{c}_i^{s_i} + (\gamma_i)^{s-s_i}\left(\hat{c}_i^{s_i} - c_i^{s_i}\right)\right| \\
&\le \left|(\hat{\gamma}_i)^{s-s_i} - (\gamma_i)^{s-s_i}\right| \hat{c}_i^{s_i} + (\gamma_i)^{s-s_i} \left|\hat{c}_i^{s_i} - c_i^{s_i}\right| \\
&\le \left((s-s_i)^2 + (s-s_i)\right)\Delta_i + (\gamma_i)^{s-s_i}\Delta_i \\
&\le \Delta_i\left(s^2 + s + s_i^2 + (\gamma_i)^{s-s_i}\right).
\end{aligned}$$

By the previous lemma,

$$(\gamma_i)^{s-s_i} \le (\hat{\gamma}_i)^{s-s_i} + \frac{((s-s_i)^2 + (s-s_i))\Delta_i}{\hat{c}_i^{s_i}} \le (\hat{\gamma}_i)^{s-s_i} + \frac{s^2 + s + s_i^2 \Delta_i}{\hat{c}_i^{s_i}},$$

and we obtain

$$|\hat{c}_i^s - c_i^s| \le \Delta_i\left(s^2 + s + s_i^2 + (\hat{\gamma}_i)^{s-s_i} + \frac{(s^2 + s + s_i^2)\Delta_i}{\hat{c}_i^{s_i}}\right).$$

A simple application of the union bound results in Theorem 10 immediately.

[6] If $\hat{\gamma}_i$ happens to be greater than 1 (which is possible), then we can safely set it to 1. This change can only make the estimate more accurate, since we know $\gamma_i \in (0, 1]$.

G Proof of Theorem 11

Theorem 11 can be obtained by applying Lemma 12 below followed by Lemma 11.

Lemma 10. *Let Y be some subset of a Euclidean space and let $f : Y \to \mathbb{R}$ and $g : Y \to \mathbb{R}$ be two functions. Then, assuming all the maxima below exist,*

$$|\max_{y\in Y} f(y) - \max_{y\in Y} g(y)| \le \max_{y\in Y} |f(y) - g(y)|.$$

Proof. Observe that

$$|\max_{y\in Y} f(y) - \max_{y\in Y} g(y)| = \begin{cases} \max_y f(y) - \max_y g(y) \text{ if } \max_y f(y) \ge \max_y g(y), \\ \max_y g(y) - \max_y f(y) \text{ if } \max_y g(y) \ge \max_y f(y). \end{cases}$$

In the first case,

$$\max_{y\in Y} f(y) - \max_{y\in Y} g(y) \le \max_{y\in Y}(f(y) - g(y)) \le \max_{y\in Y} |f(y) - g(y)|.$$

Similarly, in the second case,

$$\max_{y\in Y} g(y) - \max_{y\in Y} f(y) \le \max_{y\in Y}(g(y) - f(y)) \le \max_{y\in Y} |g(y) - f(y)| = \max_{y\in Y} |f(y) - g(y)|.$$

□

Lemma 11. *Let Y be some subset of a Euclidean space and let $f : Y \to \mathbb{R}$ and $g : Y \to \mathbb{R}$ be two functions such that $\|f - g\|_\infty \le \gamma$. Suppose that $\operatorname{argmax}_{y\in Y} f(y)$ and $\operatorname{argmax}_{y\in Y} g(y)$ exist. Then*

$$|g(\operatorname*{argmax}_{y\in Y} g(y)) - g(\operatorname*{argmax}_{y\in Y} f(y))| \le 2\gamma \ .$$

Proof.

$$\begin{aligned}
&\left|g(\operatorname*{argmax}_{y\in Y} g(y)) - g(\operatorname*{argmax}_{y\in Y} f(y))\right| \\
&= \left|g(\operatorname*{argmax}_{y\in Y} g(y)) - f(\operatorname*{argmax}_{y\in Y} f(y)) + f(\operatorname*{argmax}_{y\in Y} f(y)) - g(\operatorname*{argmax}_{y\in Y} f(y))\right| \\
&\le \left|f(\operatorname*{argmax}_{y\in Y} f(y)) - g(\operatorname*{argmax}_{y\in Y} f(y))\right| + \left|g(\operatorname*{argmax}_{y\in Y} g(y)) - f(\operatorname*{argmax}_{y\in Y} f(y))\right| \\
&\le \gamma + |g(\operatorname*{argmax}_{y\in Y} g(y)) - f(\operatorname*{argmax}_{y\in Y} f(y))| \\
&\le \gamma + \max_{y\in Y} |g(y) - f(y)| \qquad (8) \\
&\le \gamma + \|f - g\|_\infty \le 2\gamma,
\end{aligned}$$

where Equation 8 follows from Lemma 10. □

Fix b_{-i} (e.g., to equilibrium b^*_{-i}). To simplify notation, we remove b_{-i} from expressions below and leave only the response of bidder i, b_i.

Now, recall that x denotes context and X the space of all possible context realizations. During exploration, the prices $p_i(x, b_i)$ will remain as before; thus, only the effective click-through rates will change. We let $c'_i(x, b_i)$ be the effective CTR of player i during exploration and let $\hat{C}^s_i = E_{x\sim D}[\hat{c}^s_i(x)]$ and $C^s_i = E_{x\sim D}[c^s_i(x)]$. Let $\bar{r} = \max_{x\in X} r_1(x)$ and $\bar{\alpha} = \max_s \alpha_s$. Finally, let $\bar{c} = \sup_{x,a,s,b_i} c^s_i(x, b_i) \le 1$, since CTR is just the probability of a click and therefore cannot exceed 1. We define the expected utility of a player i during exploration as

$$U'_i(b) = E_{x\sim D}[c'_i(x, b_i)(v_i - p_i(x, b_i))] \qquad (9)$$

Lemma 12. *Suppose for player i, $|\hat{C}_i^s - C_i^s| \leq \epsilon$ for all s and suppose that there is $\bar{p} < \infty$ such that $E[p_i(x, b_i)] \leq \bar{p}$ for all $b_i \in \mathbb{R}$. Then $\|U_i - U_i'\|_\infty \leq 2\bar{r}\bar{\alpha}(v_i\hat{C}_i^1 + \bar{p} + \epsilon)$ uniformly on $b_i \in \mathbb{R}$.*

Proof. We will demonstrate the uniform bound by bounding the greatest utility loss and gain at any point $b \in \mathbb{R}$. Take advertiser i, some $b \in \mathbb{R}$ and let $X = X^1 \cup \cdots \cup X^K$ be a partition of X into subsets such that for all $x \in X^s$ player i gets slot s. Then

$$E_x[c_i(x, b_i)p_i(x, b_i)] = \sum_{s=1}^{K} \int_{X^s} c_i^s(x)p_i^s(x)dF(x)$$

and

$$\begin{aligned}
E_x[c_i'(x, b_i)p_i(x, b_i)] &= \sum_{s=1}^{K} \int_{X^s} c_i^{'s}(x)p_i^s(x)dF(x) \\
&= \sum_{s=1}^{K} \int_{X^s} \Big(([1 - r_{s-k}(x) - r_s(x)]c_i^s(x) \\
&\qquad\qquad + r_{s-k}(x)c_i^{s-k}(x) + r_s(x)c_i^{s+k}(x))p_i^s(x)\Big)dF(x) \\
&= E_x[c_i(x, b_i)p_i(x, b_i)] + \sum_{s=1}^{K} \int_{X^s} \Big(r_{s-k}(x)(c_i^{s-k}(x) - c_i^s(x))p_i^s(x)\Big)dF(x) \\
&\qquad - \sum_{s=1}^{K} \int_{X^s} \Big(r_s(x)(c_i^s(x) - c_i^{s+k}(x))p_i^s(x)\Big)dF(x) \\
&\leq E_x[c_i(x, b_i)p_i(x, b_i)] + \bar{r}\bar{\alpha} \sum_{s=1}^{K} \int_{X^s} c_i^{s-k}(x)p_i^s(x)dF(x) \\
&\leq E_x[c_i(x, b_i)p_i(x, b_i)] + \bar{r}\bar{\alpha}E[p_i(x, b_i)] \\
&\leq E_x[c_i(x, b_i)p_i(x, b_i)] + \bar{r}\bar{\alpha}\bar{p} \\
&\leq E_x[c_i(x, b_i)p_i(x, b_i)] + 2\bar{r}\bar{\alpha}\bar{p}.
\end{aligned}$$

Alternatively,

$$\begin{aligned}
E_x[c_i'(x, b_i)p_i(x, b_i)] &= \sum_{s=1}^{K} \int_{X^s} c_i^{'s}(x)p_i^s(x)dF(x) \\
&= \sum_{s=1}^{K} \int_{X^s} \Big(([1 - r_{s-k}(x) - r_s(x)]c_i^s(x) \\
&\qquad\qquad + r_{s-k}(x)c_i^{s-k}(x, b_i) + r_s(x)c_i^{s+k}(x))p_i(x)\Big)dF(x) \\
&= E_x[c_i(x, b_i)p_i(x, b_i)] + \sum_{s=1}^{K} \int_{X^s} \Big(r_{s-k}(x)(c_i^{s-k}(x) - c_i^s(x, b_i))p_i^s(x)\Big)dF(x) \\
&\qquad - \sum_{s=1}^{K} \int_{X^s} \Big(r_s(x)(c_i^s(x) - c_i^{s+k}(x))p_i^s(x)\Big)dF(x) \\
&\geq E_x[c_i(x, b_i)p_i(x, b_i)] - 2\bar{r}\bar{\alpha} \sum_{s=1}^{K} \int_{X^s} c_i^s(x)p_i^s(x)dF(x) \\
&\geq E_x[c_i(x, b_i)p_i(x, b_i)] - 2\bar{r}\bar{\alpha}\bar{p}.
\end{aligned}$$

Similarly,

$$v_i E_x[c_i(x,b_i)] - 2\bar{r}\bar{\alpha}E[c_i^1(x)] \le v_i E_x[c_i'(x,b_i)] \le v_i E_x[c_i(x,b_i)] + 2\bar{r}\bar{\alpha}E[c_i^1(x)].$$

Consequently,

$$\begin{aligned}U_i(b) - U_i'(b) &\ge v_i E_x[c_i(x,b_i)] - E_x[c_i(x,b_i)p_i(x,b_i)] - v_i E_x[c_i(x,b_i)] - 2\bar{r}\bar{\alpha}\bar{c} + E_x[c_i(x,b_i)p_i(x,b_i)] - 2\bar{r}\bar{\alpha}\bar{c}\bar{p} \\ &\ge -2\bar{r}\bar{\alpha}(v_i E[c_i^1(x)] + \bar{p})\end{aligned}$$

and

$$\begin{aligned}U_i(b) - U_i'(b) &\le v_i E_x[c_i(x,b_i)] - E_x[c_i(x,b_i)p_i(x,b_i)] - v_i E_x[c_i(x,b_i)] + 2\bar{r}\bar{\alpha}\bar{c} + E_x[c_i(x,b_i)p_i(x,b_i)] + 2\bar{r}\bar{\alpha}\bar{c}\bar{p} \\ &\le 2\bar{r}\bar{\alpha}\bar{c}(v_i E[c_i^1(x)] + \bar{p}).\end{aligned}$$

Thus,

$$|U_i(b) - U_i'(b)| \le 2\bar{r}\bar{\alpha}(v_i E[c_i^1(x)] + \bar{p}) \le 2\bar{r}\bar{\alpha}(v_i E[\hat{c}_i^1(x)] + \bar{p} + \epsilon)$$

Since this bound is independent of b_i, it is uniform on $b_i \in \mathbb{R}$. □

By Lemma 12, then, we have that

$$\|U_i - U_i'\|_\infty \le 2\bar{r}\bar{\alpha}(v_i \hat{C}_i^1 + \bar{p} + \epsilon).$$

We have assumed that the players' bids are in equilibrium prior to exploration. Consequently, the bid b_i of advertiser i prior to exploration maximizes U_i. Let us denote by b_i' the bid which accrues to i the maximum utility under the new effective CTR. Then, the maximum benefit that i gains from deviating under the new CTR is $|U_i'(b_i') - U_i'(b_i)|$. By Lemma 11, we double the bound of Lemma 12 to make it the upper bound on the maximal benefit to deviation for player i, keeping the bids of other advertisers constant:

$$|U_i'(b_i') - U_i'(b_i)| \le 4\bar{r}\bar{\alpha}(v_i \hat{C}_i^1 + \bar{p} + \epsilon).$$

This yields the desired result.