
Volume 30, issue 1, January 2012 ISSN 0167-7187

INTERNATIONAL JOURNAL OF
INDUSTRIAL ORGANIZATION

30 (1) 1–126
(2012)

CONTENTS

International Journal of Industrial Organization 30 (2012) 80–88

Contents lists available at ScienceDirect

International Journal of Industrial Organization

journal homepage: www.elsevier.com/locate/ijio

Comparing predicted prices in auctions for online advertising

Eric Bax [a,*], Anand Kuratti [b], Preston Mcafee [c], Julian Romero [d]

[a] Yahoo! Marketplace Architecture, 3333 Empire Blvd., Burbank, CA 91504, USA
[b] Yahoo! Software Development India Pvt. Ltd., Torrey Pines, Embassy Golflink Business Parks, Off Koramangala-Indiranagar, Intermediate Ring Road, 560 071, India
[c] Yahoo! Research, 3333 Empire Blvd., Burbank, CA 91504, USA
[d] Purdue University 100 S. Grant St. West Lafayette, IN 47907, USA

ARTICLE INFO

Article history:
Received 21 December 2009
Received in revised form 9 June 2011
Accepted 9 June 2011
Available online 14 July 2011

JEL classification:
C13
C44
C45
D81
D84

Keywords:
Reversion
Validation
Bias
Auction
Prediction

ABSTRACT

Online publishers sell opportunities to show ads. Some advertisers pay only if their ad elicits a user response. Publishers estimate response rates for ads in order to estimate expected revenues from showing the ads. Then publishers select ads that maximize estimated expected revenue.

By taking a maximum among estimates, publishers inadvertently select ads based on a combination of actual expected revenue and inaccurate estimation of expected revenue. Publishers can increase actual expected revenue by selecting ads to maximize a combination of estimated expected revenue and estimation accuracy.

1. Introduction

Online publishers use auctions to sell opportunities to advertise, called *ad calls*, to online advertisers. There are two broad categories of online advertising auctions: search and display. In search advertising auctions the advertiser pays only if their ad elicits a click. In display advertising auctions, advertisers may select a basis for payment. Some advertisers pay when the ad is shown, others pay only when showing the ad elicits a user response such as a click or a purchase. (For details on auctions for online advertising, refer to Varian (2006, 2009), Edelman et al. (2007), and Lahaie and Pennock (2007).)

When advertisers pay per click or other user response, the revenue received by the publisher for showing an ad is random. Since user response rates are not known exactly but must be estimated, there is uncertainty in addition to randomness. The estimation accuracy of response rates varies. One reason is that the amount of historical data varies. Another reason is that the response rates themselves vary, and more data is required to estimate smaller rates with the same relative accuracy.

With randomness, a risk-neutral seller seeks to maximize expected revenue. Facing uncertainty, the seller may select an offer having maximum estimated expected revenue. However, this is not necessarily the best policy for maximizing actual expected revenue.

The reason is that selecting a maximum estimate selects for a combination of having an over-estimate and having a large actual expected revenue. Some classes of ads are more likely to have inaccurate estimates, such as ads with lower response rates and ads for which there is less historical data. Even if the individual response rate estimates are unbiased, these classes are more likely to have the largest response rate over-estimates. So selecting a maximum estimate can favor these classes even if they offer less expected revenue than other classes.

Having more buyers in the auction exacerbates the problem, because more estimates means more and more extreme over-estimates. However, having many buyers is not sufficient for selecting a maximum estimate to be a sub-optimal policy for maximizing expected revenue. Varying levels of uncertainty about revenue distributions is also required.

This paper is organized as follows. Section 2 describes related work. Section 3 presents some theory on selection bias for estimated offer values. Section 4 explores correcting selection bias for online display advertising auctions. Section 5 focuses on corrections for search advertising auctions. Section 6 discusses opportunities for future work.

* Corresponding author. Tel.: +1 626 296 1946.
E-mail addresses: ebax@yahoo-inc.com (E. Bax), kuratti@yahoo-inc.com (A. Kuratti), mcafee@yahoo-inc.com (P. Mcafee), jnromero@purdue.edu (J. Romero).

0167-7187/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijindorg.2011.06.001

2. Related work

There are a few areas of work related to this paper. One is work by Athey and Levin (2001) on U.S. Forest Service timber auctions. In that work, as in this paper, the seller selects an offer (ex ante) based on estimated values but is paid (ex post) based on actual values. The forest service work focuses on how buyers can use private information to exploit the seller's estimation and selection process.

Another area of related work, by Wilson (1969) and Thaler (1988), concerns the winner's curse. The winner's curse occurs when multiple bidders estimate the value of an item and submit bids based on those estimates. The auctioneer, by selecting the highest bid, tends to select a bid based on an overestimate of value. As a result, the winner tends to realize less value than their bid. Both the winner's curse and the revenue loss studied in this paper are the result of the difference between actual values and first order statistics of estimates of values. (For more on order statistics, refer to David and Nagaraja (2003).) The revenue loss studied in this paper is borne by the seller or market-maker, because the seller or market-maker must estimate the values of bids and bears the exposure from misestimation. When the bidders, rather than the market-maker, bear the risk, Wilson (1969) gives a method to correct for bias.

Another area of related work is machine learning, where uniform error bounds are used extensively to predict whether a model selected on the basis of limited training data is likely to fit as-yet-unseen data drawn from the same distribution. As limited training data is used to estimate the test performance of more models, it becomes less likely that a model that maximizes estimated expected performance will perform nearly as well as its estimate on test data. (For background on machine learning, see Duda et al. (1973), Valiant (1984), and Devroye et al. (1996). Work on uniform error bounds includes Vapnik and Chervonenkis (1971), Audibert et al. (2007), Langford (2005), and Bax and Callejas (2008).) This effect is similar to the gap between a maximum estimated expected revenue ad and the actual expected revenue from that ad. Both are manifestations of regression to the mean, studied by Galton (1886) and Samuels (1991).

The nested classes of classifiers used in support vector machines and other kernel classifiers are similar to classes of ads with different estimation accuracies in this paper. Kernel methods favor classifiers from classes with more certain bounds on test data performance, even if their estimated expected performance is slightly inferior to classifiers from classes with more uncertainty. For more on support vector machines, refer to Vapnik (1998). For other kernel methods, refer to Shawe-Taylor and Cristianini (2004).

In statistics, Hsu and Chen (1996), Wilcox (1984), and Bechhofer and Turnbull (1978) study procedures to select populations with maximum means among sets of populations. In this paper, offers play the role of populations and awarding an ad call to an offer plays the role of a sample. Their work focuses on determining the number of samples needed to confidently select a population with maximum mean, while this paper focuses on selecting an offer before any further sampling.

3. Theory of selection bias for estimated offers

This section shows that favoring offers that have more accurately estimated offer values improves revenue, under the following model. Actual offer values $\mu_1,\ldots,\mu_n$ are drawn i.i.d. from some distribution. The auctioneer does not know these actual values. Instead, the auctioneer receives unbiased estimates $X_1,\ldots,X_n$ of the offer values. The estimation errors are normal, and the auctioneer knows their standard deviations $\sigma_1,\ldots,\sigma_n$.

For simplicity, we assume a first-price auction in this section. Expected revenue is the actual offer value μ_i of the winning offer. In subsequent sections we focus on second-price auctions.

Suppose the auctioneer selects a parameter value c and selects an offer that maximizes $X_i - c\sigma_i$ as the winning offer. Let $r(c)$ be the expected revenue from the winning offer. Then

$$r(c) = E\left[\mu_{\arg\max_i (X_i - c\sigma_i)}\right],$$

where the expectation is over the distribution of $(\mu_1,\ldots,\mu_n,X_1,\ldots,X_n)$.

The following theorem shows that selecting an offer based on the combination of estimated value and accuracy of estimation $X_i - c\sigma_i$ increases expected revenue over simply selecting an offer with maximum estimated value X_i. Specifically, when there are near ties for $max(X_1,\ldots,X_n)$, expected first-price revenue increases if we break ties in favor of offers with lower σ_i.

Theorem 3.1. *Let unknown actual offer values $\mu_1,\ldots,\mu_n$ be i.i.d. random variables. Let estimated offer values $X_1 \sim \mathcal{N}(\mu_1,\sigma_1),\ldots,X_n \sim \mathcal{N}(\mu_n,\sigma_n)$ be normal random variables with actual offer values $\mu_1,\ldots,\mu_n$ as means and known standard deviations $\sigma_1,\ldots,\sigma_n$. Assume $n \geq 3$ and $\sigma_1,\ldots,\sigma_n$ are not all equal. Then*

$$\left.\frac{\partial r(c)}{\partial c}\right|_{c=0} > 0.$$

The proof is in Appendix A. Here is a sketch of the proof based on a small example. Let $n = 3$, with $\sigma_1 = 0$, $\sigma_2 = 1$, and $\sigma_3 = 2$. Let μ_1, μ_2, and μ_3 be drawn independently and uniformly at random from $\{7,10\}$. Then $X_1 \sim \mathcal{N}(\mu_1,0)$, $X_2 \sim \mathcal{N}(\mu_2,1)$, and $X_3 \sim \mathcal{N}(\mu_3,2)$. Define $X^* = \max(X_1,X_2,X_3)$.

Informally, the theorem says that when X_i and X_j are nearly tied for X^*, if we break the near tie in favor of the value with lower σ, then we increase the expectation of the selected μ. We can ignore cases where $\mu_i = \mu_j$, because selecting either value produces the same winner's μ. (We also ignore three-way ties, because they have probability $O(c^2)$ as $c \to 0$.)

Without loss of generality, assume $\sigma_j < \sigma_i$. Let T be the condition that X_i and X_j are nearly tied for X^*. Let W be the condition that breaking the near tie in favor of lower σ selects the greater μ: $W = \{\mu_i = 7 \wedge \mu_j = 10\}$. Let $\overline{W}$ be the condition that we select the lower-μ value as winner: $\overline{W} = \{\mu_i = 10 \wedge \mu_j = 7\}$. We want to show

$$Pr\{W|T\} > Pr\{\overline{W}|T\}.$$

Using Bayes' Theorem:

$$Pr\{T|W\} = \frac{Pr\{W|T\}Pr\{T\}}{Pr\{W\}},$$

and

$$Pr\{T|\overline{W}\} = \frac{Pr\{\overline{W}|T\}Pr\{T\}}{Pr\{\overline{W}\}}.$$

Since μ_1, μ_2, and μ_3 are i.i.d., $Pr\{W\} = Pr\{\overline{W}\}$. So we only need to show

$$Pr\{T|W\} > Pr\{T|\overline{W}\}.$$

Consider near ties between X_1 and X_2. The correction breaks near ties in favor of X_1 since $\sigma_1 < \sigma_2$. Because $\sigma_1 = 0$, $X_1 = \mu_1$, which is 7 or 10, and any near tie is near X_1. A near tie with $\mu_1 = 10$ and $\mu_2 = 7$ is just as likely as a near tie with $\mu_1 = 7$ and $\mu_2 = 10$; in either case, X_2 is $3\sigma_2$ from μ_2. However, for X_1 and X_2 to be a near tie for X^* requires X_3 less than X_1 and X_2. This is more likely when X_1 and X_2 are near 10 than near 7. As a result, breaking ties in favor of X_1 over X_2 is more likely to occur when $X_1 = \mu_1 = 10$ than when $X_1 = \mu_1 = 7$.

Similar reasoning applies to near ties between X_1 and X_3. They must occur with $X_1 = \mu_1$ at 7 or 10. If $\mu_1 \neq \mu_3$, then a near tie is equally likely at 7 or 10, because both require X_3 to be $\frac{3}{2}\sigma_3$ from u_3. However, a near tie for X^* is more likely with $X_1 = 10$ because it only requires X_2 to be less than 10, which is more likely than X_2 being less than 7.

Now consider near ties between X_2 and X_3. In Fig. 1, the dotted line shows the pdf of near ties given $\mu_2 = 10$ and $\mu_3 = 7$. The dotted-and-dashed line shows the pdf of near ties given $\mu_2 = 7$ and $\mu_3 = 10$.The pdfs peak closer to μ_2 than to μ_3, because $\sigma_2 < \sigma_3$. The pdfs are reflections of each other around the midpoint 8.5 between 7 and 10. Above 8.5, near ties are more likely the result of $\mu_2 = 10$ and $\mu_3 = 7$ than of $\mu_2 = 7$ and $\mu_3 = 10$. For a near tie between X_2 and X_3 to be a near tie for X^*, X_1 must be below the near tie. This is a certainty above 10, where the near tie is more likely the result of $\mu_2 = 10$, and it is impossible below 7, where the near tie is more likely the result of $\mu_2 = 7$. (Between, it is 50%.) So breaking the near tie in favor of X_2 is more likely to select X_2 when $\mu_2 = 10$ than when $\mu_2 = 7$.

4. Selection bias in display advertising

This section focuses on selection bias in display advertising, which is the portion of online advertising with graphical ads rather than text-only ads. (The next section focuses on text only ads, which includes most search engine advertising.) Subsection 4.1 discusses the role of estimated offer values in display advertising. Subsection 4.2 explores how estimated offer values impact revenue. Subsection 4.3 uses simulated auctions to evaluate a correction for selection bias.

4.1. Display advertising and estimated offer values

Marketplaces for display advertising such as the RightMedia Exchange host auctions where publishers sell ad calls – opportunities to advertise – and advertisers buy them. Advertisers have a choice of price types, including cost-per-impression (CPM), cost-per-click (CPC), and cost-per-action or cost-per-acquisition (CPA). CPM offers pay when their ad is displayed. (The abbreviation CPM represents cost per *mille*, or thousand impressions; in this paper we treat CPM prices as per-impression prices.)

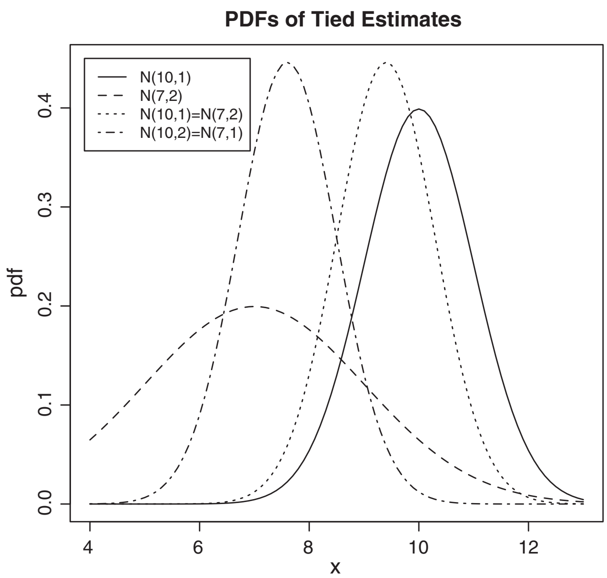

Fig. 1. The dotted line is the pdf for a tie between $\mathcal{N}(10,1)$ and $\mathcal{N}(7,2)$. (These two distributions are shown for reference.) The other line is the pdf for a tie after swapping means — a tie between $\mathcal{N}(7,1)$ and $\mathcal{N}(10,2)$. That pdf is a reflection of the first, around the midpoint, 8.5. Each tie pdf peaks closer to the mean of the normal with the lower standard deviation. So higher-valued ties are more likely when the higher mean has the lower standard deviation.

CPC and CPA offers pay only if displaying the ad elicits a user response. For CPC offers, the response is a user clicking on the ad. For CPA offers, the advertiser specifies the response; examples include a user completing an online purchase, filling out a form, or visiting a web page.

An auctioneer can use expected offer values to compare offers with different price types. Let p^* be the response rate for each ad — the probability that displaying the ad will elicit the user response required for the advertiser to pay. For CPM offers, $p^* = 1.0$. For CPC offers, p^* is the click-through rate. For CPA offers, p^* is the probability that the user will complete the action specified by the advertiser. Let b be the bid amount, the amount an advertiser pays for responses. Then the expected offer value is bp^*. The auctioneer's goal is to select an offer with maximum bp^*.

Since response rates p^* are unknown for CPC and CPA offers, the auctioneer uses estimated response rates p. There are many methods to estimate probabilities of clicks and conversions. The simplest method is Bernoulli sampling, where the fraction of auction wins that result in a click or conversion is the estimated probability. Generally, methods begin with a prior based on results for similar ads and content. Then, Bernoulli sampling is used to modify the estimated probability. In essence, most methods try to "partially borrow" samples from other ads and content, for which there is plenty of data, and then tune the estimate based on actual performance of the ad on the same or similar content. As samples accumulate, the estimated probability is based more and more on Bernoulli sampling. The following analysis focuses on Bernoulli sampling, but the general principles also apply to more complex prediction methods.

Let p^* be the actual probability of action for a performance ad. Let p be the estimate of p^* based on Bernoulli sampling. The estimate p has a binomial distribution, with mean p^*. If the ad is shown n times, resulting in k actions, then

$$p \equiv \frac{k}{n}.$$

The estimate p is unbiased:

$$\mu(p) = p^*.$$

The variance is

$$\sigma^2(p) = \frac{np^*(1-p^*)}{n^2} = \frac{p^*(1-p^*)}{n}.$$

The standard deviation is

$$\sigma(p) = \sqrt{\frac{p^*(1-p^*)}{n}}.$$

For click prediction, p^* is on the order of 0.01. For conversion prediction, it can be on the order of 0.001. In both cases, the square root of $1-p^*$ is very close to one. Also, for normal variables, the ratio between mean absolute deviation and standard deviation is $\sqrt{\frac{2}{\pi}} \approx 0.8$, and we can apply this as an approximation for Bernoulli sampling. Hence,

$$E\left[\left|p-p^*\right|\right] \approx 0.8\sqrt{\frac{p^*}{n}}.$$

Now consider how the estimation error in response rates affects estimation error in expected offer values. There is no bias:

$$E\left[bp-bp^*\right] = bE\left[p-p^*\right] = 0.$$

However,

$$E\left[\left|bp-bp^*\right|\right] = bE\left[\left|p-p^*\right|\right] \approx 0.8b\frac{\sqrt{p^*}}{\sqrt{n}}.$$

Consider this as a fraction of the actual expected offer value:

$$E\left[\frac{\left|bp-bp^*\right|}{bp^*}\right] \approx 0.8\frac{1}{\sqrt{np^*}}.$$

So the relative error due to estimation grows as p^* shrinks.

As a result, CPA ads are likely to have much less accurate estimates than CPC ads. For example, with $n=10{,}000$ samples and $p^*=0.001$ for a CPA ad, the estimated expected offer value is expected to differ from the actual by about 25%. In contrast, for a CPC ad with $p^*=0.01$ and the same number of samples, the expected relative error is only about 8%. (For a CPM ad the relative error is 0%, since the probability of payout is known with certainty.)

The gap in accuracy between CPA and CPC ads can be even worse in practice than in these examples. Conversions for one ad may be based on different actions than conversions on other ads. So using conversions from one ad to estimate conversion probabilities for other ads is usually less effective than doing so for click probabilities.

Another way to view the formula above is to keep relative error constant, if the probability of action p^* shrinks by some factor, then the number of auctions needed to learn p^* must grow by the same factor. For our examples with $p^*=0.01$ for CPC ads and $p^*=0.001$ for CPA ads, ten times as many learning auctions must be devoted to each CPA ad as to each CPC ad in order to have the same expected relative error for both.

4.2. How selection bias impacts display auction revenue

The auctioneer selects a winner by maximizing the estimated expected offer value. Ideally, the auctioneer would select an offer that maximizes actual expected value. This subsection examines how using estimated rather than actual expected values impacts expected revenue.

Let *maximal expected revenue* r^* be the revenue obtained by an auction based on actual response rates, and let *expected revenue* r be the revenue obtained based on estimates. Define *revenue impact* R to be the portion of maximal expected revenue foregone by using estimates:

$$R = \frac{r^*-r}{r^*}.$$

We use second-price auctions here and in simulations in the next subsection. These auctions are common in online advertising (see Varian (2006, 2009) and Edelman et al. (2007)), though they are not necessarily revenue-optimizing (see Lahaie and Pennock (2007) and Myerson (1981)). For general information on auction mechanisms, refer to Milgrom (2004) or Krishna (2002).

For the auctions, let

$$w = \arg\max_i \; b_i p_i,$$

where i indexes offers, with bids b_i and estimated response rates p_i. Offer w wins the auction. (In case of a tie, select w uniformly at random from indices of tied expected payouts.) Let

$$s = \arg\max_{i \neq w} \; b_i p_i.$$

Call offer s the second-place offer. The charge for the winning offer is

$$a_w = min\left(\frac{p_s b_s}{p_w} + \frac{\varepsilon}{p_w}, b_w\right),$$

where ε is the minimum bid increment, usually \$0.01. If offer w is a CPM offer, then the advertiser is charged a_w. For a CPC offer, the advertiser is charged a_w if the ad is clicked. For a CPA offer, the advertised is charge a_w if showing the ad elicits the specified response.

Let p_w^* be the actual response rate for which p_w is an estimate. Then the expected revenue from the auction is

$$r = a_w p_w^*.$$

In our analysis, we will ignore the added ε in a_w. Then

$$a_w = \frac{p_s b_s}{p_w}.$$

So expected revenue is

$$r = \frac{p_s b_s}{p_w} p_w^*.$$

To compute maximal expected revenue r^*, the expected revenue if the auctioneer could select winning and second-place offers based on actual response rates rather than estimates, define the maximal winning index

$$w^* = \arg\max_i \; p_i^* b_i.$$

Define the maximal second-place index

$$s^* = \arg\max_{i \neq w} \; p_i^* b_i.$$

Then the maximal expected revenue is

$$r^* = \frac{p_{s^*}^* b_{s^*}}{p_{w^*}^*} p_{w^*}^*.$$

Since the actual response rates are unknown, it is not possible to observe maximal revenue in real auctions. However, since responses are observed for auction winners, it is possible to measure the average difference between estimated and observed response rates for winners. This difference between p_w and p_w^*, scaled by a_w, is the difference between estimated expected revenue when a winner is selected and actual revenue received. In the RightMedia exchange, if there were no corrections, the difference for CPA winners would be about 20%. The difference for CPC winners would be less than 10%.

4.3. Correcting for selection bias

In this subsection, we use simulations to examine how much we can improve revenue and selectivity by adjusting estimated offer values for selection bias. We estimate the standard deviation

$$\sqrt{\frac{p^*(1-p^*)}{n}},$$

by using the estimate p in place of (the unknown) p^*. The adjusted probability estimate is

$$\hat{p} = p - c\sqrt{\frac{p(1-p)}{n}}.$$

We experiment with a variety of values for the coefficient c. In practice, the coefficient c can be selected through empirical observations and experiments to optimize some combination of revenue and selectivity.

Figs. 2 and 3 show how the number of offers in each auction affects revenue and selectivity with adjustments based on standard deviations. Both figures are based on the same simulations. For each number of offers in $4, 8, \ldots, 40$, a different set of one million simulated auctions is generated and used for all values of c. For each auction, the specified number of offers are generated independently at random. For each offer, whether it is CPM, CPC, or CPA is determined uniformly at random. For CPM offers, $p^* = 1.0$. For CPC offers, p^* is drawn uniformly at random from $\{0.005, 0.01, 0.02, 0.05\}$. For CPA offers, p^* is drawn uniformly at random from $\{0.0002, 0.0005, 0.001, 0.002\}$. For all offer types, the number n of simulated learning auctions is drawn uniformly at random from $\{5{,}000, 10{,}000, 50{,}000, 100{,}000\}$. Then the estimated probability of action p is determined by drawing from a binomial distribution based on p^* and n. Actual values are drawn at random from a normal distribution with mean \$1.00 and standard deviation \$0.10, and bids are set by dividing actual values by probabilities of action p^*.

Fig. 2 shows revenue impact, defined as the fraction of maximal expected revenue lost due to using estimated probabilities:

$$\hat{R} = \frac{r^* - \hat{r}}{r^*},$$

where r^* is the maximal expected revenue and $\hat{r}$ is the expected revenue when using $\hat{p}$ to estimate response rates. Fig. 3 shows *selectivity*, defined as the fraction of auctions won by the offer with highest actual expected value. The values for $c = 0$, on the y-axes, are the baseline values achieved without bias correction. As c increases from zero, revenue impact and selectivity improve smoothly. Then further increases in c yield poorer results.

For each auction size, revenue-optimizing c values are close to selectivity-optimizing values. The effects of optimal adjustments are more pronounced for larger auctions, and optimal c values increase with auction size. Even for optimal c values, the corrections leave room for improvement: for large auctions, revenues are about 5% less than maximal, and the best offer is selected in only about 50% of auctions.

5. Selection bias in search advertising

This section focuses on text advertising, the type of online advertising most commonly associated with search engines. Subsection 5.1 describes the slot auctions used for text advertising. Subsection 5.2 uses simulations to examine the impact of adjusting these auctions for selection bias.

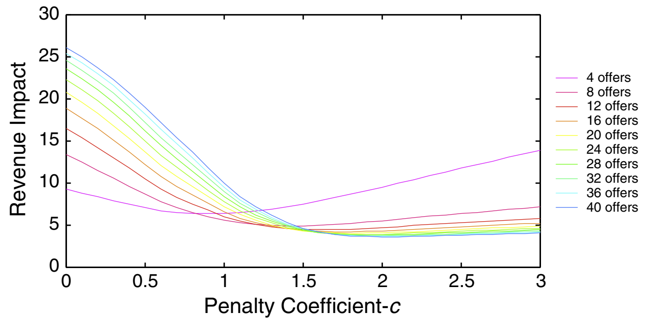

Fig. 2. Revenue impact over auction sizes.

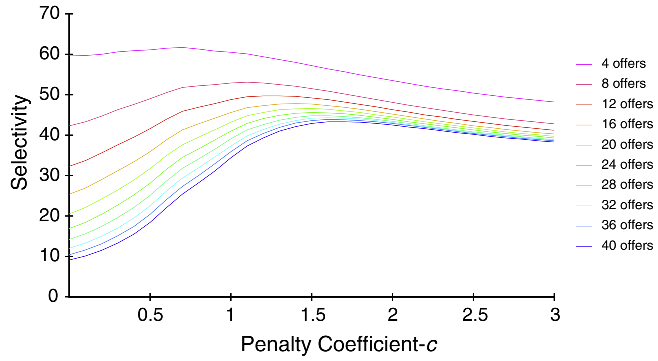

Fig. 3. Impact on selectivity over auction sizes.

5.1. The search advertising auction

In the search advertising auctions, there are multiple slots for ads on each page and hence multiple auction winners. The auctioneer orders offers by estimated expected offer value (breaking ties randomly.) The most desirable slot is awarded to the first offer, the second most desirable slot is awarded to the second offer, and so on. The charge for each winning offer is based on the estimated expected offer value of the next offer, with the intention of charging the first winner the second price, the second winner the third price, and so on. Typically, search auctions have only CPC pricing.

We will use notation similar to that for display auctions. As in display auctions, for offer i, let b_i be the bid, p_i be the estimated probability of action, and p_i^* be the actual probability of action. Also, the estimated expected offer value is $b_i p_i$ and the actual expected offer value is $b_i p_i^*$.

We will use a generalized second-price auction model (Edelman et al., 2007; Varian, 2009.) Let m be the number of offers, and let k be the number of ad slots. For $j \in \{1, \ldots, m\}$, let w_j be the original index $i \in \{1, \ldots, m\}$ of the offer in position j after ordering offers by estimated expected offer value. For example, w_1 is the index of the winner of the first ad slot. Similarly, let w_j^* be the original index i of the offer in position j after ordering offers by actual expected offer value. Then expected revenue is

$$r = \sum_{j=1}^{k} \frac{b_{w_{j+1}} p_{w_{j+1}}}{p_{w_j}} p^*_{w_j}.$$

Similarly, maximal revenue is

$$r^* = \sum_{j=1}^{k} \frac{b_{w^*_{j+1}} p^*_{w^*_{j+1}}}{p^*_{w^*_j}} p^*_{w^*_j} = \sum_{j=1}^{k} b_{w^*_{j+1}} p^*_{w^*_{j+1}}.$$

As in display, define the revenue impact for search as

$$R = \frac{r^* - r}{r^*}.$$

In practice, response rates decrease as an ad moves from more to less desirable slots. (See Varian (2006), Blumrosen et al. (2008), Kempe and Mahdian (2008); and Gomes et al. (2008) for more detail.) For simplicity, we ignore this effect in our simulations. Including this effect would increase the revenue impact from the early ad slots and decrease the impact from the later slots. When the effect is strong, the revenue from the top slot overwhelms the revenue from other slots, resembling the single-slot display auction. When the effect is weak, the effects on revenue resemble those in this section.

5.2. Adjusting estimated probabilities

This subsection explores a method to increase revenue by adjusting estimated probabilities in search auctions. The method uses different adjustment coefficients c for different ad slots, because top ad slots have more competing offers than subsequent slots. In each case, each adjusted estimate $\hat{p}$ is

$$\hat{p} = p - c\sqrt{\frac{p(1-p)}{n}}.$$

Let $\mathbf{c} = (c_1,\ldots,c_k)$ be the sequence of c values for ad slots. Then the auction procedure is as follows. Start with slot 1. Adjust probabilities of action for all offers using c_1. Order by adjusted estimated expected offer values to determine a winner for the first slot and a charge based on the second offer in the ordering. Remove the winner. Then repeat this process, using c_2 for the second slot, c_3 for the third slot, and so on.

Define $\hat{r}_d$ to be the expected revenue using this procedure. Define

$$\hat{R}_d = \frac{r^* - \hat{r}_d}{r^*}$$

to be the revenue impact.

Table 1 shows results of simulations to determine the revenue impact. Each column is based on a set of 10,000 simulated five-slot auctions. Each offer is generated independently, with:

- Actual value determined at random from a normal distribution with mean \$1 and standard deviation \$0.10.
- p^* selected uniformly at random from $\{0.005, 0.01, 0.015, \ldots 0.05\}$.
- p drawn at random based on a binomial distribution simulating n learning auctions, with n selected uniformly at random from $\{1,000, 2,000, 3,000, \ldots, 10,000\}$.

For each column, an optimal value of $\mathbf{c}$, called $\mathbf{c}^*$, is computed using gradient descent over a different set of auctions than those used for the results shown in Table 1.

In Table 1, for each number of offers, optimal adjustments are greatest for the top slot and decrease from slot to slot, because expected selection bias is greater when selecting a maximum than a runner-up and so on. When selecting a minimum, expected selection bias is negative. For six offers, the lower slots have negative optimal adjustments. Selecting an offer among six for the fifth slot is similar to selecting a runner-up for the minimum-value offer.

Table 1
Revenue impact in simulated search auctions.

		Number of Offers					
		6	10	15	20	25	30
	r^*	4.87	5.20	5.38	5.48	5.55	5.61
$c=0$	r	4.69	5.00	5.13	5.19	5.23	5.26
	R	3.77	3.91	4.63	5.28	5.81	6.21
Variable c	$\mathbf{c}^*$	0.41	0.78	1.04	1.23	1.37	1.46
		0.10	0.54	0.87	1.07	1.24	1.31
		−0.31	0.26	0.75	0.94	1.18	1.22
		−0.68	0.12	0.67	0.86	1.14	1.16
		−1.15	−0.01	0.58	0.83	1.11	1.12
	$\hat{r}$	4.72	5.01	5.15	5.23	5.29	5.33
	$\hat{R}_d$	3.22	3.72	4.23	4.54	4.77	4.91
$R-\hat{R}_d$		0.55	0.19	0.40	0.74	1.04	1.30

For each slot, optimal adjustments increase as the number of offers increases, because expected selection bias increases with more competition. In the bottom row of Table 1, as the number of offers increases beyond ten, the value of using the adjustment increases. (For six offers, the adjustment has a strong effect because it corrects for the large negative bias in the lower slots.)

6. Conclusion

This paper explores the impact of using estimates of offer values in an auction. Using estimates introduces a bias that can significantly reduce revenue and selectivity. This paper also outlines a method to correct for the bias, improving revenue and selectivity. The method selects an auction winner based on a combination of estimated offer value and an estimate of the estimation error.

The method in this paper has free parameters. To apply the method in practice, it is possible to use simulations to select starting points for the parameters. Then use statistical optimization techniques, as in Box et al. (2005) to adjust the parameters, optimizing for any desired combination of revenue and selectivity. Fortunately, the simulations in this paper indicate that revenue-optimal parameter settings are similar to selectivity-optimal ones.

Our simulations showed that optimal values of c depend on the number of offers in each auction. This is similar to classical shrinkage methods such as James–Stein estimation (James and Stein, 1961; Stein, 1955). Most shrinkage methods are designed to minimize average error over the quantities being estimated; see for example Brown (1966) and Bock (1975). For auctions with a single winner, it would be interesting to explore whether there are estimators that tend to select the offer with highest actual mean directly, rather than first applying shrinkage methods and then selecting the maximum estimate.

In practice, many auctions contain some offers that are not competitive. Those offers should be removed before applying corrections or shrinkage. Uncompetitive offers can be identified using uniform error bound methods from machine learning, such as Hoeffding (1963) bounds or Audibert et al.'s empirical Bernstein (Audibert et al., 2007; Mnih et al., 2008) bounds. Offers with upper bounds on value less than the maximum offer value lower bound can be declared uncompetitive and removed.

When the response probability estimates are based on sampling, it would be interesting to explore whether the method in this paper could be improved by using more sophisticated methods to estimate confidence intervals for binomial proportions than the technique based on sample standard deviation used in this paper, which is called the *normal approximation interval*. Some alternative methods are the Wilson score interval (Wilson, 1927) and the Clopper–Pearson interval (Clopper and Pearson, 1934). There are several papers that compare different methods, such as Agresti and Coull (1998), Brown et al. (2001), and Ross (2003).

In general, it is possible to use any of a variety of machine learning approaches to determine functional forms for the adjustments and set parameters for those forms. Inputs can include the number of offers, their estimated values, and any available information about the distributions of actual values, such as how much frequencies of action have varied over time for each offer or for sets of offers. Since online advertising marketplaces hold many auctions, the amount of data needed for machine learning approaches is available to them. One such approach is to use Bayesian principles, basing adjustments on priors developed using empirical data from past auctions. For details on Bayesian methods, refer to Berger (1985), Duda et al. (1973), and Gelman (2004).

The simulations for search advertising auctions indicate that using different correction factors for different ad slots can improve revenue and selectivity. It would be interesting to explore whether a similar tactic can improve methods to correct for uncertainty in portfolio allocations for financial markets, as discussed in Jorion (1986), Jobson et al. (1979), and Lintner (1965). For example, it may be useful to apply one correction to all available investments, select one or a few investments to receive a portion of the resource allocation, remove those investments, apply a weaker correction to those remaining, and then select among them to receive the remaining resource allocation.

In the future, it would be interesting to examine interactions between correcting for selection bias and bidding strategy. For example, bidders may respond to corrections by choosing to submit firmer bids, such as using CPM pricing instead of CPC or CPA. In this case, though, bidders shift some risk from sellers to themselves, and they also incur the computational burden of estimating response probabilities. Alternatively, bidders may offer more than their actual values initially to win auctions, generate responses, and reduce uncertainty about their offer values. Then the bidders may reduce their bids, using selection bias corrections as a barrier for competition.

Another direction for future work is to explore interactions between learning, correcting for selection bias, and maximizing revenue over time. Multi-armed bandit literature, including Gittins (1979), Auer et al. (2002), and Audibert et al. (2007), examines strategies to optimize revenue over time by awarding some wins to offers with uncertain values now to reduce uncertainty about their values in the future. In general, corrections for selection bias may discourage learning by awarding more wins to offers with more certain values. So it may be useful to select winners based on a combination of selection bias correction and the value of learning.

Appendix A. Proof of Theorem 3.1

Proof. We want to show

$$\left.\frac{\partial r(c)}{\partial c}\right|_{c=0} > 0.$$

By the definition of a derivative,

$$\left.\frac{\partial r(c)}{\partial c}\right|_{c=0} = \lim_{c\to 0}\frac{1}{c}[r(c)-r(0)].$$

By the definition of $r()$ and linearity of expectations, this is

$$\left.\frac{\partial r(c)}{\partial c}\right|_{c=0} = \lim_{c\to 0}\frac{1}{c}E\left[\mu_{\arg\max_i(X_i-c\sigma_i)} - \mu_{\arg\max_i X_i}\right]. \quad (1)$$

The difference between μ values is nonzero only if applying the adjustment $-c\sigma$ alters who wins:

$$\arg\max_i(X_i-c\sigma_i) \neq \arg\max_i X_i.$$

Define E_{ij} to be the event that the adjustment promotes X_j over X_i, in other words, $X_i = max(X_1,\ldots,X_n)$ and $X_j > X_i - c(\sigma_i - \sigma_j)$. Ignore ties for $max(X_1,\ldots,X_n)$ because they have zero support, and ignore intersections among E_{ij} for different ij pairs because they have probability $O(c^2)$ as $c\to 0$.

Note that E_{ij} requires $\sigma_j<\sigma_i$. Let $g()$ be the distribution of μ values. Let $f()$ be the standard normal pdf. (So $\frac{1}{\sigma_i}f\left(\frac{X_i-\mu_i}{\sigma_i}\right)$ is the pdf for X_i.) Integrate over $\mu_i, \mu_j, and\ X_i$:

$$\left.\frac{\partial r(c)}{\partial c}\right|_{c=0} = \lim_{c\to 0}\frac{1}{c}\sum_{(i,j):\sigma_j<\sigma_i}\int_{\mu_i}\int_{\mu_j}\int_{X_i} g(\mu_i)g\left(\mu_j\right)\frac{1}{\sigma_i}f\left(\frac{X_i-\mu_i}{\sigma_i}\right) \quad (2)$$
$$\times\left[\left(\mu_j-\mu_i\right)Pr\left\{E_{ij}|\mu_j,X_i\right\}\right]d\mu_i d\mu_j dX_i.$$

Note that

$$\lim_{c\to 0}\frac{1}{c}Pr\left\{E_{ij}\middle|\mu_j,X_i\right\} = \lim_{c\to 0}\frac{1}{c}Pr\left\{X_j\in\left(X_i-c\left(\sigma_i-\sigma_j\right),X_i\right)\middle|X_j\sim\mathcal{N}\left(\mu_j,\sigma_j\right)\right\}$$
$$\times Pr\{\forall k\notin\{i,j\}: X_k<X_i\}$$
$$= \left(\sigma_i-\sigma_j\right)\frac{1}{\sigma_j}f\left(\frac{X_i-\mu_j}{\sigma_j}\right)\prod_{k\notin\{i,j\}}Pr\{X_k<X_i\}.$$

So, from Eq. (2),

$$\left.\frac{\partial r(c)}{\partial c}\right|_{c=0} = \sum_{(i,j):\sigma_j<\sigma_i}\left(\sigma_i-\sigma_j\right)\int_{\mu_i}\int_{\mu_j}g(\mu_i)g\left(\mu_j\right)\left(\mu_j-\mu_i\right)$$
$$\times\int_{X_i}\frac{1}{\sigma_i}f\left(\frac{X_i-\mu_i}{\sigma_i}\right)\frac{1}{\sigma_j}f\left(\frac{X_i-\mu_j}{\sigma_j}\right)\prod_{k\notin\{i,j\}}Pr\{X_k<X_i\}d\mu_i\ d\mu_j\ dX_i.$$

Let $a = max(\mu_i,\mu_j)$ and $b = min(\mu_i,\mu_j)$. Integrate over a and b:

$$\left.\frac{\partial r(c)}{\partial c}\right|_{c=0} = \sum_{(i,j):\sigma_j<\sigma_i}\left(\sigma_i-\sigma_j\right)\int_a\int_{b<a}g(a)g(b)(a-b)da\ \ db$$
$$\times\int_{X_i}\left[\frac{1}{\sigma_j}f\left(\frac{X_i-a}{\sigma_j}\right)\frac{1}{\sigma_i}f\left(\frac{X_i-b}{\sigma_i}\right)-\frac{1}{\sigma_j}f\left(\frac{X_i-b}{\sigma_j}\right)\frac{1}{\sigma_i}f\left(\frac{X_i-a}{\sigma_i}\right)\right]$$
$$\times\prod_{k\notin\{i,j\}}Pr\{X_k<X_i\}\ dX_i.$$

The terms before the integral over X_i are nonnegative, and they are all positive for some values of i, j, and $a>b$. So to show the RHS is positive, we only need to show that the integral over X_i is positive for all i, j, and $a>b$.

Define

$$p_a(x) = \frac{1}{\sigma_j}f\left(\frac{x-a}{\sigma_j}\right)\frac{1}{\sigma_i}f\left(\frac{x-b}{\sigma_i}\right),$$

$$p_b(x) = \frac{1}{\sigma_j}f\left(\frac{x-b}{\sigma_j}\right)\frac{1}{\sigma_i}f\left(\frac{x-a}{\sigma_i}\right),$$

and

$$m_3 = \max_{k\notin\{i,j\}}X_k.$$

Then the integral over X_i becomes

$$D = \int_{x=-\infty}^{\infty}Pr\{m_3<x\}(p_a(x)-p_b(x))dx.$$

The midpoint $h = \frac{a+b}{2}$ is an equal number of standard deviations from the mean of X_j whether $\mu_j = a$ or $\mu_j = b$. The same holds for X_i. So $p_a(h) = p_b(h)$. Also, since $f(x) = f(-x)$:

$$\forall s: p_a(h+s) = p_b(h-s).$$

Center at h:

$$D = \int_{s=-\infty}^{\infty}Pr\{m_3<h+s\}(p_a(h+s)-p_b(h+s))ds.$$

Fold at $s=0$:

$$D=\int_{s=0}^{\infty}\left[Pr\{m_3<h+s\}(p_a(h+s)-p_b(h+s))+Pr\{m_3<h-s\}\times(p_a(h-s)-p_b(h-s))\right]\mathrm{d}s.$$

Use $p_a(h+s)=p_b(h-s)$:

$$D=\int_{s=0}^{\infty}\left[Pr\{m_3<h+s\}(p_a(h+s)-p_b(h+s))-Pr\{m_3<h-s\}\times(p_a(h+s)-p_b(h+s))\right]\mathrm{d}s.$$

Collect terms:

$$D=\int_{s=0}^{\infty}[p_a(h+s)-p_b(h+s)][Pr\{m_3<h+s\}-Pr\{m_3<h-s\}]\mathrm{d}s.$$

For $x>h$, $p_a(x)>p_b(x)$, so the first bracketed term is positive. (We prove this as Lemma A.1 below.) For $n\geq 3$, the second bracketed term is positive, because the cdf of m_3 is an increasing function. So $D>0$. ■

The proof shows that some adjustment based on standard deviations increases expected revenue. In addition, it shows that $\sigma_j<\sigma_i$ implies $\mu_j>\mu_i$ is more likely than $\mu_{\backslash j}<\mu_i$, given that X_j and X_i are nearly tied for $max(X_1,\ldots,X_n)$. In other words, it shows that some adjustment increases *selectivity* — the probability of selecting an offer with maximum actual expected value.

Now we prove $p_a(h+s)>p_b(h+s)$ for $s>0$:

Lemma A.1. *Let* $\sigma_j<\sigma_i$, $a>b$, $h=\frac{a+b}{2}$,

$$p_a=\frac{1}{\sigma_j\sigma_i}f\left(\frac{x-a}{\sigma_j}\right)f\left(\frac{x-b}{\sigma_i}\right),$$

and

$$p_b=\frac{1}{\sigma_j\sigma_i}f\left(\frac{x-a}{\sigma_i}\right)f\left(\frac{x-b}{\sigma_j}\right),$$

where $f()$ *is the standard normal pdf.*

Then

$$\forall s>0: p_a(h+s)>p_b(h+s).$$

Proof. Let $z_1=\frac{|x-a|}{\sigma_j}$, $z_2=\frac{|x-b|}{\sigma_i}$, $z_3=\frac{|x-a|}{\sigma_i}$, and $z_4=\frac{|x-b|}{\sigma_j}$. Since $f(x)=f(-x)$, $p_a=\frac{1}{\sigma_j\sigma_i}f(z_1)f(z_2)$ and $p_b=\frac{1}{\sigma_j\sigma_i}f(z_3)_f(z_4)$.We need to show $f(z_1)f(z_2)>f(z_3)f(z_4)$. Using the standard normal pdf:

$$f(z_1)f(z_2)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z_1^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z_1^2}=\frac{1}{2\pi}e^{-\frac{1}{2}(z_1^2+z_2^2)}.$$

Similarly,

$$f(z_3)f(z_4)=\frac{1}{2\pi}e^{-\frac{1}{2}(z_3^2+z_4^2)}.$$

So we need to show $z_1^2+z_2^2<z_3^2+z_4^2$.

First consider the case $h+s\geq a$. Let $m=a-h=h-b$, and let $d=h+s-a$. Then $z_1=\frac{d}{\sigma_j}$, $z_2=\frac{2m+d}{\sigma_i}$, $z_3=\frac{d}{\sigma_i}$, and $z_4=\frac{2m+d}{\sigma_j}$. So

$$z_1^2+z_2^2-z_3^2-z_4^2$$

$$=\frac{d^2}{\sigma_j^2}+\frac{(2m+d)^2}{\sigma_i^2}-\frac{d^2}{\sigma_i^2}-\frac{(2m+d)^2}{\sigma_j^2}$$

$$=\frac{4m}{\sigma_j^2\sigma_i^2}\left[\left(\sigma_j^2-\sigma_i^2\right)(m+d)\right].$$

This is negative since $d\geq 0$, all other variables are positive, and $\sigma_j<\sigma_i$.

Now consider the case $h<h+s<a$. Define m as before. Then $z_1=\frac{m-s}{\sigma_j}$, $z_2=\frac{m+s}{\sigma_i}$, $z_3=\frac{m-s}{\sigma_i}$, and $z_4=\frac{m+s}{\sigma_j}$. So

$$z_1^2+z_2^2-z_3^2-z_4^2$$

$$=\frac{(m-s)^2}{\sigma_j^2}+\frac{(m+s)^2}{\sigma_i^2}-\frac{(m-s)^2}{\sigma_i^2}-\frac{(m+s)^2}{\sigma_j^2}$$

$$=\frac{4ms}{\sigma_j^2\sigma_i^2}\left[\sigma_j^1-\sigma_i^2\right].$$

This is negative since all variables are positive and $\sigma_j<\sigma_i$. ■

References

Agresti, A., Coull, B.A., 1998. Approximate is better than 'exact' for interval estimation of biniomial proportions. The American Statistician 52, 119–126.

Athey, S., Levin, J., 2001. Information and competition in U.S. forest service timber auctions. Journal of Political Economy 109 (2), 375–417.

Audibert, J.-Y., Munos, R., Szepesvari, C., 2007. Variance estimates and exploration function in multi-armed bandit. CERTIS Research Report 07–31.

Auer, P., Cesa-Bianchi, N., Fischer, P., 2002. Finite-time analysis of the multi-armed bandit problem. Machine Learning 47, 235–256.

Bax, E., Callejas, A., 2008. An error bound based on a worst likely assignment. Journal of Machine Learning Research 9, 859–891.

Bechhofer, R.E., Turnbull, B.W., 1978. Two $(k+1)$-decision selection procedures for comparing k normal means with a specified standard. Journal of the American Statistical Association 73, 385–392.

Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis, 2nd Edition. Springer.

Blumrosen, L., Hartline, J.D., Nong, S., 2008. Position auctions and non-uniform conversion rates. ACM EC Workshop on Advertisement Auctions.

Bock, M.E., 1975. Minimax estimators of the mean of a multivariate distribution. The Annals of Statistics 3 (1), 209–218.

Box, G., Hunter, W., Hunter, J., 2005. Statistics for Experimenters: Design, Innovation and Discovery. Wiley.

Brown, L.D., 1966. On the inadmissibility of invariant estimators of one or more location parameters. Annals of Mathematical Statistics 37, 1087–1136.

Brown, L.D., Cai, T.T., DasGupta, A., 2001. Interval estimation for a binomial proportion. Statistical Science 16 (2), 101–133.

Clopper, C.J., Pearson, E.S., 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 26, 404–413.

David, H.A., Nagaraja, H.N., 2003. Order Statistics, 3rd Edition. Wiley.

Devroye, L., Györfi, L., Lugosi, G., 1996. A Probabilistic Theory of Pattern Recognition. Springer.

Duda, R., Hart, P., et al., 1973. Pattern Classification and Scene Analysis. Wiley.

Edelman, B., Ostrovsky, M., Schwarz, M., 2007. Internet advertising and the generalized second-price auction: selling billions of dollars worth of keywords. The American Economic Review 97, 242–259.

Galton, F., 1886. Regression towards mediocrity in hereditary stature. Journal of the Anthropological Institute of Great Britain and Ireland 246–263.

Gelman, A., 2004. Bayesian Data Analysis. CRC Press.

Gittins, J.C., 1979. Bandit processes and dynamic allocation indices. Journal of the Royal Statistical Society. Series B (Methodological) 41 (2), 148–177.

Gomes, R., Immorlica, N., Markakis, V., 2008. Externalities in keyword auctions: an empirical and theoretical assessment. ACM EC Workshop on Advertisement Auctions.

Hoeffding, W., 1963. Probability inequalities for sums of bounded random variables. Journal of the American Statistical Association 58 (301), 13–30.

Hsu, L., Chen, P., 1996. Selecting the best population, provided it is better than a control: the unequal variance case. Biometrical Journal 4, 425–432.

James, W., Stein, C., 1961. Estimation with quadratic loss. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability 1, 361–379.

Jobson, J.D., Korkie, B., Ratti, V., 1979. Improved estimation for Markowitz portfolios using James–Stein type estimators. Proceedings of the American Statistical Association, Business and Economics Statistics Section 41, 279–284.

Jorion, P., 1986. Bayes–Stein estimation for portfolio analysis. Journal of Financial and Quantitative Analysis 279–292.

Kempe, D., Mahdian, M., 2008. A cascade model for externalities in sponsored search. ACM EC Workshop on Advertisement Auctions.

Krishna, V., 2002. Auction Theory. Academic Press.

Lahaie, S., Pennock, D.M., 2007. Revenue analysis of a family of ranking rules for keyword auctions. ACM Conference on Electronic Commerce, pp. 50–56.

Langford, J., 2005. Tutorial on practical prediction theory for classification. Journal of Machine Learning Research 6, 273–306.

Lintner, J., 1965. The valuation of risk assests and the selection of risky investments in stock portfolios and capital budgets. The Review of Economics and Statistics 47 (1), 13–39.

Milgrom, P., 2004. Putting Auction Theory to Work. Cambridge University Press.

Mnih, V., Szepesvari, C., Audibert, J.-Y., 2008. Empirical Bernstein stopping. Proceedings of the 25th International Conference on Machine Learning, pp. 672–679.

Myerson, R.B., 1981. Optimal auction design. Mathematics of Operations Research 6 (1), 58–73.

Ross, T.D., 2003. Accurate confidence intervals for binomial proportion and poisson rate estimation. Computers in Biology and Medicine 33, 509–531.

Samuels, M., 1991. Statistical reversion toward the mean: more universal than regression toward the mean. The American Statistician 344–346.

Shawe-Taylor, J., Cristianini, N., 2004. Kernel Methods for Pattern Analysis. Cambridge University Press.

Stein, C., 1955. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proceedings of the Third Berkeley Symp. on Prob. and Stat. 1, 197–206.

Thaler, R.H., 1988. Anomalies: the winner's curse. The Journal of Economic Perspectives 2 (1), 191–202.

Valiant, L., 1984. A theory of the learnable. Communications of the ACM 27, 1134–1142.

Vapnik, V., 1998. Statistical Learning Theory. Wiley.

Vapnik, V.N., Chervonenkis, A.Y., 1971. On the uniform convergence of relative frequencies of events to their probabilities. Theory of Probability and its Applications 16 (2), 264–280.

Varian, H.R., 2006. Position auctions. International Journal of Industrial Organization 25, 1163–1178.

Varian, H.R., 2009. Online ad auctions. The American Economic Review 99, 430–434.

Wilcox, R., 1984. Selecting the best population, provided it is better than a standard: the unequal variance case. Journal of the American Statistical Association 79 (388), 887–891.

Wilson, E.B., 1927. Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association 22, 209–212.

Wilson, R.B., 1969. Competitive bidding with disparate information. Management Science 15 (7), 446–448.