Modeling the Price-Demand Relationship Using Auction Bid Data

Alex Zhang, Dirk Beyer, Julie Ward, Tongwei Liu, Alan Karp, Kemal Guler, Shailendra Jain, Hsiu-Khuern Tang
Software Technology Laboratory
HP Laboratories Palo Alto
HPL-2002-202
July 16th , 2002*

pricing, demand function, auction, estimation

This report documents two mathematical models for extracting the price-demand relationship, commonly known as the demand function or demand curve, for consumer products using on-line auction bid data. The first model assumes that all bidding sequences are equally likely; the second model assumes that the bidders arrive in Poisson streams and utilizes the time of bid information. Our goal is to estimate, from recorded bids in *n* auctions, the proportions of potential (in addition to the known) bidders who are willing to bid a price.

* Internal Accession Date Only
Approved for External Publication

Contents

1 Introduction

The concept of demand-price relationship is fundamental to economic theories, in which the form and specification of the "demand function" is almost always assumed to be explicitly known. And yet the empirical determination of the demand function for any given product has been difficult. Product managers and marketing executives are constantly faced with the question, "If I cut my price on Product X by 10%, what would be my volume increase?" Managers in many different contexts whom we have talked to seem to rely on their personal intuition, experience and judgment in their pricing work, and seem to be reactive to competitors' pricing moves, rather than proactive in leading the market pricing.

Data Availability and Granularity. Existing approaches to figuring out the demand function include econometric estimation models and procedures. However, these approaches are mainly developed for macro-economic (economy-wide) studies. More micro-economic oriented studies suitable for adoption by a single firm in an industry, on the other hand, require extensive sets of data, which include competitors' prices, sales volume and advertising expenditure, in order to achieve a reasonable statistical accuracy. The availability of data might seem a non-issue for a modern enterprise operating computerized data collection systems; however, we have found that it is not the data from the enterprise itself, but the data from competitors that has made the existing approaches difficult to apply in practice. Often times, competitive information is available through third-party market research firms or industry associations; these data sets, however, usually are available only in aggregate levels (aggregated by time, such as monthly and quarterly sales, or by product models, such as all inkjet printers and all laser printers), making them inappropriate for the level of data atoms required in these approaches. We have experimented with econometric models without the competitive information; the resulting level of statistical significance and predictive power was generally disappointing in our trials.

Practical Approaches. Many managers have adopted their own practical approaches for figuring out the price-demand relationship — through formal or informal consumer surveys (questionnaires, polls, etc.) and through real experiments conducted in the form of limited-time offers of discounts. These practical approaches are in general costly to plan and execute (giving out 15% discounts for the purpose of finding out what would happen to demand can be expensive, even after considering the increased unit volume for the discounts), are time-consuming (involving time to change price labels and time to advertise the sales), with questionable accuracy of price-demand relationship, as many variables such as competing stores' responses during the survey or real experiment and even the local weather and traffic conditions are uncontrolled for.

Our Idea. Auction websites such as eBay.com and Amazon.com have opened a new source of data. If our goal is to obtain a demand-price relationship of a product, then going directly to individual buyers of this product for price-demand information seems to be a viable approach. Indeed, since auction websites capture and display a host of useful information such as auction format and individual bids from each bidder (eBay, for example, displays a "bid history" which includes each bidder's identity, dollar amount of bid, time of bid, and units desired), we might be able to glean for information on willingness-to-pay, and hence on the demand-price relationship.

Data Censoring. It is dangerous, however, to report the bids in some aggregated terms as a demand function without a careful examination of the underlying auction mechanism and the bidder behavior. A key difference between a traditional, off-line (in-person) auction and a web-based on-line auction is that not all bidders in a web auction would be present at the start of the auction; many bidders at a web auction "arrive" after the auction has already started. Indeed, many bidders find out about a particular web auction after they performed a search to find the auction, or were informed by the web host's e-mail notification. The average web auction lasts significantly longer (ranging from 3 days to 10 days on eBay, for example) than a typical off-line auction; the longer duration of the on-line auction is to account for the random arrivals of potential bidders.

The random, sequential arrival of bidders leads to omission of some bids which might have been recorded in an off-line auction but will not be registered in an on-line auction. For example, Mr. Smith is interested in buying a particular HP inkjet printer model and is willing to pay at most $100 for it. When Mr. Smith finds out about an auction for the printer model on eBay, the bidding has already started, and has now progressed to $105 (the "current price"). Disappointed, Mr. Smith does not submit his bid, which would have been $100 if he had joined the bidding a little earlier. EBay's "bid history" will not record Mr. Smith's $100 bid.

We call this phenomenon "self-censoring of bids" (or *truncation* in the standard econometric jargon) in an on-line auction. Since our goal is to determine the number of **all** buyers who are willing to pay for the product at each price level (such as $100), including Mr. Smith who would be willing to pay $100 but was not recorded in the bid history, we will need to recover the missing bids.

Related Literature. Auctions have been studied extensively in the economics literature; for an overview of auction theory and its historical development, we refer to the survey paper by Klemperer (1999). The subject of empirical analysis of auction data has recently received considerable attention; Perrigne and Vuong (1999) contains a survey of methods in estimating value distributions of first-price auctions through structural econometrics. These methods generally deal with issues in valuation dependency (e.g. common value and private value), information (a)symmetry, and bidder risk attitude.

Our models differ from these econometric models in that we focus on one particular aspect: that of random times of bids, and ignore the aspects of bidder valuation and bidders' strategic behaviors. Our goal is to estimate the bid distribution from a partially observable (recorded) bid history, rather than estimating the value distribution from the full set of bids. Hence, in assuming that bidders will bid their true willingness-to-pay, we are able to accomplish our goal without imposing a structural model which entails a host of assumptions on valuation dependency, information (a)symmetry, and risk attitude.

Particularly on using eBay auction data, there are several recent papers. Lunking-Reiley et al (1999) perform an empirical study of the determinants of price for collectible coins in over 20,000 eBay auctions between July and August 1999. They found that the seller's feedback rating (reported by other buyers participating on eBay auctions), the duration of the auction, and the presence of a reserve price and a higher starting bid, all have a positive impact on the final price reached in the auction. Bajari and Hortacsu (2000) discuss endogenous entry in eBay auctions in accounting for the valuations of bidders who decided not to bid in a particular auction because they found the reserve price too high; Bajari and Hortacsu study the case of common value and examine the collectible coin auctions on eBay.

Like the above empirical papers on eBay auctions, this paper also uses eBay-like data. However, we differ in that this paper explicitly considers a temporal entry process. Our focus is on the effect of random timing of bidding on the observed bids.

Our General Approach. We will attempt to recover the censored bids by models of the auction bidding process. In this report, we describe two models on bidder arrival process. The first model assumes that all possible bidder sequences are equally likely; the second model assumes that bidders at each price level arrive at the auction based on a Poisson process.

Our emphasis is on the inference of the truncated bids in the continually arrived bid streams; we will ignore the strategic implications of bidding and gaming behavior for each individual bidder by making an important assumption that each bidder will bid exactly his true valuation (or willingness-to-pay).

Our general approach is two-step: First assuming that we know the number of bidders at each price level, we use the model to determine the statistical properties of the observed bid history. Second, given a sample of bid histories (say, $n = 100$), we use statistical inference to "reverse engineer" the proportion of the underlying bidder pools at each price level.

In both models, we discrete the prices so that bids are represented by integers, which allow for a concise expression of the results, but do not materially hinder the applicability of the models in a world of continuous prices. The results from these models are implemented in computer code to produce a numerical list and a graph that represents the demand function.

Organization of This Report. In Section 2, we present the equally likely bidder sequences model. We show a number of interesting properties of the bidding model, and give a maximum likelihood estimator of the relative demand function; i.e. the proportion of all bidders who (potentially) want to bid a certain price level. In Section 3, we present the Poisson bidding stream model, and again derive a maximum likelihood estimator for the relative demand, i.e. the arrival rate of bidders who would bid a price level as a percentage of the total arrival rate of all bidders. We conclude and point out future research directions in Section 4.

2 The Equally Likely Bidder Sequences Model

2.1 Model Description and Assumptions

In this model, we assume that there is a fixed number of potential bidders (N), say bidders A, B, C, D and E for $N = 5$, each willing to pay a known price, which ranges from 1, 2, to K (we have used integers as prices; these could be replaced by $p_1, p_2, \ldots,$ p_K). The total number of bidders who are willing to pay price i is m_i, with $m_i \geq 0$ known and $m_1 + m_2 + \cdots + m_K = N$.

Depending on who arrives first, second, and last, the recorded bid history would be different. Given N bidders, there are $N!$ (N factorial) number of possible sequences of bidder arrivals (such as ABCDE, ACDEB, or EDCBA). For example, if bidder A is willing to pay the price of 1, B is 2, C is 3, D is 4, and E is 5, then the sequence ADEBC would result in a recorded bid history of bidders ADE and prices 145 (B and C will not be recorded). Define the *bid history length* as the number of bidders recorded in the bid history, then the bid history length of the above sequence ADEBC would be 3.

Our questions are:

1. What is the probability of a price i being recorded in the bid history?
2. What is the probability distribution of the bid history length?

To simplify the mathematical derivation, we make the following assumptions:

(A1) (Equally likely sequences) Each sequence of bidder arrivals is equally likely.

(A2) (True valuation bid) Each bidder will bid exactly (not above or below) his or her true willingness-to-pay.

(A3) (No-wait bid) Each bidder will attempt to submit his bid as soon as he arrives at the auction; he will not "wait and see" (until the last minute of the scheduled auction ending time, for example) to submit his bid.

(A4) (Known current high bid) The currently highest bid (but not necessarily the bidder identity) is announced to all bidders.

(A5) (First price auction) The winner will pay what he bids.

(A6) (Ascending price auction) A bid is recorded only if it is strictly greater than the previous bid.

(A7) (Zero starting bid) The first bid is required to be greater than zero.

(A8) (Negligible bid increment) Any bid that is strictly greater than the currently highest bid is recorded; the required increment of the new bid over the current high bid is assumed to be negligible.

Note that our assumption on bidder behavior (A2) implies that each bidder will submit at most one bid; (A2) might be the most restrictive assumption as the rest of the assumptions are simply rules of the auction. The above auction rules (A2) through (A8) do not describe a typical eBay auction; we will discuss in a later section how to convert an eBay auction bid history to one described above. We will attempt to relax some of these assumptions later so that our results can be applied to situations closer to real eBay auctions.

2.2 Probability That Price i Is Recorded in Bid History

Let us first consider the probability of a particular bid price i being recorded. The following is a key result.

Lemma 1.

$$\Pr\{\text{Price } i \text{ is recorded in bid history}\} = \frac{m_i}{m_i + m_{i+1} + \cdots + m_K}.$$

Proof. We note that price K (the highest bid of the N bidders) will appear in the bid history with probability 1.

Consider the bidder group i which has a bid price i. Think about the first bidder in bidder group i to arrive at the auction. None of the other bidders from group i will be recorded in the bid history, by auction rule (A5). This bidder can arrive at the auction as the first in sequence, second in sequence, third in sequence, etc.

If he is the first in the bidder sequence, then he will be included in the bid history. The number of sequences where the first in the sequence is a bidder from group i is $m_i(N-1)!$ (recall that m_i is the number of bidders who are willing to bid price i).

If he is the second bidder in the bidder sequence, then he will be included in the bid history if and only if the first bidder in the sequence is from groups $i-1$ or lower. The number of such sequences is $(m_1 + m_2 + \cdots + m_{i-1}) \cdot m_i \cdot (N-2)!$.

If he is the third bidder in the bidder sequence, then he will be included in the bid history if and only if the first and second bidders in the sequence is from groups $i-1$ or lower. Then number of such sequences is $(m_1 + m_2 + \cdots + m_{i-1}) \cdot m_i \cdot (N-3)!$.

Let $M_i \equiv m_1 + m_2 + \cdots + m_i$.

If he is the $(M_{i-1}+1)$-st bidder in the bidder sequence, then he will be included in the bid history if and only if all the first M_{i-1} bidders are from groups $i-1$ or lower. The number of such sequences is $M_{i-1} \cdot (M_{i-1}-1) \cdot 2 \cdot 1 \cdot m_i \cdot (N - M_{i-1} - 1)!$.

Beyond the $(M_{i-1}+1)$-st position, there will be a bidder from a higher price group to arrive before the first bidder from group i.

So, the total number of sequences in which a bidder from group i will appear before any bidder from a higher group is

$$m_i(N-1)! + M_{i-1}m_i(N-2)! + M_{i-1}(M_{i-1}-1)m_i(N-3)! + \\ + \cdots + (M_{i-1})!m_i(N-M_{i-1}-1)!$$

The total number of bidder sequences is $N!$, so the probability that price i will appear in the bid history is the above expression divided by $N!$.

Utilizing the identity (for proof please see Appendix A)

$$\sum_{j=1}^{M+1} \frac{(N-j)!}{(M-j+1)!} = \frac{N!}{M!(N-M)}$$

we get

$$\Pr\{\text{Price } i \text{ is in bid history}\} = \frac{m_i}{N - M_{i-1}} = \frac{m_i}{m_i + m_{i+1} + \cdots + m_K}.$$

Clearly, when $i = K$, the above probability is equal to 1. □

Corollary 1. The probability that a specific bidder whose price is i will be recorded in the bid history is

$$\Pr\{\text{A particular bidder with price } i \text{ will be recorded in bid history}\} = \frac{1}{m_i + \cdots + m_K}.$$

2.3 Probability Independence

Now consider any two bid prices i and j. We want the probability that both prices are recorded in the bid history. The following is a powerful result which we will utilize later to derive the mean and variance of the demand function.

Lemma 2. For prices $i < j$,

$$\begin{aligned}&\Pr\{\text{Prices } i \text{ and } j \text{ are both recorded in bid history}\} = \\ &\quad \Pr\{\text{Price } i \text{ is recorded in bid history}\} \cdot \\ &\quad \cdot \Pr\{\text{Price } j \text{ is recorded in bid history}\}.\end{aligned}$$

Proof. Consider the conditional probability

$$\Pr\{\text{Price } j \text{ is recorded in bid history} \,\Big|\, \text{Price } i \text{ is recorded in bid history}\},$$

we shall show that this is equal to $\Pr\{\text{Price } j \text{ is recorded in bid history}\}$. Suppose the first bidder from price group i is the k-th bidder in the bidder sequence. Now fix k and consider the probability that price j will be recorded in the bid history. The probability can now be thought of as the problem with new parameters $(m_1, \cdots, m_{i-1}, m_i - 1, m_{i+1}, \cdots, m_K)$ (the size of price group i is reduced by 1 while all other price groups remain the same). But in this new problem, the probability of price j being recorded is still $m_j/(m_j + \cdots + m_K)$, independent of price i and position k. Hence,

$$\begin{aligned}\Pr\{\text{Price } j \text{ is recorded in bid history} \,\Big|\, \text{Price } i \text{ is recorded in bid history}\} \\ = \Pr\{\text{Price } j \text{ is recorded in bid history}\}.\end{aligned}$$

□

The above results can easily be extended to 3-dimensional and n dimensional:

Lemma 3. Given n prices $p_1, p_2, p_3, \cdots, p_n$,

$$\begin{aligned}\Pr\{\text{Prices } p_1, p_2, \cdots, p_n \text{ appear in bid history}\} = \\ = \prod_{j=1}^{n} \Pr\{\text{Price } p_j \text{ appears in bid history}\}.\end{aligned}$$

2.4 The Probability of Observing a Recorded Bid History

We now turn to the final question in this series of looking at recorded bids: Given a bid history containing a subset of prices (say $A = \{1, 3, 4, K\}$) in an auction with bidder groups $m_1, m_2, \cdots, m_K$ (only prices in A are recorded; i.e. prices not in A are not recorded), what is the probability of observing such a recorded bid history?

Theorem. Denote by A the subset of prices recorded in the bid history. Then the probability of observing a bid history containing exactly the set A of bid prices is

$$P(A) = \prod_{i \in A} P(i) \cdot \prod_{j \notin A} [1 - P(j)]$$

where $P(i) = \dfrac{m_i}{m_i + m_{i+1} + \cdots + m_K}$.

For example, consider 5 bid groups A, B, C, D and E, each willing to bid prices 1, 2, 3, 4 and 5. There are 2 bidders in group A, 2 bidders in group B, and 1 bidder in the other three groups. Then we have $m_1 = m_2 = 2$, $m_3 = m_4 = m_5 = 1$. The probability of observing a recorded bid history of bid prices 1, 3 and 5 would be $P(1, 3, 5) = P(1) \cdot P(3) \cdot P(5) \cdot [1 - P(2)] \cdot [1 - P(4)] = (2/7)(1/3)(1/1)[1 - (2/5)][1 - (1/2)] = 2/70 = 0.02857$.

Proof. Denote $P(i, j)$ as the probability of observing recorded i and j, and $P(i, \bar{j})$ as the probability of observing recorded i but not j. Then we have $P(i, j) + P(i, \bar{j}) = P(i)$. From Lemma 2 and Lemma 3, we have $P(i, \bar{j}) = P(i) - P(i, j) = P(i) - P(i)P(j) = P(i)[1 - P(j)]$. Extending this argument for all bids in A, we obtain the result. □

Finally, we point out that one can obtain the probability distribution of the length of the bid history (i.e. number of bids recorded in an auction log); for details, see Appendix B.

2.5 Reverse Engineering: Solving for Bidder Group Sizes

In previous sections, we discussed Step 1 of our two-step process: Assuming that we know the population characteristics (m_1, m_2, ..., m_K), what is the characteristics of the resulting bid history (probability of price i appearing in the bid history, and the probability distribution of the bid history length). In this section, we describe Step 2, in which we observe a sample of bid histories and solve for the population characteristics.

The idea we use here is to relate the probability of price i appearing in a bid history to the relative frequency of price i counted in a sample of bid histories. Let the relative frequency of the price i appearing in the sample be f_i ($0 \le f_i \le 1$). Then from the results of Section 2.2, we have the following system of the simultaneous equations (total of K equations):

$$\begin{aligned}
\frac{m_1}{m_1 + m_2 + \cdots + m_K} &= f_1; \\
\frac{m_2}{m_2 + \cdots + m_K} &= f_2; \\
&\cdots \\
\frac{m_{K-1}}{m_{K-1} + m_K} &= f_{K-1}; \\
m_1 + m_2 + \cdots + m_K &= 1.
\end{aligned}$$

The last equation replaces the redundant equation $m_K/m_K = f_K = 1$ and is used to normalize the bid group sizes. The solution will be closed-form as follows:

$$\begin{aligned}
m_1 &= f_1; \\
m_2 &= f_2(1 - f_1); \\
m_3 &= f_3(1 - f_2)(1 - f_1); \\
m_4 &= f_4(1 - f_3)(1 - f_2)(1 - f_1); \\
&\cdots \\
m_i &= f_i \prod_{j=1}^{i-1} (1 - f_j), \quad i = 2, 3, \cdots, K - 1; \\
&\cdots \\
m_K &= \prod_{j=1}^{K-1} (1 - f_j).
\end{aligned}$$

This solution is simple enough to allow easy implementation in an Excel spreadsheet.

2.6 Demand Function Under the Basic Model

The "demand function" $D(p)$ is the number of bidders who are willing to pay a price p. The relative demand function is the percentage of all bidders who are willing to pay a price p. Therefore, for integer p between 1 and K,

$$D(p) = m_p + m_{p+1} + \cdots + m_K = m_p/f_p = \prod_{j=1}^{p-1} (1 - f_j), \quad p = 2, 3, \ldots, K;$$

with $D(1) = 1$.

Strictly speaking, for a given p, the above $D(p)$ is a random variable as $\tilde{f}_j$, $j = 1, 2, \cdots, p-1$, are random variables. Utilizing the independence result from Section 2.3, we have that the estimated demand function is given by

$$E[D(p)] = E\left[\prod_{j=1}^{p-1}(1-\tilde{f}_j)\right] = \prod_{j=1}^{p-1} E[(1-\tilde{f}_j)] = \prod_{j=1}^{p-1}(1-\bar{f}_j)$$

where $\bar{f}_j \equiv E[\tilde{f}_j]$.

We next derive the variance of $D(p)$. The variance will enable us to construct a lower and upper confidence level for the estimated $D(p)$.

Variance Result. The variance of $D(p)$ is given by

$$\begin{aligned}
\mathrm{Var}[D(1)] &= 0, \\
\mathrm{Var}[D(2)] &= f_1(1-f_1)/S, \\
\mathrm{Var}[D(p)] &= \mathrm{Var}\left[\prod_{j=1}^{p-1}(1-\tilde{f}_j)\right] = \prod_{j=1}^{p-1}\left[(1-f_j)^2 + f_j(1-f_j)/S\right] - \prod_{j=1}^{p-1}(1-f_j)^2, \quad p > 1,
\end{aligned}$$

where, for notational convenience, we have written $\bar{f}_j$ as f_j, and S is the number of auctions (sample size of bid histories).

Proof. $D(1) = 1$ is a constant, so $\mathrm{Var}[D(1)] = 0$. Consider $D(2) = 1 - \tilde{f}_1$. $\mathrm{Var}[D(2)] = \mathrm{Var}[\tilde{f}_1]$. For $j = 1, 2, \ldots,$, $\tilde{f}_j$ is a binomial random variable with number of trials S and probability of success f_j. Hence, $\mathrm{Var}[D(2)] = \mathrm{Var}[\tilde{f}_1] = \sqrt{f_1(1-f_1)/S}$.

Now consider $\mathrm{Var}[D(p)]$.

$$\begin{aligned}
\mathrm{Var}[D(p)] &= \mathrm{Var}\left[\prod_{j=1}^{p-1}(1-\tilde{f}_j)\right] \\
&= E\left[\left\{\prod_{j=1}^{p-1}(1-\tilde{f}_j)\right\}^2\right] - \left\{E\left[\prod_{j=1}^{p-1}(1-\tilde{f}_j)\right]\right\}^2 \\
&= E\left[\left\{\prod_{j=1}^{p-1}(1-\tilde{f}_j)^2\right\}\right] - \left\{\prod_{j=1}^{p-1}E[(1-\tilde{f}_j)]\right\}^2 \\
&= \prod_{j=1}^{p-1}E[(1-\tilde{f}_j)^2] - \left\{\prod_{j=1}^{p-1}(1-f_j)\right\}^2
\end{aligned}$$

$$= \prod_{j=1}^{p-1} \left\{ \left(E[1-\tilde{f}_j]\right)^2 + \text{Var}[1-\tilde{f}_j] \right\} - \prod_{j=1}^{p-1} (1-f_j)^2$$

$$= \prod_{j=1}^{p-1} \left\{ (1-f_j)^2 + f_j(1-f_j)/S \right\} - \prod_{j=1}^{p-1} (1-f_j)^2,$$

where we have repeatedly utilized the independence result that

$$E\left[\prod_{j=1}^{p-1}(1-\tilde{f}_j)\right] = \prod_{j=1}^{p-1} E[(1-\tilde{f}_j)]$$

and the identity $(\prod_j A_j)^2 = \prod_j (A_j^2)$.

Results verified by simulation. We wrote a simulation code (see Section 2.9) of auctions following assumptions (A1) through (A8) using Excel, with 5 bidders (A, B, C, D, and E) and number of price segments (K) from 2 to 5. Both the demand function recovery $E[D(p)]$ and the variance estimation $\text{Var}[D(p)]$ has been numerically verified.

2.7 An Alternative Method of Estimating Bidder Group Sizes

Our method described above utilizes the probability of price i appearing in the recorded bid history, in which price i can appear as the first recorded bid, or the second, third, up to i-th bid (recall that the auction rules ensure that the first bid will always be recorded in the bid history). An alternative method is to look at the *first* bid ever recorded in the bid history. The assumption that all bid sequences are equally likely implies that the probability of the first bid being price i is equal to $m_i/(m_1 + m_2 + \cdots + m_K)$. Again equating this probability to the relative frequency of price i appearing as the first bid in the bid history, denoted by $f_i^{(1)}$, we obtain $m_i = f_i^{(1)}$ (after we normalize $m_1 + m_2 + \cdots + m_K$ to 1), with resulting demand function estimate $D^{(1)}(p) = \sum_{i=p}^K m_i = \sum_{i=p}^K f_i^{(1)}$.

To evaluate the alternative method against our previously described method, we compare the variance of $D(p)$ to that of $D^{(1)}(p)$ (as both methods are unbiased estimators). It is immediately clear that both methods produce the same $D(p)$ for $p = 1$ (with $D(1) = 1$). Now consider $p = 2$. Since the lowest price of 1 is recorded *only if* it is the first bid, we have $f_1^{(1)} = f_1$, hence $D(2) = 1 - f_1 = 1 - f_1^{(1)} = D^{(1)}(2)$.

For $p \geq 3$, we now derive the variance of $D^{(1)}(p)$. We divide all possible bids that can potentially be the first bid to arrive into two groups: those that are greater than or equal to p, or less than p. Then $D^{(1)}(p)$ will be the probability that the first bid falls into the first group. It is now clear that $D^{(1)}(p)$ is a binomial random variable with number of trials S (number of all bid histories, our sample size) and probability of success $(m_p + m_{p+1} + \cdots + m_K)/(m_1 + m_2 + \cdots + m_K)$. Hence, the binomial variance is

$$\mathrm{Var}[D^{(1)}(p)] = \frac{(m_1 + m_2 + \cdots + m_{p-1})(m_p + m_{p+1} + \cdots + m_K)}{(m_1 + m_2 + \cdots + m_K)^2} \cdot \frac{1}{S}$$

We can now compare the above variance to that variance given in Section 2.6, numerically. For example, when $N = 5$, $K = 5$, $m_i = 1$ for $i = 1, 2, 3, 4, 5$, and sample size $S = 100$, we get the following result:

Price p	1	2	3	4	5
$\mathrm{Var}[D(p)]^{1/2}$	0	0.0400	0.0459	0.0417	0.0290
$\mathrm{Var}[D^{(1)}(p)]^{1/2}$	0	0.0400	0.0490	0.0490	0.0400

In all the numerical cases that we have run, we see that our method of using the full set of bids recorded yields a lower standard error than the alternative method of using only the first bid recorded. Intuitively, this should be so as our method utilizes more information in the data. We will analytically show that $\mathrm{Var}[D(p)] \leq \mathrm{Var}[D^{(1)}(p)]$, as follows.

In each of S independent auctions, there are m_i bidders with price i, $i = 1, \ldots, K$. The outcome of each auction may be described by a binary vector of length K, with a 1 in the i-th component if price i was recorded, and 0 otherwise. Let X be the $S \times K$ matrix formed from these (row) vectors. We assume that $S > 1$.

Let $M_i = \sum_{j \geq i} m_j$ and $u_i = m_i/M_i$. The likelihood function of X is

$$\prod_{n=1}^{S} \prod_{i=1}^{K} u_i^{X_{ni}} (1 - u_i)^{1 - X_{ni}} = \prod_{i=1}^{K} u_i^{X_{\cdot i}} (1 - u_i)^{S - X_{\cdot i}},$$

where $X_{\cdot i} = \sum_n X_{ni}$ is the i-th column sum. From the form of the likelihood, we conclude that conditionally given $X_{\cdot 1}, \ldots, X_{\cdot K}$, all binary matrices X with these column sums are equally likely. In other words, under this conditional probability, X is a binary matrix with independent columns where the i-th column is a random combination of $X_{\cdot i}$ 1's and $S - X_{\cdot i}$ 0's.

The two estimators of u_p are

$$\hat{D}(p) = (1 - X_{\cdot 1}/S) \cdots (1 - X_{\cdot p}/S)$$

and

$$\hat{D}^{(1)}(p) = S^{-1} \sum_{n=1}^{S} 1_{\{X_{ni}=0 \text{ for } i=1,\ldots,p-1\}}.$$

Since

$$\begin{aligned} & P\{X_{ni} = 0 \text{ for } i = 1, \ldots, p-1 \mid X_{\cdot 1}, \ldots, X_{\cdot K}\} \\ = \; & (1 - X_{\cdot 1}/S) \cdots (1 - X_{\cdot p}/S) = \hat{D}(p), \end{aligned}$$

we conclude that

$$\hat{D}(p) = E[\hat{D}^{(1)}(p) \mid X_{\cdot 1}, \ldots, X_{\cdot K}]$$

and hence $\mathrm{Var}(\hat{D}(p)) \leq \mathrm{Var}(\hat{D}^{(1)}(p))$. The inequality is strict whenever $\hat{D}^{(1)}(p)$ is not a function of $X_{\cdot 1}, \ldots, X_{\cdot K}$, i.e., for $p > 2$.

2.8 Extending the Basic Model (1): Non-Zero Starting Bid

We now relax the assumption (A7) (zero starting bid). Specifically, each auction will now have a known starting bid s which can be 1, 2, or K; the first bid recorded must be equal to or greater than s; any bids lower than s will not be recorded. All other rules still apply.

We note that *not all* auctions in the sampled bid histories have the same starting bid; some auctions might have a zero starting bid, while some others might have a very high starting bid such as $K-1$ or even K (partly because some sellers use the starting bid as a reserve price — in fact, the terminology "starting bid" here is referred to as reserve price, or more exactly, announced and enforced reserve price, in the auction theory literature). If all auctions have the same starting bid, then we could simply re-classify our price set $\{1, 2, \cdots, K\}$ so that the price 1 is equal to or above the starting bid.

Surprisingly (or perhaps not so surprisingly), our basic results still hold with a very minor modification to the tallying of the frequency f_i for price i to appear in the bid histories. Suppose we have a sample of S recorded bid histories. Since the starting bid for each auction is known, we count the number of auctions with a starting bid $s \leq i$ as S_i, $i = 1, 2, \cdots, K$ (with $S_K = S$ — assuming no auctions have starting bids greater than the highest possible bid). Then we redefine the relative frequency of price i appearing in bid histories as

$$f_i = (\text{Number of times price } i \text{ appears in the bid histories})/S_i,$$

all our previous results on $D(p)$ still hold.

To prove this extension result, we look at an auction with a starting bid s. Any bids i less than s will not be recorded (will be censored); any bid i that is greater than s will appear with probability $m_i/(m_i+m_{i+1}+\cdots+m_K)$ (see the results in Section 2.2), which is independent of m_j for all $j < s$. That is, to bidders with a bid i greater than or equal to s, the auction will appear to have zero starting bid, while to bidders with a bid lower than s, the auction simply does not exist.

To summarize the results so far, we have

$$\begin{aligned} E[D(p)] &= \prod_{j=1}^{p-1}(1-f_j), \quad p > 1; \\ \mathrm{Var}[D(p)] &= \prod_{j=1}^{p-1}\left[(1-f_j)^2 + f_j(1-f_j)/S_j\right] - \prod_{j=1}^{p-1}(1-f_j)^2; \quad p > 1, \end{aligned}$$

with $E[D(1)] = 1$ and $\mathrm{Var}[D(p)] = 0$.

The effect of the starting bids is that the variance of $D(p)$ for lower p values is increased.

Again, the above results have been verified by our simulation code (5 bidders).

2.9 Simulation Code to Verify Analytical Results

We wrote a simple ($N = 5$ bidders) code on the Microsoft Excel platform to simulate the sampling of bid histories and reverse engineering the demand function $D(p)$. The following is a brief description of the simulation code.

Step 0: Initialization and preparation. Manually prepare the list of all $5! = 120$ possible sequences of the 5 bidders (ABCDE, ABCED, ..., EDCBA). Assign a sequence number, from 1 to 120, to each sequence. Assign a price between 1 and 5 (if $K = 5$; K can be chosen to be 2,3,4, or 5) to each bidder.

Step 1: Generate a (uniform, or equally likely) random number on the set of integers between 1 and 120 (using the Excel function RANDBETWEEN(1,120)); check this random number against the master list of all 120 possible sequences, and retrieve the sequence that corresponds to the sequence number thus generated. Also, generate a random starting bid from the set of integers $\{1, 2, \cdots, 5\}$ (using the Excel function RANDBETWEEN(1,5) or some other desired statistical patterns).

Step 2: Generate the bid history given the bidder sequence. For each bidder in the bidder sequence, starting from the first bidder, its price will be recorded if the price is strictly greater than all the previous bids AND greater than or equal to the starting bid.

Some bid history will contain only 1 bid (price 5, if the bidder with the highest bid happens to be the first one in the bidder sequence), some bid history will contain all $K = 5$ bids (if the bidders arrive based on sorted order of their bid prices)

Step 3: Repeat Steps 1 and 2 for $S = 25$ times (or $S = 100$, where S is the sample size of bid histories). We obtain a list of S bid histories.

Step 4: Summary statistics. On the list of S bid histories, for each price $i = 1$ through 5, we count the number of times the price i occurred, and the number of auctions where the starting bid is less than or equal to i (this number is S_i). Divide the two numbers, we obtain the relative frequency f_i.

Step 5: Reverse engineering. For $p = 1$ through K, output $E[D(p)]$ and $\text{Var}[D(p)]$ following the equations in Section 2.8.

Step 6: Verification of $E[D(p)]$. Check the output $E[D(p)]$ against the originally assigned prices to the 5 bidders. For example, if A,B,C,D,E is assigned prices 1,2,3,4,5 respectively, then the originally assigned demand function is $D(1) = 1$, $D(2) = 0.8$, $D(3) = 0.6$, $D(4) = 0.4$, $D(5) = 0.2$. The reverse engineered $E[D(p)]$ might be 1, 0.805, 0.597, 0.402, 0.201, respectively for $p = 1, 2, 3, 4, 5$.

Step 7: Verification of $\text{Var}[D(p)]$. We derive the variance by repeating Steps 1 through 5, say, 250 times. Each of the 250 times produces an estimated vector ($E[D(1)]$, $E[D(2)]$, $E[D(3)]$, $E[D(4)]$, $E[D(5)]$). The variance among the 250 $E[D(1)]$ values is the standard error of $E[D(1)]$; and so on for $E[D(2)]$, $E[D(3)]$, $E[D(4)]$, $E[D(5)]$. We check these 5 variances against the reverse engineered $\text{Var}[D(p)]$.

2.10 Extending the Basic Model (2): English Auctions or Second-Price Auctions

Based on our first extension of the basic model to auctions with starting bids, we now extend further to auctions more resembling the eBay auction. The basic assumptions

(A4) through (A7) are now changed as follows:

(A4) (Known current second-highest bid) The currently second-highest bid (but not necessarily the bidder identity) is announced to all bidders.

(A5) (Second price auction) The winner will pay what the second-highest bidder bids.

(A6) (Ascending price auction) A bid is recorded only if it is strictly greater than the previous second-highest bid.

(A7) (Non-Zero starting bid) The first bid is required to be equal to or greater than a known starting bid.

Note that Assumption (A8) (negligible bid increment) still holds here. For non-zero bid increment, Assumption (A5) can be modified to "the winner will pay what the second-highest bidder bids *plus the bid increment*" as eBay auction rules stipulate. Also, on eBay auctions, the winner may actually pay what he bids; this occurs when the second-highest bidder happens to bid the same price as the winner (eBay rules stipulate that in case of such a tie, the earlier bidder wins).

We note that in this type of auction, the bids recorded might not be increasing over time; i.e. a lower bid than a previous bid might actually be recorded as long as it is greater than the previous second-highest bid. In the first-price auction, we would always observe increasing bids.

There are essentially two approaches to handle this new situation. One approach is to start a new model specifically for the second-price auction. Another approach is to first convert the second-price bid history into an equivalent first-price bid history, then apply our previously developed models and results to it. We opted for the second approach.

The Bid History Conversion Algorithm. Given a second-price bid history of length n, denoted by $(b_1, b_2, \cdots, b_n)$ sorted by time of bid, we convert it into a first-price bid history $(b'_1, b'_2, \cdots, b'_\ell)$ by the following procedure:

1. Step 0. Set $\ell = 1$; set $b'_1 = b_1$.

2. Step 1. For $i = 2$ to n: If $b_i > b'_\ell$ Then ($\ell = \ell + 1$; $b'_\ell = b_i$); Next i.

3. Step 2. Output ℓ and history $(b'_1, b'_2, \cdots, b'_\ell)$.

Again, the above reverse engineering approach to obtain $E[D(p)]$ and $\mathrm{Var}[D(p)]$ has been verified by a simulation code.

3 The Poisson Arrival Model

3.1 Model Description and Assumptions

Consider bidders arriving at the auction site one at a time, with inter-arrival times following the statistical distribution of negative exponential, which is used extensively in the theories of queues and telecommunications. We shall call this arrival process by "Poisson arrivals", as the number of bidders to arrive during any fixed time interval follows the Poisson distribution.

Instead of assuming fixed and known sizes of price segments (bidder pools), we work with *arrival rate*, denoted by λ_i, for bidders who are willing to bid price i.

Again, we make the following basic assumptions, and will attempt to relax or extend some assumption in later sections.

(B1) (Independent arrival streams) The arrival process of bidders with bid price i is independent of the arrival processes of all other bidders with bid prices.

(A2) (True valuation bid) Each bidder will bid exactly (not above or below) his or her true willingness-to-pay.

(A3) (No-wait bid) Each bidder will attempt to submit his bid as soon as he arrives at the auction ; he will not "wait and see" (till the last minute of the scheduled auction ending time, for example) to submit his bid.

(A4) (Known current high bid) The currently highest bid (but not necessarily the bidder identity) is announced to all bidders.

(A5) (First price auction) The winner will pay what he bids.

(A6) (Ascending price auction) A bid is recorded only if it is strictly greater than the previous bid.

(A7) (Zero starting bid) The first bid is required to be greater than zero.

(A8) (Negligible bid increment) Any bid that is strictly greater than the currently highest bid is recorded; the required increment of the new bid over the current high bid is assumed to be negligible.

We will first present some analytical results on the characteristics of the bid histories, then discuss how we can derive the demand function through a statistical estimation procedure. The organization is similar to the section on the fixed population model.

3.2 Probability That Price i Is Recorded in Bid History

A convenient approach is to think in terms of bidders segmented by their bid price. We call a bidder who is willing to bid price i as a bidder of type i (or segment i).

Define additional notation: X_i as the arrival time of the first bidder of type i. Then X_i follows an exponential distribution with mean $1/\lambda_i$ (recall that λ_i is the arrival rate of bidders of type i). Let T be the duration of the auction (known and fixed before the auction starts). Then

$$\begin{aligned}
& \Pr\{\text{Price } i \text{ is recorded in bid history}\} \\
= \; & \Pr\{\text{First bidder of type } i \text{ arrives before any bidders of types } i+1 \text{ and higher} \\
& \quad \text{AND first bidder of type } i \text{ arrives before auction ends}\} \\
= \; & \Pr\{X_i < \min_{j>i} X_j \text{ AND } X_i \leq T\}
\end{aligned}$$

Before proceeding further, denote $\sigma_i \equiv \sum_{j=i}^{K} \lambda_j$ as the aggregate arrival rate of bidders willing to pay price i or higher. From a basic results on the exponential distribution, $\min_{j>i} X_j$ is another exponentially distributed random variable with rate $\sum_{j>i} \lambda_j$. Conditioning on the arrival time X_i, we obtain

$$\begin{aligned}
& \Pr\{\text{Price } i \text{ is recorded in bid history}\} \\
= \; & \int_{x=0}^{\infty} \Pr\{X_i < \min_{j>i} X_j \text{ AND } X_i < T \;\Big|\; X_i = x\} \cdot \lambda_i e^{-\lambda_i x} \, dx \\
= \; & \int_{x=0}^{T} \Pr\{X_i < \min_{j>i} X_j \;\Big|\; X_i = x\} \cdot \lambda_i e^{-\lambda_i x} \, dx \\
= \; & \int_{x=0}^{T} e^{-\sum_{j>i} \lambda_j x} \cdot \lambda_i e^{-\lambda_i x} \, dx
\end{aligned}$$

$$= \int_{x=0}^{T} \lambda_i e^{-\sigma_i x}\, dx$$
$$= \frac{\lambda_i}{\sigma_i}\left[1 - e^{-\sigma_i T}\right]$$

Unlike the previous fixed population model, in which the highest recorded bid is always K, in the Poisson arrival model we might observe a random highest bid recorded. The probability that price i is recorded as the highest bid is as follows:

$$\Pr\{\text{Highest bid recorded is } i\} = \Pr\{X_i \le T \text{ AND } \min_{j>i} X_j > T\}$$
$$= \int_{x=0}^{\infty} \Pr\{X_i \le T \text{ AND } \min_{j>i} X_j > T \,\big|\, X_i = x\} \cdot \lambda_i e^{-\lambda_i x}\, dx$$
$$= \int_{x=0}^{T} e^{-\sum_{j>i} \lambda_j T} \cdot \lambda_i e^{-\lambda_i x}\, dx$$
$$= e^{-(\sigma_i - \lambda_i)T}(1 - e^{-\lambda_i T}).$$

It is clear that as $T \to \infty$, the above probability approaches 1 if $i = K$ (as $\sigma_i - \lambda_i = 0$), and approaches 0 if $i < K$.

3.3 Maximum Likelihood Estimation Procedure

We now propose a statistical procedure to estimate the Poisson arrival rates $\lambda_1, \lambda_2, \cdots, \lambda_K$. We will derive the Maximum Likelihood Estimator (MLE). Let X_i be the time of arrival of the first bidder from price segment i. Then we have X_i is exponentially distributed with rate λ_i.

3.3.1 Two Bidders

Auction Outcome. When the auction is completed, we will observe the following: BidderID, Bid Amount (i), and Time of Bid (X_i). In addition, we observe the duration of the auction, T. There could be no bidders, one bidder, two bidders, and more than two bidders in the recorded bid history. For the sake of exposition, let's assume that there are no more than two potential bidders. The mathematical derivation and result applies to any number of bidders, as we will show subsequently.

Outcome 2: Exactly two bidders recorded, with bid amount i and j ($i < j$), and time of bid x_i and x_j ($x_i < x_j$), respectively. The likelihood for this situation to occur is

$$\Pr\{x_i < X_i \le x_i + dx_i, x_j < X_j \le x_j + dx_j\} = \lambda_i e^{-\lambda_i x_i}\, \lambda_j e^{-\lambda_j x_j}\, dx_i\, dx_j.$$

Outcome 1(i): Exactly one bidder is recorded, and it is bid i at time x_i. The likelihood for this to occur is

$$\Pr\{x_i < X_i \le x_i + dx_i, X_2 > T\} = \lambda_i e^{-\lambda_i x_i}\, dx_i\, e^{-\lambda_2 T}.$$

Outcome 1(j): Exactly one bidder is recorded, and it is bid j at time x_j. The likelihood for this to occur is

$$\Pr\{X_i > X_j, x_j < X_j \le x_j + dx_j\} = \Pr\{X_i > x_j, x_j < X_j \le x_j + dx_j\} = e^{-\lambda_i x_j}\, \lambda_j e^{-\lambda_j x_j}\, dx_j.$$

Outcome 0: No bidder is recorded. The likelihood for this to occur is

$$\Pr\{X_i > T, X_j > T\} = e^{-\lambda_i T}\, e^{-\lambda_j T}.$$

Likelihood Function from Multiple, Independent Auctions. Now assume that we have observed a series of independent auctions, with realized outcomes falling into any one of the above four outcomes. For example, let's say we observed four completed auctions, the outcomes of the four auctions are: $(x_i^{(1)}, x_j^{(1)})$; $(x_j^{(2)})$; $(x_i^{(3)}, x_j^{(3)})$; Empty (no bids). Then the likelihood function is

$$L(\lambda_i, \lambda_j) = \lambda_i e^{-\lambda_i x_i^{(1)}}\, \lambda_j e^{-\lambda_j x_j^{(1)}} \cdot e^{-\lambda_i x_j^{(2)}}\, \lambda_j e^{-\lambda_j x_j^{(2)}} \cdot \lambda_i e^{-\lambda_i x_i^{(3)}}\, \lambda_j e^{-\lambda_j x_j^{(3)}} \cdot e^{-\lambda_i T^{(4)}}\, e^{-\lambda_j T^{(4)}}.$$

The (λ_i, λ_j) that maximizes $\ln L$ is

$$\begin{aligned} \lambda_i &= \frac{2}{x_i^{(1)} + x_j^{(2)} + x_i^{(3)} + T^{(4)}}; \\ \lambda_j &= \frac{3}{x_j^{(1)} + x_j^{(2)} + x_j^{(3)} + T^{(4)}}. \end{aligned}$$

3.3.2 More Than Two Bidders

From the above two-bidder case, we see that for each bid price $p = 1, 2, \cdots, K$, we might or might not see p in the recorded bid history. If p appears in the bid history, then it will contribute a multiplicative term, $\lambda_p e^{-\lambda_p x_p}$, to the likelihood function. If p does not appear in the bid history, then we conclude that (1) bidder p arrives later than the next higher recorded bidder i (i.e. $X_p > X_i$), in which case it contributes a multiplicative term $e^{-\lambda_p x_i}$ to the likelihood function; or (2) bidder p arrives later than the auction ending time T (if there are no recorded bidders higher than p), in which case it contributes a multiplicative term $e^{-\lambda_p T}$ to the likelihood function. Summarizing (1) and (2): If we denote X_{p+} as the arrival time of the next higher recorded bidder or the auction ending time T if no higher bidders are recorded, then the contribution of price p to the likelihood function, if price p is not recorded in the bid history, is simply $e^{-\lambda_p x_{p+}}$.

For example, let a recorded bid history be 3 bidders (P_i, P_j, P_k) with auction duration T. Note that i, j, k might not be consecutive. Then the likelihood of having observed this auction is

$$\begin{aligned}
&\Pr\{X_1 > x_i, X_2 > x_i, \cdots, X_i = x_i, X_{i+1} > x_j, X_j = x_j, X_{j+1} > x_j, X_k = x_k, X_{k+1} > T\} \\
&\qquad = e^{-\lambda_1 x_i} e^{-\lambda_2 x_i} \cdots e^{\lambda_{i-1} x_i} \cdot \lambda_i e^{-\lambda_i x_i} \cdot e^{-\lambda_{i+1} x_j} \\
&\qquad\quad \cdot \lambda_j e^{-\lambda_j x_j} \cdot e^{-\lambda_{j+1} x_k} \\
&\qquad\quad \cdot \lambda_k e^{-\lambda_k x_k} \cdot e^{-\lambda_{k+1} T}.
\end{aligned}$$

Finally, we multiply together the likelihood terms from multiple (total S, independent auctions, and solve for the optimal λ_p which maximizes $\ln L$, we obtain the general result

$$\lambda_p = \frac{\text{Number of auctions in which price } p \text{ was recorded}}{x_{p+}^{(1)} + x_{p+}^{(2)} + \cdots + x_{p+}^{(S)}}$$

where $x_{p+}^{(s)}$ denotes the arrival time of bidder p in the s-th auction if bidder p is recorded in the s-th auction, or the arrival time of the next higher bidder, or the auction duration if no bidder higher than p is recorded.

For numerical stability issues, we refer to our discussion in Appendix C.

4 Conclusions, Implementation and Future Research

We have considered two models of the bidding process of on-line auctions. Both models account for the different outcomes in the recorded bids due to randomness in the bidder arrivals; the equally likely bid sequence model makes reference with only the relative timing and not the exact timing of the bids, while the Poisson arrival model takes into consideration of the lengths of the intervals between two bids. From this, we are able to derive estimates of the relative proportions of bidders at each price level.

We have ignored the strategic bidding behavior that is possible in an on-line auction. Prior research has shown that the optimal bidding strategy for a bidder may not be to bid his true valuation; this points to further research of demand-price estimation due to random arrivals with individually optimal bidding strategy.

We have also ignored the fact that there are usually multiple on-line auctions for the same item, and that a bidder can bid more than once in the same auction, and bid in more than one auction, simultaneously or sequentially. We are currently working on the model with a single auction but multiple bids from the same bidder.

The implementation of the models has been carried out in a software prototype. Since the bid amounts in a real auction can be continuous (in cents), the bids are first discretized. The computation time for the estimation is found satisfactory, with almost instantaneous output of the demand curve once the data has been collected. These results are promising for future refinements of the models we presented here.

5 References

Bajari, P. and A. Hortacsu. 2000. Winner's Curse, Reserve Price and Endogenous Entry: Empirical Insights from eBay Auctions. Working paper, Department of Economics, Stanford University.

Lucking-Reiley D., D. Bryan, N. Prasad, and D. Reeves. 1999. Pennies from eBay: the Determinants of Price in Online Auctions. Working paper, Vanderbilt University.

Klemperer, P. 1999. Auction Theory: A Guide to the Literature. *Journal of Economic Surveys*, Vol. 13(3), 227-286.

Perrigne, I. and Q. Vuong. 1999. Structural econometrics of first-price auctions: A survey of methods. *Canadian Journal Of Agricultural Economics*, Vol.47(3): 203-223.

A Proof of Combinatorial Identity

We prove a slightly more general identity: For $k \geq 1$,

$$\sum_{j=0}^{k} \frac{(a+j)!}{j!} = \frac{a!}{0!} + \frac{(a+1)!}{1!} + \frac{(a+2)!}{2!} + \frac{(a+3)!}{3!} + \cdots + \frac{(a+k)!}{k!} = \frac{(a+1+k)!}{(a+1)\cdot k!}.$$

Consider the sum of the first two terms ($j = 0$ and $j = 1$):

$$\frac{a!}{0!} + \frac{(a+1)!}{1!} = a!(1+a+1) = a!(a+2) = \frac{(a+2)!}{a+1}.$$

Now consider the sum of the first three terms ($j = 0, 1, 2$):

$$\frac{(a+2)!}{a+1} + \frac{(a+2)!}{2!} = (a+2)!\left(\frac{1}{a+1} + \frac{1}{2!}\right) = (a+2)!\frac{(a+3)!}{(a+1)\cdot 2!} = \frac{(a+3)!}{(a+1)\cdot 2!}.$$

Now consider the sum of the first four terms ($j = 0, 1, 2, 3$):

$$\frac{(a+3)!}{(a+1)\cdot 2!} + \frac{(a+3)!}{3!} = (a+3)!\left(\frac{1}{(a+1)\cdot 2!} + \frac{1}{3!}\right) = (a+3)!\frac{(3+a+1)}{(a+1)\cdot 3!} = \frac{(a+4)!}{(a+1)\cdot 3!}.$$

Now we can easily apply induction on k. Assume that this identity holds for k. For $k+1$, we have

$$\begin{aligned}
\sum_{j=0}^{k+1} \frac{(a+j)!}{j!} &= \sum_{j=0}^{k} \frac{(a+j)!}{j!} + \frac{(a+k+1)!}{(k+1)!} \\
&= \frac{(a+1+k)!}{(a+1)\cdot k!} + \frac{(a+k+1)!}{(k+1)!} \\
&= (a+1+k)!\left[\frac{1}{(a+1)\cdot k!} + \frac{1}{(k+1)!}\right] \\
&= (a+1+k)!\left[\frac{(k+1)+(a+1)}{(a+1)\cdot(k+1)!}\right] \\
&= (a+1+k)!\frac{a+1+k+1}{(a+1)\cdot(k+1)!} \\
&= \frac{(a+1+k+1)!}{(a+1)\cdot(k+1)!}.
\end{aligned}$$

For our intended summant,

$$\sum_{j=1}^{M+1} \frac{(N-j)!}{(M-j+1)!} = \frac{(N-M-1)!}{0!} + \frac{(N-M)!}{1!} + \cdots + \frac{(N-1)!}{M!},$$

we have $a \equiv N - M - 1$ and $k \equiv M$, hence $a + 1 \equiv N - M$ and $a + 1 + k = N$. The result is

$$\frac{(a+1+k)!}{(a+1) \cdot k!} = \frac{N!}{(N-M) \cdot M!}.$$

B Length of Bid History

By Lemma 1, we can compute the expected length of the bid history (number of bids recorded) of an auction as

$$E[\text{Length of bid history}] = \sum_{i=1}^{K} \Pr\{\text{Price } i \text{ is in bid history}\} = \sum_{i=1}^{K} \frac{m_i}{m_i + \cdots + m_K}.$$

We next derive the complete probability distribution of the length of the bid history, in a recursive procedure. But first, we introduce a preliminary result.

B.1 A Preliminary Result

Given m bidders who must appear in a predefined sequence. Now we want to add n bidders, so that in the new bidding sequence thus formed, we will maintain the predefined sequence of the original m bidders. In how many possible ways can we form the new sequence?

Example: Given 3 bidders A, B and C who must appear in the sequence A before B before C. We now add 2 bidders D and E, so that the new sequence ABCDE, or DEABC, or AEBDC, is a legitimate sequence. How many legitimate sequences are there?

Answer: $(m+1)(m+2)\cdots(m+n)$ (total n terms in the multiplication).

Proof: Consider adding one new bidder in the predefined sequence of m bidders. There are $(m+1)$ possible positions for the new bidder without upsetting the original sequence.

Now consider adding the second new bidder to any one of the $(m+1)$ sequences formed with the first new bidder. There are $(m+1)+1$ positions for the second new bidder.

So on and so forth. QED.

In the above example, $m = 3$, $n = 2$, hence we will have $4 \cdot 5 = 20$ legitimate sequences. There are total $5! = 120$ possible sequences.

B.2 The Main Result on the Length of Bid History

Question: How many bidder sequences will result in a bid history of length h?

Solution: We first solve the simplest case: $h = 1$. The bid history will contain only one bid if and only if the first bidder in a bidder sequence is price K. Since there are m_K bidders with price K, the number of bidder sequences beginning with K will be $m_K \cdot (N-1)!$. The probability of observing a bid history of only 1 bidder is therefore $m_K \cdot (N-1)!/N! = m_K/N$.

Now, consider the general case $h > 1$. Let $f(h; m_1, m_2, \ldots, m_K)$ be the number of bidder sequences that will result in a bid history of length h from bidder groups $m_1, m_2, \ldots, m_K$. We derive $f(\cdot;\cdot)$ recursively.

Consider the first bidder in a bidder sequence. The first bidder can be any integer between (and including) 1 and K.

If the first bidder is 1, then we take the $N - m_1$ bidders whose bid prices are greater than 1 and form a history of length $h-1$; adding the remaining $m_1 - 1$ bidders whose bid price is 1 to any of the subsequence with history of length $h-1$, we will obtain a bid history of length h.

The number of possible bidder sequences if the first bid is 1 is:

$$f(h-1; m_2, \cdots, m_K) \quad \text{if } m_1 = 1;$$

$$m_1 \cdot f(h-1; m_2, \cdots, m_K) \cdot (N - m_1 + 1)(N - m_1 + 2) \cdots (N-1) \quad \text{if } m_1 > 1.$$

Defining $\prod_{i=1}^{0} \equiv 1$ and noting $N = m_1 + m_2 + \cdots + m_K$, we can rewrite the above as

$$f(h-1; m_2, \cdots, m_K) \cdot m_1 \prod_{i=1}^{m_1 - 1} (i + m_2 + \cdots + m_K).$$

If the first bid is 2, then we take the $N - m_1 - m_2$ bidders whose bid prices are greater than 2 and form a history of length $h-1$; adding the remaining $m_1 + m_2 - 1$ bidders whose bid prices are 1 or 2 to any of the subsequence with history of length $h-1$, we will obtain a bid history of length h.

The number of possible bidder sequences if the first bid is 2 is:

$$m_2 \cdot f(h-1; m_3, \cdots, m_K) \cdot (N - m_1 - m_2 + 1) \cdots (N-2)(N-1)$$

$$= f(h-1; m_3, \cdots, m_K) \cdot m_2 \prod_{i=1}^{m_1+m_2-1} (i + m_3 + \cdots + m_K).$$

If the first bid is $K-h+1$, then we take the $N - m_1 - m_2 - \cdots - m_{K-h+1}$ bidders whose bid prices are greater than $K-h+1$ and form a history of length $h-1$; adding the remaining $m_1 + m_2 + \cdots + m_{K-h+1} - 1$ bidders whose bid prices are less than or equal to $K-h+1$ to any of the subsequence with history of length $h-1$, we will obtain a bid history of length h.

The number of possible bidder sequences if the first bid is $K-h+1$ is:

$$m_{K-h+1} \cdot f(h-1; m_{K-h+2}, \cdots, m_K) \cdot (N - m_1 - \cdots - m_{K-h+1} - 1) \cdots (N-2)(N-1)$$
$$= f(h-1; m_{K-h+2}, \cdots, m_K) \cdot m_{K-h+1} \prod_{i=1}^{m_1+\cdots+m_{K-h+1}-1} (i + m_{K-h+2} + \cdot + m_K).$$

If the first bid is greater than $K-h+1$, then it is impossible to form a bid history of length h. Hence, the total number of possible bidder sequences with a bid history of length h is the sum of all the above expressions:

$$f(h; m_1, \cdots, m_K) =$$
$$f(h-1; m_2, \cdots, m_K) \cdot m_1 \prod_{i=1}^{m_1-1} (i + m_2 + \cdots + m_K) +$$
$$+ f(h-1; m_3, \cdots, m_K) \cdot m_2 \prod_{i=1}^{m_1+m_2-1} (i + m_3 + \cdots + m_K) +$$
$$+ \cdots +$$
$$+ f(h-1; m_{K-h+2}, \cdots, m_K) \cdot m_{K-h+1} \prod_{i=1}^{m_1+\cdots+m_{K-h+1}-1} (i + m_{K-h+2} + \cdot + m_K).$$

The boundary condition is $f(1; m_j, \cdots, m_K) = m_K \cdot (m_j + \cdots + m_K - 1)!$ (as we have shown previously).

Example 1: We have $N = 3$ bidders named Alex, Bob, Chuck. Their bid prices are 1,1,2, respectively. Then $K = 2$; $m_1 = 2$ and $m_2 = 1$.

We start with $h = 1$:

$f(1, m_1, m_2) = m_2 \cdot (m_1 + m_2 - 1)! = 1 \cdot (1 + 2 - 1)! = 2$ (Chuck, Alex, Bob or Chuck, Bob, Alex; the bid history contains one bidder – Chuck).

$f(1, m_2) = m_2 \cdot (m_2 - 1)! = 1 \cdots (1 - 1)! = 1.$

Next, consider $h = 2$:

$f(2, m_1, m_2) = f(1, m_2) \cdot m_1 \cdot (1 + m_2) = 1 \cdot 2 \cdot 2 = 4.$ (ABC, ACB, BAC, BCA).

$h = 3$: Impossible because h now exceeds $K = 2$.

Example 2: We have $N = 4$ bidders named Alex, Bob, Chuck and Dirk (A,B,C,D). Their reservation prices are 1,1,2,2, respectively. Then $K = 2$; $m_1 = 2$, $m_2 = 2$.

We start with $h = 1$:

$f(1, m_1, m_2) = m_2 \cdot (m_1 + m_2 - 1)! = 2 \cdots (4 - 1)! = 12$ (they are: CABD, CADB, CBAD, CBDA, CDAB, CDBA, DABC, DACB, DBAC, DBCA, DCAB, DCBA).

$f(1, m_2) = m_2 \cdot (m_2 - 1)! = 2 \cdot (2 - 1)! = 2$ (they are: CD, DC).

Next, consider $h = 2$:

$f(2, m_1, m_2) = f(1, m_2) \cdot m_1 \cdot (1 + m_2) = 2 \cdot 2 \cdot (1 + 2) = 12$ (they are: ABCD, ACBD, ACDB, ABDC, ADBC, ADCB, BACD, BADC, BCAD, BCDA, BDAC, BDCA).

Since we have assumed that each of the $N!$ arrival sequences is equally likely, the probability of observing a bid history of length h is simply $f(h; \cdot)/N!$.

C Numerical Stability Issues in the Poisson Arrival Model

We have derived a relationship between the rate at which bids of a certain value λ_i arrive and p_i, the probability of seeing a bid with a value of i in an auction of duration T. We are interested in the λ_is because they can be translated into the demand curve.

The formula is

$$p_i = \frac{\lambda_i}{\lambda_i + \gamma_i} \left(1 - e^{-(\lambda_i + \gamma_i)T}\right),$$

where

$$\gamma_i = \sum_{j=i+1}^{N} \lambda_j.$$

If we assume that the probabilities, p_i, are measured, we can compute the rates, λ_i. The purpose of this exercise is to understand how uncertainties in the probabilities affect the accuracy with which we can find the rates.

The normal problem is that round-off errors grow exponentially or as some large power of the number of terms in the recursion. That's not the issue here. Here, p_i is a sampled value, and getting even a few percent accuracy is a problem. Hence, even modest growth factors can give values for the λ_i with no precision at all.

There are two problems. First, we need to make sure that the error in any one of the p_is isn't magnified too much. Secondly, we need to make sure that each step in the recursion doesn't multiply the error from previous steps by too much.

C.1 Start of Recursion

We can find all the λ_is using a recursive procedure starting with $i = N$. Let's look at the error in computing λ_N. Clearly, $\gamma_N = 0$, so we have

$$p_N = 1 - e^{-\lambda_N T}.$$

We want to know how much an error in p_N affects the value of λ_N, so we write

$$p_N + \delta_p = 1 - e^{-(\lambda_N + \delta_\lambda)T}.$$

We can rearrange terms to get

$$\lambda_N + \delta_\lambda = -\frac{1}{T} \ln(1 - p_N - \delta_p).$$

Expanding the logarithm, and noting that $\lambda_N T = -\ln(1 - p_N)$ gives

$$\delta_\lambda = \frac{1}{T} \frac{\delta_p}{1 - p_N} = \frac{\delta_p}{T} e^{\lambda_N T}.$$

The problem is that p_N may be arbitrarily close to 1, as will be the case when T is large or when there are many bidders at bid N. (**Question:** How does δ_p change as $p_N \to 1$?) In other words, the error in λ_N can be quite a bit larger than the error in p_N. Fortunately, we can save the day by using the probability of *not* finding bid N, $\tilde{p}_N$. The result is

$$\delta_\lambda = -\frac{1}{T} \frac{\delta_{\tilde{p}}}{\tilde{p}},$$

which relates the error in the rate to the relative error in the probability, a quantity controlled by how the statistics are gathered.

C.2 Infinite Auctions

Of course, there's no such thing as an infinitely long auction.[1] However, any auction in which all parties are given an opportunity to bid can be considered infinite. In this limit, Dirk's equation becomes

$$p_i = \frac{\lambda_i}{\lambda_i + \gamma_i}.$$

We need a recursive solution starting with p_N, but there's a problem; that equation carries no information. Since the highest bid is guaranteed to appear in an infinite auction, the equation becomes simply $p_N = 1$, which has no dependence on λ_N. In other words, λ_N is undetermined.

We have another problem for infinite auctions, the scale. For a finite auction, the duration T sets the scale. We can solve both our problems by using λ_N to set the scale. In what follows, I'll take $\lambda_N = 1$, but I'll still write λ_N so the formulas look more like the ones that follow.

Let's look at $N-1$, namely

$$p_{N-1} = \frac{\lambda_{N-1}}{\lambda_{N-1} + \lambda_N},$$

and look at the errors,

$$p_{N-1} + \delta_p = \frac{\lambda_{N-1} + \delta_\lambda}{\lambda_{N-1} + \delta_\lambda + \lambda_N}.$$

Expanding and dropping terms of order δ^2 gives

$$\frac{\delta_\lambda}{\lambda_{N-1} + \lambda_N} = \frac{\delta_p}{1 - p_{N-1}} = \delta_p \left(1 + \frac{\lambda_{N-1}}{\lambda_N} \right).$$

which has a problem since $\lambda_{N-1} > \lambda_N$. Fortunately, we can use the same cure as before, namely,

$$\frac{\delta_\lambda}{\lambda_{N-1} + \lambda_N} = -\frac{\delta_{\tilde{p}}}{\tilde{p}}.$$

[1] I'll pay \$10,000 for the best joke. No time limit.

or

$$\frac{\delta_\lambda}{\lambda_{N-1}} = \frac{\lambda_{N-1} + \lambda_N}{\lambda_{N-1}} \frac{\delta_{\tilde{p}}}{\tilde{p}}.$$

This result is even nicer than the last one because it gives the relative error in λ_{N-1}.

What about other values of i? The expansion becomes

$$p_i + \delta_p = \frac{\lambda_i + \delta_\lambda}{\lambda_i + \delta_\lambda + \gamma_i + \delta_\gamma}.$$

As before, I'll expand and drop terms of order δ^2 to get

$$\delta_\lambda = \frac{\lambda_i + \gamma_i}{1 - p_i}\delta_p + \frac{p_i}{1 - p_i}\delta_\gamma,$$

which reduces to

$$\frac{\delta_\lambda}{\lambda_i} = \frac{\lambda_i + \gamma_i}{\lambda_i} \frac{\delta_{\tilde{p}}}{\tilde{p}} + \frac{\delta_\gamma}{\gamma_i}$$

which nicely bounds the error. In other words, there is no accumulation in the relative error of γ_i.

C.3 Full Equation

Now we can look at the behavior of the full equation

$$p = \frac{\lambda}{\lambda + \gamma}(1 - e^{-(\lambda+\gamma)T}),$$

where I've dropped the subscript for clarity. We can write this as

$$(\lambda + \gamma)T = -\ln(1 - p\frac{\lambda + \gamma}{\lambda}).$$

Rearranging a bit gives

$$(\lambda + \gamma)T = \ln \lambda - \ln(\lambda - p\lambda - p\gamma).$$

Next, we add the error in each of λ, γ, and p to get

$$(\lambda + \gamma)T + (\delta_\lambda + \delta_\gamma)T = \ln(\lambda + \delta_\lambda) - \ln(\lambda - p\lambda - p\gamma + \delta_\lambda - p\delta_\lambda - \lambda\delta_p - p\delta_\gamma - \gamma\delta_p).$$

Factoring, expanding the logarithms, and eliminating terms gives

$$(\delta_\lambda + \delta_\gamma)T = \frac{\delta_\lambda}{\lambda} - \frac{\delta_\lambda - p\delta_\lambda - \lambda\delta_p - p\delta_\gamma - \gamma\delta_p}{\lambda - p\lambda - p\gamma}.$$

We can now solve for the unknown error, δ_λ, in terms of the know errors, δ_p and δ_γ.

$$\frac{\delta_\lambda}{\lambda}\left(\lambda T - 1 + \frac{1-p}{1-p-p\gamma/\lambda}\right) = \delta_p \frac{1+\gamma/\lambda}{1-p-p\gamma/\lambda} + \frac{\delta_\gamma}{\lambda}\left(\frac{p}{1-p-p\gamma/\lambda} - \lambda T\right).$$

We can now subsitute for some of the terms. For example,

$$1 - p - p\gamma/\lambda = e^{-(\lambda+\gamma)T}.$$

We now have

$$\frac{\delta_\lambda}{\lambda}\left[\left(\lambda T - 1 - \frac{\lambda}{\lambda+\gamma}\right)e^{-(\lambda+\gamma)T} + \frac{\gamma}{\lambda+\gamma}\right] =$$
$$\frac{\lambda+\gamma}{\lambda}\delta_p + \frac{\delta_\gamma}{\lambda}\left[\frac{\lambda}{\lambda+\gamma} + \left(\frac{\lambda}{\lambda+\gamma} - \lambda T\right)e^{-(\lambda+\gamma)T}\right],$$

which is quite a mess.

Could it possibly be right? Let's let $T \to \infty$ and see if we get the infinite auction result. We see that

$$\frac{\gamma}{\lambda+\gamma}\frac{\delta_\lambda}{\lambda} = \frac{\lambda+\gamma}{\lambda}\delta_p + \frac{\lambda}{\lambda+\gamma}\frac{\delta_\gamma}{\lambda}$$

becomes

$$\frac{\delta_\lambda}{\lambda} = \frac{\lambda+\gamma}{\lambda}\frac{\lambda+\gamma}{\gamma}\delta_p + \frac{\delta_\gamma}{\gamma}.$$

In this limit, $\gamma/(\lambda+\gamma) = 1-p$, so we end up with

$$\frac{\delta_\lambda}{\lambda} = \frac{\lambda+\gamma}{\lambda}\frac{\delta_p}{1-p} + \frac{\delta_\gamma}{\gamma},$$

which is the infinite auction result.

C.4 The Gotcha

All of this looks good, but there's a *progressive* problem.[2] Let's look at the infinite auction result.

$$\frac{\delta_{\lambda_i}}{\lambda_i} = \frac{\lambda_i+\gamma_i}{\lambda_i}\frac{\delta_{\tilde{p}_i}}{\tilde{p}_i} + \frac{\delta_{\gamma_i}}{\gamma_i},$$

where I've added subscripts on the error terms. The problem is the last term because

$$\delta_{\gamma_i} = \sum_{j=i+1}^{N} \delta_{\lambda_j},$$

[2]How many gifts did my true love give me on the twelfth day of Christmas?

and δ_{λ_j} has a term δ_{γ_j}.

In order to save some space, I'll define

$$\alpha_i = (\lambda_i + \gamma_i)\frac{\delta_{\tilde{p}_i}}{\tilde{p}_i},$$

and use

$$\delta_{\lambda_i} = \alpha_i + \frac{\lambda_i}{\gamma_i}\delta_{\gamma_i}.$$

Substituting for δ_{γ_i} gives

$$\delta_{\lambda_i} = \alpha_i + \frac{\lambda_i}{\gamma_i}\sum_{j=i+1}^{N}\delta_{\lambda_j}$$

which is

$$\delta_{\lambda_i} = \alpha_i + \frac{\lambda_i}{\gamma_i}\sum_{j=i+1}^{N}\left(\alpha_j + \frac{\lambda_j}{\gamma_j}\delta_{\gamma_j}\right).$$

Continuing we get

$$\delta_{\lambda_i} = \alpha_i + \frac{\lambda_i}{\gamma_i}\sum_{j=i+1}^{N}\left(\alpha_j + \frac{\lambda_j}{\gamma_j}\sum_{k=j+1}^{N}\delta_{\lambda_j}\right),$$

and

$$\delta_{\lambda_i} = \alpha_i + \frac{\lambda_i}{\gamma_i}\sum_{j=i+1}^{N}\left[\alpha_j + \frac{\lambda_j}{\gamma_j}\sum_{k=j+1}^{N}(\alpha_k + \delta_{\gamma_k})\right].$$

We keep going until we reach the point where the only term left is δ_{γ_N}, which is zero. At that point, we'll have α_{i+1} appearing once, α_{i+2} twice, *etc.* In other words, the error in δ_{p_N} appears $N-2$ times in the error for δ_{λ_1}; the error for $\delta_{p_{N-1}}$ appears $N-3$ times, and so on.

The factors λ_i/γ_i might save the day, since each subsequent term is multiplied by the product of the preceding λ_i/γ_is. Unfortunately, there is no way to guarantee that $\lambda_i/\gamma_i < 1$. Hence, some of these factors, *e.g.*, those for steep parts of the demand curve, may magnify the errors. We can monitor this factor during the recurrence procedure, but there will still be demand curves where we have a problem.

We conclude that the recursive procedure will not be useful for more than a few points on the price curve. The situation can be improved by binning. Instead of taking each price as a separate point, break the interval into 5 or so bins and aggregate the statistics. This approach has two benefits. The statistics are improved so that the relative error in p_i is reduced, and the recurrence is shorter, leading to less error build up. Besides, 5-8 points should be more than enough precision on the price curve.