

特別研究報告書

広告オークションにおける入札者の
評価値分布の推定

指導教員 松原 繁夫 准教授

京都大学工学部情報学科

伊藤 優也

平成26年1月31日

広告オークションにおける入札者の評価値分布の推定

伊藤 優也

内容梗概

近年、インターネットの普及に伴いディスプレイ広告の市場は拡大の一途を辿っている。ディスプレイ広告とはオンライン広告の一形式であり、バナー広告やパネル広告などのサイト上に埋め込まれて画像や動画が表示される広告のことである。

ディスプレイ広告の多くはオークションによって取引される。このオークションの特徴は、買い手である広告主が直接入札をして財(広告を掲載する広告枠)を手に入れるのではなく、DSP(Demand Side Platform)とよばれるプラットフォームが広告主と売り手である媒体社(メディア)の仲介に立ち、代理入札を行うという点である。DSPは複数の広告主を顧客として管理し、広告主の予算などの制約を満たすうえで最も高い入札ができる広告を選択し入札を行う。DSPの立場からすれば、自身の利益が高くなるような入札戦略の獲得が重要な課題となっている。

ディスプレイ広告オークションでは第二価格入札が採用されている。また、財となる広告枠は一日に何度もオークションによって販売されているので、逐次オークションとみなすことができる。一度きりの第二価格入札においては、財に対する評価値をそのまま提示することが最適戦略である。しかしこれが逐次オークションとなると評価値をそのまま提示し続けることは均衡戦略とはならず、一般的な最適戦略は存在しない。よって試行錯誤で良い入札戦略を考案する必要がある。他入札者の財に対する評価値分布を知ることが出来れば、より良い入札戦略を獲得することができるだろう。そこで本研究では、DSP業者から入手した実オークションデータを分析して、他の入札者の評価値分布を推定することを目的とする。

オークションにおける評価値分布の推定は一部の先行研究でも議論されているが、先行研究とは取り扱う状況が異なっており、ディスプレイ広告を対象とする場合は以下の二つの課題が存在する。

1. 部分的なデータからの評価値分布の推定

ディスプレイ広告オークションで観測可能なデータは各広告枠に対するオークションの開催時間(入札要求のあった時間)、自己の入札額、オークショ

ンの勝敗の結果であり，オークションに勝利した場合はさらに落札額(支払額)が観測できるが，敗北した場合は敗北したという結果のみしか観測できない．このようにYahoo!オークションなどの公開オークションと異なり，他の入札者の入札額や落札額の全てを観測することができない状況下で入札者の評価値分布を推定する必要がある．

2. 経時変化を考慮した評価値分布の推定

1日の経時変化を考慮する必要がある．各広告には予算上の制限あり，時間が経過するにつれて予算が消化されていくので，いくら広告枠に対して高い評価値を持っていても実際に入札することができなくなる．よって単純に評価値分布と言っても，例えば朝の8時と夜の8時では異なる評価値分布を持っているはずである．

本研究では，前者の課題を解決するために，オークション理論を援用して，評価値分布と落札確率，落札できた場合の落札額分布の関係を明らかにする．また，後者の問題を解決するために，複数のモデルを作成し，実オークションデータを用いて最尤推定を行う．そして，どのモデルが最も優れているかを調べるため，モデル選択を行う．具体的にはオークションが開催される時間と入札者の残り予算それぞれの影響の有無を考慮したモデル4つを作成し，最尤推定によって評価値分布を決定した．そしてそれらのモデルの予測の精度を比較するためにAICという指標を用いてモデル選択を行った．

本研究の貢献は以下の2点である．

1. 部分的なデータのみを用いた評価値分布モデルの獲得

オークション理論を援用することで，他の入札者が財に対して同一の評価値分布を持つという仮定のもとでの評価値分布の推定を可能にした．それによって自己の入札額に対する勝利確率や，敗北した場合の1位の入札額の期待値を求める計算式を導出できた．

2. 経時変化を考慮した評価値分布モデルの獲得

実オークションデータを用いて最尤推定を行うことによって，予算制約による評価値分布を考慮したモデルよりも，予算を直接考慮せずに時間帯毎にモデルを作成する方が，より実データに合致する結果が得られることを明らかにした(AIC基準で約20%の改善)．これは各DSPが複数の広告を管理し，それぞれが予算を持っているため時間と共に予算が単調に減少していくとは一概に言えないことが理由として考えられる．

Estimating the distribution of bidder's evaluation values in advertisement auctions

Yuya ITOH

Abstract

Recently, markets of display advertisements are growing with the increase of the Internet user. Display advertisements are a kind of online advertisements such as banner advertisements, which are shown in the sides of web pages.

Many display advertisements are sold by auctions(ad auction). The feature of this auction is that buyers, advertisers do not directly take bids to purchase ad spaces where their advertisements show. In ad auctions, platforms called DSP(Demand Side Platform) play roles like mediators and take bids instead of buyers. Each DSP has advertisers as clients and selects the advertisement which can take the highest bids within constraints of budget of advertisers. Considering from the position of one DSP, it is important problem to develop its bidding strategies to maximize its revenue.

Ad auction applies second-price auction, and it can also be regarded as sequential auction because ad spaces are sold many times in a day. In one-time second-price auction, it is a dominant strategy for every bidder to bid his true evaluation value. In sequential second-price auction, however, continuously bidding true evaluation values is not a dominant strategy and there is no general best strategy in this auction. Therefore, it is necessary to develop better strategies through a trial and error process. If the distribution of bidders' evaluation values can be observed, better strategies will be developed more easily. So, this study estimates the distribution of bidders' evaluation values in ad auctions by analyzing actual auction data from a certain company which has a DSP.

There are studies about estimating the distribution of evaluation values in auctions, but the situation in this study is different from previous studies as follows.

1. Estimating distribution from partial data

Observable data in ad auction is time each auction starts, your own bids, and results of auctions(win or lose). In addition, winning bids can be observed when you win auctions. In this way, it is necessary to estimate the

distribution of bidders' evaluation values in the situation where opponent bidders' bid cannot be observed.

2. Estimating the distribution considering the change of time

Ad auction is held through a day and it is necessary to consider the change of time in a day. Each advertisement has constraints of budget. Because budgets are decreasing as time passes, taking high bid in the end of the day is less probable however high distribution a DSP has.

To settle the former problem, I use auction theory to find the relation among the distribution of the evaluation values, the probability of winning auctions, and the distribution of the winning bids when win. To settle the latter problem, I construct four models and estimate them by maximum likelihood estimation method. Specifically, I construct four models, considering the influence of time that each auction starts, and the influence of budget that each advertisement has, on the distribution of bidder's evaluation values. After estimating those models, I use AIC value to select the best model.

The contributions of this study are following.

1. Acquisition of the distribution of evaluation values from partial data

Based on the auction theory, it becomes possible to estimate the distribution of bidder's evaluation values assuming that every bidder's distribution are i.i.d.. By using this distribution, I calculated the probability of winning when you take a certain bid, and the expectation of the highest bid when you lose.

2. Acquisition of the distribution model of evaluation values considering changes of time

By using maximum likelihood estimation method with actual auction data, I found that the model where budget is not considered, and time each auction starts is considered is the best model of the four (AIC value is better than the worst by 20%). This is because budgets are allocated by each advertisement, so budgets are not monotonically decreasing as time passes.

広告オークションにおける入札者の評価値分布の推定

目次

第1章	はじめに	1
第2章	オークションの基礎	4
2.1	第二価格入札	4
2.2	逐次オークション	5
第3章	最尤法	6
3.1	最尤法とは	6
3.2	利用可能なデータ	7
3.3	オークション理論に基づく対数尤度関数の設定	7
第4章	モデル作成	9
4.1	敵入札者の入札モデル	9
4.2	最尤法によるモデル推定	11
4.3	対数尤度の比較	14
第5章	モデル選択	15
5.1	AICとは	15
5.2	AICによるモデル選択	18
5.3	時間帯の分割数	20
第6章	入札額および落札額の推定	20
6.1	オークション勝利時の落札額	21
6.2	オークション敗北時の1位の入札額および落札額	22
第7章	おわりに	24
	謝辞	25
	参考文献	25
	付録：推定データ	A-1
A.1	モデルの推定データ	A-1
A.2	推定された評価値分布	A-4

第1章 はじめに

インターネットの普及に伴い、ディスプレイ広告などのオンライン広告の市場が拡大している。ディスプレイ広告とはオンライン広告の一形式であり、バナー広告やパネル広告などのサイト上に埋め込まれて画像や動画が表示される広告のことである。

以前までは、広告主は期間保証型と呼ばれる形式で媒体社(広告を掲載するメディア)に対価を支払っていた。期間保証型では、広告主は一定の金額を事前に媒体社に支払い、広告を一定期間掲載することができる。また、期間保証型の他にも、一定期間広告を表示し、その間にユーザが広告をクリックした回数に応じて課金額が決定するクリック課金型(Cost per Click:CPC)なども登場した。しかしこれらの方式では、広告主はターゲットとなりえないユーザに対しても広告を表示してしまうという問題点があった。例えば、酒やたばこなどの20歳以上対象の商品の広告を20歳未満のユーザに表示しても、広告主の利益に繋がるとは考えにくく、無駄な広告費が発生してしまう。

そこで、この状況を打破するプレイヤーとして、DSP(Demand Side Platform)と呼ばれるプラットフォームが2008年頃に誕生した。DSPは期間単位で広告枠を買い付けるのではなく、インプレッション¹⁾ごとに買い付け判断を行うことができる。そのため、販売対象外のユーザを取り除き、広告主にとって意味のあるインプレッションのみを買い付けることが可能になった。DSPは複数の広告主をクライアントとして保持し、各広告主の利益が最大になるように支援する。DSPの誕生によって、広告主は媒体社からリアルタイムで広告枠を買い付ける仕組みへと変わっていった。

また、DSPの誕生と時を同じくして、媒体社側の利益向上を支援するSSP(Supply Side Platform)と呼ばれるプラットフォームが誕生した。SSPは財となる広告枠を、Ad Exchangeと呼ばれる広告の取引市場に1インプレッション単位で提供する。Ad Exchange内ではディスプレイ広告オークションが行われ、最も高い入札を行ったDSPの広告をWEBページに表示する。

ここでディスプレイ広告オークションの流れを図1に示す。ディスプレイ広告オークションにおいて、DSPは複数の広告主とオークシヨニアの間に立って、代理入札を行う仲介者の役割を果たす。1日にDSPが利用可能な予算は広告毎

¹⁾ 広告をWEBページに掲載すること

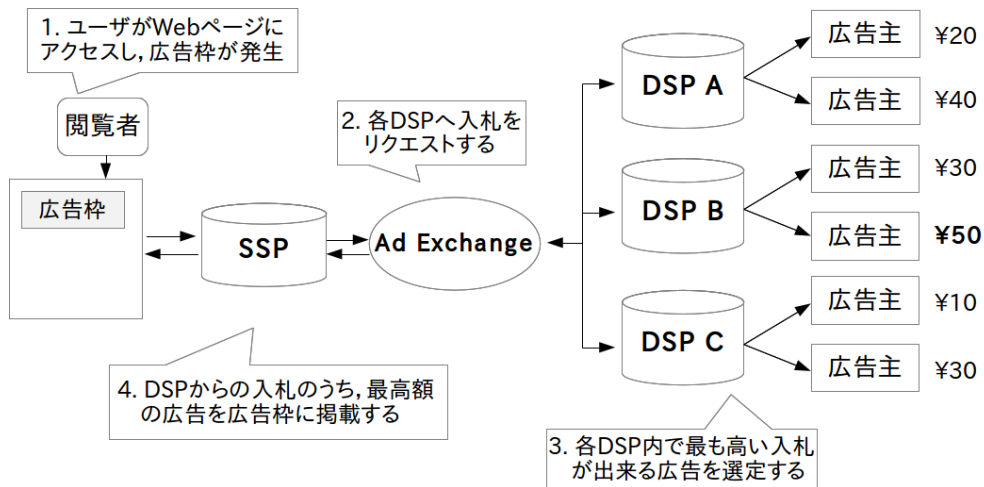


図 1: デisplay広告オークションの流れ

に割り当てられている。例えば、1日のオークションが0時から開始されているとすると、同日24時で1日のオークションが終了し、広告の予算が再び割り当てられる仕組みとなっている。

実際にユーザがWEBページを訪れてから広告が表示されるまでには以下の過程が踏まれる。

1. ユーザがWEBページを訪れると、WEBページに広告を掲載する枠が発生する。
2. Ad Exchangeが広告枠毎にオークション開催を各DSPに通知する。
3. 各DSPはSSPからの情報と独自の情報を元に、条件を満たした上で最も高い入札が可能な広告を選出し、入札を行う。
4. SSPは各DSPからの入札のうち、最も高い入札額を行ったDSPを判断し、そのDSPの選出した広告をWEBページの広告枠に掲載する。

ディスプレイ広告オークションでは第二価格入札が採用されている。また1日に同じ財(広告枠)が何度も販売されることから、逐次オークションとみなすことができる。一度きりの第二価格入札においては、財に対する評価値を素直に入札額として提示することが最適戦略となるのがオークション理論によって証明されている[1][2]が、第二価格入札の逐次オークションとなると、評価値を素直に入札し続けることが最適戦略とはならない。そのためDSPの利益を上げるためには、試行錯誤的に良い入札戦略を探索する必要がある。入札戦略を考案する際に、他DSPの財に対する評価値分布を知ることが出来ればより良い

戦略を獲得できるであろう。そこで本研究では、DSP 業者から入手した実オークションデータを分析し、他 DSP の評価値分布を推定することを目的とする。

評価値分布の推定に関する先行研究としては、渡邊 [3] が挙げられる。渡邊は Yahoo!オークションに出品された 1 円銀貨のオークション 68 件を分析することによって入札者の評価値分布モデルを推定した。そして推定されたモデルを用いてオークションの潜在入札者数の推定を行なっている。また、オークションの構造推定に関しての理論的な基礎や証明については Milgrom[4], Paarsch, Hong and Haley[5] などが挙げられる。

しかし、本研究では先行研究と対象としているオークションが異なるため、先行研究をそのまま当てはめることはできない。ディスプレイ広告オークションを対象とする場合には、以下の 2 つの課題が存在する。

部分的なデータからの評価値分布の推定

先行研究では Yahoo!オークションを対象としている。Yahoo!オークションは公開オークションと呼ばれる種類であり、他入札者の入札額およびオークションの落札額の全てが観測可能である。しかし、本研究の対象としているディスプレイ広告オークションでは他入札者の入札額を観測できない。ディスプレイ広告オークションで観測可能なデータは各オークションの開催時間、自己の入札額、オークションの勝敗結果であり、オークションに勝利した場合には落札額が観測できるが、敗北した場合は敗北という結果のみ観測可能である。このため、一部のデータしか観測することができない状況下で入札者の評価値分布を推定する必要がある。

経時変化を考慮した評価値分布の推定

1 日の経時変化を考慮する必要がある。各広告には予算上の制限があり、時間が経過するにつれて予算が消化されていくので、いくら広告枠に対して高い評価値を持っていても、予算が尽きれば実際に入札することができなくなる。よって単純に評価値分布と言っても、例えば朝の 8 時と夜の 8 時では異なる評価値分布を持っているはずである。

本研究では、前者の課題を解決するためにオークション理論を援用し、評価値分布と落札確率、落札できた場合の落札額分布の関係を明らかにする。評価値分布を明らかにすることによって、オークションにおける最適な留保価格¹⁾や潜

¹⁾ オークションの最低販売価格のこと

在的な入札者数の推測が可能になり、より多くの情報を用いて戦略を獲得することが可能になる。

また、後者の問題を解決するために複数のモデルを作成し、実オークションデータを用いて最尤推定を行う。そして、どのモデルが最も優れているかモデル選択を行う。具体的にはオークションが開催される時間と入札者の残り予算それぞれの影響の有無を考慮したモデル4つを作成し、最尤推定によって評価値分布を決定した。そしてそれらのモデルの精度を比較するために AIC という指標を用いてモデル選択を行った。モデル選択の結果として、入札者の予算が財に対する評価値分布に影響を与えないことを示した。また、1日を複数の時間帯に分割し、それぞれの時間帯で異なる評価値をもつと考えたモデルが最も優れたモデルであることを示した。オークション理論と最尤推定法という手法を組み合わせることで問題解決に当たるところが本研究の特徴である。

本稿の構成は以下の通りである。まず第2章でディスプレイ広告オークションに関するオークションの基礎についての説明を行い、第3章でモデル作成の際に利用する最尤法について述べ、対数尤度関数を設定する。第4章では実際に入札者が従う評価値分布のモデルを複数作成し、第5章でどのモデルが最も優れているかを議論する。第6章では選択されたモデルを用いて、自己の入札額での勝利確率や敗北した場合の1位の入札額の期待値を求める方法を述べる。最後に第7章をまとめとして本稿を締めくくる。

第2章 オークションの基礎

オークションには様々な種類があるが、本研究の対象であるディスプレイ広告オークションでは第二価格入札とよばれるオークションと、逐次オークションとよばれるオークションが複合した形となっている。本章では、本研究の基礎となる第二価格秘密入札と逐次オークションについて説明をしていき、これらのオークションの基礎知識を踏まえた上で本研究における評価値分布の推定について議論していく。

2.1 第二価格入札

第二価格入札 (Second-price auction) とは封印入札方式 (sealed bid auction) とよばれるオークションの一形式である。封印入札方式とは、入札者が相互に

入札額を知ることができない競売である。互いの入札額を知ることができないので入札は一度きりしか行われず、最も高い入札額を提示した者がオークションの勝者となる。封印入札方式の反対に位置するオークションとして公開入札方式 (open bid auction) が存在する。公開入札では入札者が相互に入札額を知ることができ、例えば英国式オークションでは相手を上回る入札額を何度も提示していくことが可能である。公開入札の例としては Yahoo!オークションが挙げられる。

封印入札方式において有名なオークションは第一価格入札 (First-price auction) と第二価格入札である。第一価格入札とは、入札者のうち最も高い金額で入札したものが自分の提示した入札額で財を落札するオークションである。一方、本研究の対象にしている第二価格入札とは、入札者のうち最も高い金額を入札したものが財を落札するという点では変わらないが、その時に支払う落札額は自分の入札額ではなく、二番目に高い入札額となる点で異なる。第二価格入札はアメリカの経済学者 Vickrey によって考案されたため、Vickrey オークションともよばれる [6]。

単一財の第二価格入札が一度行われる場合は、販売される財に対する自己の評価値を素直に入札することが支配戦略であることが知られている [1][2]。支配戦略とはゲーム理論における戦略の性質の一つで、他の合理的なプレイヤーがどの戦略を選択しても、自分にとって最も利得 (利益) が大きくなる戦略のことを指す。つまり、一度きりの第二価格入札では、入札者が入札額を決定するのに他者の戦略や評価値を考える必要はなく、素直に自分の評価値を提示することが最も良い戦略となる。

2.2 逐次オークション

逐次オークション (sequential auction) とは、同じ財が逐次的に販売されるオークションである [7]。ディスプレイ広告オークションでは複数の広告が一日に何度も販売されているので、複数財の逐次オークションとみなすことができる。

ディスプレイ広告オークションでは第二価格入札が行われているが、複数の財 (広告枠) が一日に何度も販売にかけられているので、複数財の第二価格逐次入札であるとみなすことができる。この場合、財に対する評価値を素直に入札し続けることは均衡戦略とはならない。評価値をそのまま入札すると、高い評価値を持つ入札者がオークションに勝利し、オークションから順に抜けていく

ことになる。よって、後のオークションになるほど落札額が低くなる。そのため、高い評価値を持っていても敢えて過少申告することで、より低い落札額で財を獲得することができる。どの程度過少申告するかは、他の入札者の評価値に依存するため、一般的な最適戦略は存在しない。

第3章 最尤法

本章では、評価値分布の推定の際に利用する最尤法について説明する。

3.1 最尤法とは

最尤法 (maximum likelihood estimation) とは統計学において、与えられたデータからそれが従う確率分布の母数について推測するためによく用いられる方法である。尤度という指標を用いて、データの分布形状を最もよく表現する分布関数を決定するときに利用する。

例として、正規分布 $N(\mu, \sigma)$ に従う母集団から無作為抽出された標本の実現値を $x_1, x_2, x_3, \dots, x_n$ とすると、これらが出現するためのそれぞれの確率密度は

$$f_N(x_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \quad (i = 1, 2, \dots, n)$$

となる。ここで、 μ と σ をそれぞれ θ_1, θ_2 で表し、さらに $\boldsymbol{\theta} = (\theta_1, \theta_2)$ の関数と考えば $f_N(x_i, \boldsymbol{\theta})$ と表現できる。 $x_1, x_2, x_3, \dots, x_n$ は独立な確率変数 $X_1, X_2, X_3, \dots, X_n$ の実現値より $f_N(x_i, \boldsymbol{\theta})$ ($i = 1, 2, \dots, n$) の積を $\boldsymbol{\theta}$ の関数とみて $L(\boldsymbol{\theta})$ とおく。これを尤度関数という。すなわち

$$L(\boldsymbol{\theta}) = f_N(x_1, \boldsymbol{\theta}) \cdot f_N(x_2, \boldsymbol{\theta}) \cdots f_N(x_n, \boldsymbol{\theta})$$

となる。

最尤法とは、この尤度関数 $L(\boldsymbol{\theta})$ を最大にするような $\boldsymbol{\theta}$ の推定量 $\tilde{\boldsymbol{\theta}}$ を求める方法であり、 $\tilde{\boldsymbol{\theta}}$ を“最尤推定量 (maximum likelihood estimator)” とよぶ。また、実際の計算では自然対数をとった値である対数尤度関数を用いる。

$$\begin{aligned} \ln L(\boldsymbol{\theta}) &= \ln(f_N(x_1, \boldsymbol{\theta}) \cdot f_N(x_2, \boldsymbol{\theta}) \cdots f_N(x_n, \boldsymbol{\theta})) \\ &= \sum_{i=1}^n \ln f_N(x_i, \boldsymbol{\theta}) \end{aligned}$$

先程と同様に、この対数尤度関数 $\ln L(\theta)$ の最尤推定量を求めることによって、母集団の従う確率分布を推定することができる。

3.2 利用可能なデータ

ここでは本研究で利用したデータの種類について説明する。本研究では、広告業者である株式会社マイクロアドが所有する DSP(以下自己という) から入手した 2012 年 6 月 29 日のオークションデータを利用する。分析の対象として、一日のオークションで最も販売された 50 個の広告枠を求め、それぞれの広告枠に対するオークションデータから無作為抽出した 3000 件を利用した。データの中には、自己が選出した広告の ID、自己の入札額、オークションの落札額、オークションの開催時間が含まれる。第 1 章で述べたように、ディスプレイ広告オークションでは自己が敗北した時のオークションの落札額は観測できないため、かわりに敗北を示す記号 “N” が観測されると考える。本研究では、これらのデータから自己の入札額、オークションの落札額、オークションの開催時間の 3 つを分析して評価値分布を推定していく。

3.3 オークション理論に基づく対数尤度関数の設定

ここで、本研究における対数尤度関数について考える。 n 人の入札者が同一の評価値分布に従って財の評価を行うとしたとき、その時の確率密度関数を $f(x)$ 、分布関数を $F(x)$ とする。DSP は複数の広告主を顧客として持ち、各広告主は各広告枠に対して評価値を持っている。よって、それらを集約することで DSP としての評価値分布が形成される。これを DSP が持つ評価値分布とする。ただし、本稿では議論を簡単にするため、各 DSP が同じ評価値分布を持つと仮定する。また、オークションの状況を簡単にするために、さらに以下の 2 つの仮定をおく。

1. オークションの入札者数は $n = 5$ 人
2. 入札者は常に評価値通りの入札を行う

2012 年 6 月時点における国内の主要な DSP の数はおよそ 10 程度あるが、全ての DSP が常に入札を行なっているとは考えにくい。そこで本稿では入札者数を 5 人とし、各オークションでの入札者数の変化はないものとする。また、現実のオークションでは、各入札者は独自の入札戦略を持って入札額を決定するが、本稿では全ての入札者は常に評価値を素直に入札する戦略を用いるとする。こ

の戦略は、一度きりの第二価格入札の支配戦略であり、またディスプレイ広告オークションでは複雑な戦略をとるのが難しい、というマイクロアド社からのインタビューからも一次近似として妥当であると考えることができる。以下ではこれらの仮定を用いた上で対数尤度関数を設定する。

各広告枠に対して利用可能なデータは、各オークションにおける自己の入札額と落札額、およびオークションの開催時間を含んだオークションデータ $T = 3000$ 件分である。しかし、上述のように自己が敗北した場合は落札額を観測することができない。そこで自己が勝利したときは落札額 w を、敗北したときは N という敗北記号を観測すると考えて対数尤度を考える。

利用可能なデータ $T = 3000$ 件のうち、自己が勝利したデータ、敗北したデータがそれぞれ T_W, T_L 件あるとする ($T_W + T_L = T$)。また、それぞれのデータに対するインデックスを下付きの t で表すとする。このときオークション理論に基づくと、落札額 w または敗北記号 N を観測する確率は以下のように求めることができる。

1. 自己が勝利した場合

自己が b_t を入札して勝利し、かつ落札額 w_t を観測する確率密度は

$$(n-1)f(w_t)F(w_t)^{n-2} \quad (0 \leq w_t \leq b_t)$$

である。これは自己を除く $(n-1)$ 人の入札者のうち 1 人が w_t の入札を行い、残りの $(n-2)$ 人の入札者が w_t 以下の入札を行う確率を表している。よって勝利したデータに対する対数尤度関数 $L_W(\theta)$ は

$$\begin{aligned} L_W(\theta) &= \ln \prod_{t=1}^{T_W} (n-1)f(w_t; \theta)F(w_t; \theta)^{n-2} \\ &= \sum_{t=1}^{T_W} \ln(n-1)f(w_t; \theta)F(w_t; \theta)^{n-2} \\ &= \sum_{t=1}^{T_W} (\ln(n-1) + \ln f(w_t; \theta) + (n-2)F(w_t; \theta)) \end{aligned}$$

となる。

2. 自己が敗北した場合

自己が b_t の入札をして敗北し、記号 N を観測する確率密度は

$$1 - F(b_t)^{n-1}$$

である。これは自己を除く $(n - 1)$ のうち少なくとも 1 人が b_t より大きい入札を行う確率を表している。よって敗北したデータに対する対数尤度関数 $L_L(\theta)$ は

$$\begin{aligned} L_L &= \ln \prod_{t=1}^{T_L} (1 - F(b_t; \theta)^{n-1}) \\ &= \sum_{t=1}^{T_L} \ln (1 - F(b_t; \theta)^{n-1}) \end{aligned}$$

となる。

以上より、 T 件のデータに対する対数尤度関数 $L(\theta)$ は

$$\begin{aligned} L(\theta) &= L_W + L_L \\ &= \sum_{t=1}^{T_W} (\ln(n - 1) + \ln f(w_t; \theta) + (n - 2)F(w_t; \theta)) + \sum_{t=1}^{T_L} \ln (1 - F(b_t; \theta)^{n-1}) \end{aligned}$$

となり、これを最大にする最尤推定量 $\tilde{\theta}$ を求める。

第 4 章 モデル作成

本章では入札者の財に対する評価が、ある分布関数 (評価値分布) に依存すると仮定し、この分布に対して各広告枠毎にパラメトリック推定を行う。

4.1 敵入札者の入札モデル

本研究では、JJ.Laffont et al.[8] にならって、入札者は対数正規分布に従って入札を行うと仮定する。また落札額の分布形状から見ても、対数正規分布を仮定して良いと考えられる。対数正規分布とは確率論および統計学における連続確率分布の一種であり、この分布に従う確率変数の対数をとったとき、対応する分布が正規分布になる性質を持つ。対数正規分布の密度関数は

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma x}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right) \quad (x > 0)$$

で与えられ、平均 $E(x)$ および分散 $V(x)$ はそれぞれ

$$\begin{aligned} E(x) &= e^{(\mu + \frac{\sigma^2}{2})} \\ V(x) &= e^{(2\mu + \sigma^2)}(e^{\sigma^2} - 1) \end{aligned}$$

である。

推定を行う際には、観測可能なデータのうちのデータが分布形状に影響を与えるかを考える必要がある。本稿では、評価値分布に影響を与える可能性があるものとして以下の2つを考える。

- 各オークションの開催時間
- 入札者の残り予算

ディスプレイ広告オークションは1日を通して行われており、時間帯によって異なる性質を持つユーザが来ると考えることができる。例えば18~24時には10代の学生がよく訪れるのであれば、ある広告枠はその時間帯の評価値が大きくなる、というようなことが考えられる。また入札者の残り予算については、予算が少なくなってくると、高い評価値を持っている広告主が予算切れで入札出来なくなり、低い評価値の広告主の選出が多くなってくるため評価値分布が変化する、といった例が考えられる。しかし、入札者の残り予算に関しては直接観測できない。そこで、入札者の予算はオークションの勝利回数が増加するにつれて単調に減少していくと仮定して、1日のオークションが開始される0時からの勝利の累計回数を1日の総勝利数で正規化した値(正規勝利数)を用いて予算の間接的な観測を試みた。

これら2つのデータを考慮したモデルを作成する。まず入札者の1日のオークションが開催される0時からの勝利数に関して以下の2つの仮説を立て、これに対応するモデルを用いて推定する。

モデル0(予算無依存型)

分布形状がオークションの勝利回数に影響を受けない場合、入札額 X の密度関数はパラメータ $\theta = (\theta_1, \theta_2)$ に対して

$$f_0(x, \theta) = \frac{1}{\sqrt{2\pi\theta_2}x} \exp\left(-\frac{(\log x - \theta_1)^2}{2\theta_2^2}\right)$$

で与えられる。

モデル1(予算無依存型)

分布形状がオークションの勝利回数に影響を受ける場合、入札額 X の密度関数はパラメータ $\theta = (\theta_{11}, \theta_{12}, \theta_2)$ および敵の正規勝利数 W に対して

$$f_1(x, \theta; W) = \frac{1}{\sqrt{2\pi\theta_2x}} \exp\left(-\frac{(\log x - (\theta_{11} + \theta_{12}W))^2}{2\theta_2^2}\right)$$

で与えられる。

この2つのモデルを比較することによって、入札者の予算の多寡が評価値分布に影響を与えるかどうかを調べることができる。また、各オークションの開催時間による評価値分布の影響を考えるため、さらに以下の2つのモデルを考える。

モデル full(時間無依存型)

分布形状が各オークションの開催時間に影響を受けない場合、各入札者の評価値分布は1日を通して同じである。

モデル quad(時間依存型)

分布形状が各オークションの開催時間に影響を受ける場合、各入札者の評価値分布は、1日を4分割した0~6時、6~12時、12~18時、18~24時のそれぞれで異なる。

時間帯に応じて評価値分布が大きく変化するか、あるいは1日を通して同じ評価値分布を用いても差が無いのかを、この2つのモデルから判断する。時間帯の分割数に関しては、例えば1時間単位のように細かく分割することは可能であるが、ここでは分割数を4とし、分割数がどのような影響を及ぼすかについては第5章で議論する。この2つのモデルによってオークションの開催時間が評価値分布に影響を与えるかを調べることができる。

以上から入札者の予算による影響およびオークションの開催時間による影響について図2のような4つの仮説が立てられる。予算の変化と時間の変化は相互に依存するものであり、完全に分離できるものではないが、ここでは評価値分布に与える主要な要因として予算の影響と時間の影響を分けて考えている。これらの4つのモデルを最尤推定量を用いて推定する。

4.2 最尤法によるモデル推定

前節で作成した4つのモデルに対して最尤法を行い、最尤推定量を求める。利用可能なデータとしては、3.2節で述べた入札額、落札額、オークションの開催

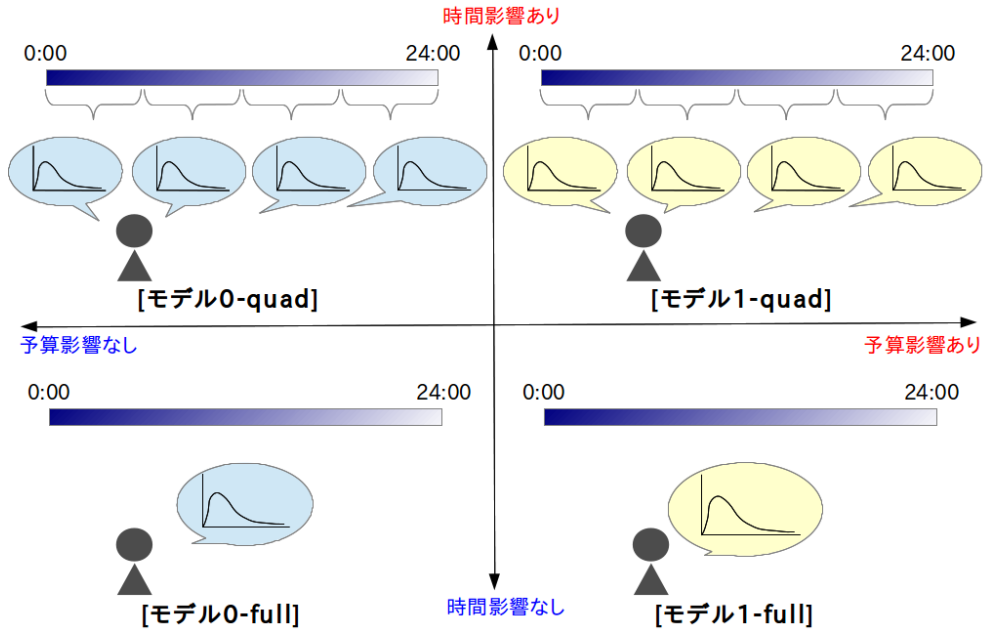


図2: 作成する4つの入札モデル

時間と，4.1節で述べた正規勝利数がセットになった3000件のオークションデータ集合 T を50個の広告枠それぞれから取ってきたものである．モデル full に対してはこの3000件のオークションデータ T を全て用いてあるひとつの評価値分布を推定する．一方モデル quad に対しては，3000件の中からオークションの開催時間が0～6時，6～12時，12～18時，18～24時のデータで分割し（それぞれのデータ集合を T_1, T_2, T_3, T_4 とする），それぞれに対して最尤法を行い，4つの評価値分布を推定する．

3.3節で導出したように，最尤法で用いる対数尤度関数は

$$\begin{aligned}
 L(\theta) &= L_W + L_L \\
 &= \sum_{t=1}^{T_W} (\ln(n-1) + \ln f(w_t; \theta) + (n-2)F(w_t; \theta)) + \sum_{t=1}^{T_L} \ln(1 - F(b_t; \theta)^{n-1})
 \end{aligned}$$

であり，これをパラメータ θ について最大化する．パラメータの推定結果は付録に回すとして，ここでは50個の広告枠毎に求められた対数尤度の平均値を比較する．表1はその結果である．

モデル full とモデル quad で尤度が大きく異なるのは，実オークションデータの標本サイズが異なるためである．またモデル quad で，時間が経つにつれて（24時に近づくとつれて）対数尤度が大きくなっているのは，オークションが時間が経つにつれて活発になり，それだけ標本数が大きくなっているからである．ま

表 1: 4つのモデルの対数尤度の平均値

		モデル 0-full	モデル 0-quad	モデル 1-full	モデル 1-quad
1 日	0~6 時	647.52	77.45	551.478	77.70
	6~12 時		109.13		108.58
	12~18 時		185.73		181.90
	18~24 時		285.16		274.41

たこの推定された分布形状の一例を図 3 に示す。これは敵入札者が model0-full モデル (予算無依存, 時間無依存モデル) に従って入札を行っていると仮定した時に, 広告枠 ID=24 に自分が 0.084 円を入札して勝利した時の落札額の分布形状である。棒グラフは実オークションデータのヒストグラムを表し, 曲線が最尤法によって推定された分布形状を表す。グラフ中には一部飛び出た値があ

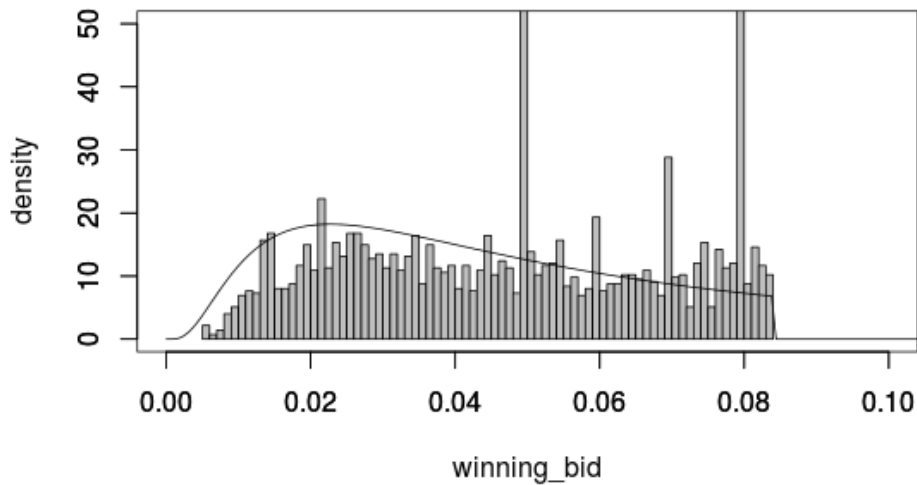


図 3: 0.084 円で勝利した時の落札額の分布形状

り, 実際の落札額分布を正確に表現できていない部分もある。この理由としては, 広告には様々なタイプがあるが, 本研究ではそれらを特に区別して扱っていない, ということが考えられる。例えば, WEB ページを訪れたユーザの性質 (年齢, 性別, 興味のある分野) などによって広告を掲載するかどうかを決定す

ることがディスプレイ広告オークションでは可能であるが、本研究ではユーザの違いを考慮していないため、落札額分布を正確に表現できなかったと考えられる。この点の詳細な分析は今後の課題となる。

表1の推定結果では model-full と model-quad の標本サイズが異なるため、優劣の比較ができない。そのため次節ではこの推定結果に対して、全てのモデルで同じ標本サイズの実オークションデータを用いて尤度を求め、モデルの優劣を議論する。

4.3 対数尤度の比較

モデル full の評価値分布に対して実データ $T_1 \dots T_4$ を当てはめることで各時間帯における評価値分布の対数尤度を求め、4つのモデルを比較する。

表2は50個の広告枠に対する対数尤度の平均を示している。対数尤度を比較

表2: 対数尤度の比較 (平均 (分散) で記載。太字はその時間帯での最大対数尤度)

	モデル 0-full	モデル 0-quad	モデル 1-full	モデル 1-quad
0～6時	74.56 (8571.37)	77.45 (8785.99)	55.37 (6795.90)	77.70 (8843.49)
6～12時	106.51 (18518.85)	109.13 (18640.26)	86.51 (15851.64)	108.58 (18554.09)
12～18時	183.93 (59085.31)	185.73 (59581.24)	177.43 (57491.95)	181.90 (58697.00)
18～24時	282.52 (143893.31)	285.16 (145047.23)	232.17 (126087.37)	274.41 (138426.21)
平均	161.88 (57517.21)	164.27 (58013.68)	137.87 (51556.72)	160.65 (56130.20)

すると、0～6時の時間帯では、予算を考慮したモデル 1-quad が最も大きく、6～24時では予算を考慮しないモデル 0-quad が最も大きい。また、図4は50個の広告枠のモデル毎の内訳である。このグラフでは、それぞれの広告枠に対して、最も対数尤度の大きいモデルを求め、それらの数を表している。グラフにはモデル 0-quad とモデル 1-quad の2つのモデルしか表されていないが、これ

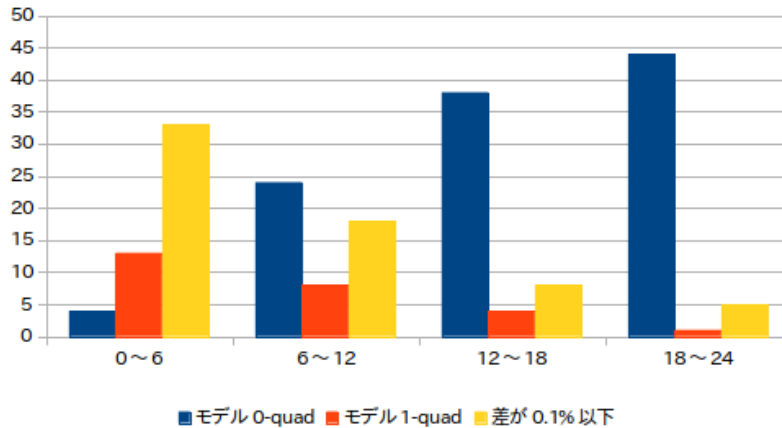


図4: 50個の広告枠に対するモデルの内訳

はモデル0-fullまたはモデル1-fullの対数尤度が最も大きかった広告枠が存在しないからである。また、モデル0-quadとモデル1-quadの対数尤度の差が、最も大きい対数尤度の差の0.1%以下である広告は、確率的な攪乱であると考えて2つのモデルに差はないとみなしている。例えば、0~6時の時間帯については、モデル0-quadの対数尤度が最も大きくなった広告枠は50個のうち4個、モデル1-quadの対数尤度が最も大きくなった広告枠は13個、2つのモデルに差が見られなかったものは33個となっている。

このグラフからも、0~6時の時間帯ではモデル1-quadの方が良い(対数尤度が大きい)広告枠の数がモデル0-quadよりも多い。対数尤度を比較するだけでは、どのモデルが最良であるか判断がつかない。そこで第5章ではモデル選択の指標として赤池情報量規準(Akaike's Information Criterion:AIC)を用いて、本研究で推定する評価値分布を一意的に決定する。

第5章 モデル選択

本章では第4章で作成した4つのモデルについて、AICという指標を用い、モデル選択を行う。

5.1 AICとは

AICとは、統計モデルの良さを評価するための指標である[9]。AICはモデルの複雑さとデータへの適合度のバランスを取るために使用されている。一般に、ある観測データを統計的に説明するモデルを作成しようとする時、パラメータ

の数や次数を増やせば増やすほどデータの当てはまりは良くなっていく。しかし、それは「たまたま得られたデータへの当てはまりの良さを向上させるための特殊化」であり、その統計モデルの予測の良さを損なっていく。この問題を過適合 (Over fitting) と言う。

過適合の例として、サイコロを 100 回投げた時を考える。それぞれの目が出た回数を表 3 に示す。このとき、 x の目が出る回数の確率変数 Y が従うモデル

表 3: サイコロを 100 回投げた時の分布

サイコロ	回数
1	18
2	13
3	22
4	11
5	19
6	17

は複数考えることができるが、ここではパラメータ b に対して

$$y = b_0$$

$$y = b_0 + b_1x + b_2x^2 + b_3x^3 + b_4x^4 + b_5x^5$$

とパラメータが最小の時 (最小モデル) と最大の時 (最大モデル) を考えてみる。このとき、それぞれのモデルを推定し、グラフを描くと図 5,6 となる。最小モ

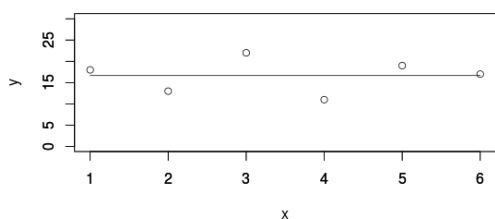


図 5: パラメータ数 1 の最小モデル

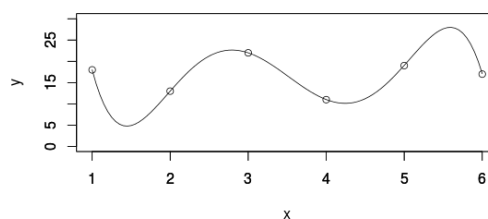


図 6: パラメータ数 6 の最大モデル

デルに比べて最大モデルの方が実データとの当てはまりが良くなっているのが

分かる。このように、一般にパラメータ数が大きいほどデータへの当てはまりは良くなっていく。しかし、これは今回たまたま観測されたデータに対しての当てはまりが良くなるように特殊化したに過ぎない。もう一度同じ試行をして観測されたデータが表4である場合、図7に示す通り最大モデルは実データにうまく当てはまらないことが分かる。

表4: サイコロを100回投げた時の分布(2回目)

サイコロ	回数
1	19
2	17
3	15
4	20
5	15
6	14

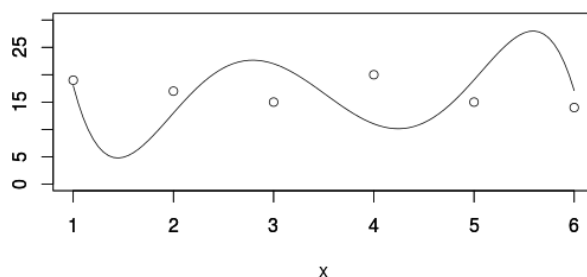


図7: 最大モデルと2回目の試行による観測データの当てはまり

これは観測データとの適合度を向上させるためにパラメータ数を多くしたことにより、ノイズや偶発的な変動にも無理に合わせるようなモデルを推定したことが原因である。観測データとの適合度は確かに向上するが、未知のデータに対する予測の精度が低下してしまう。

過適合を避けるためにはモデルのパラメータ数を抑える必要があるが、パラメータ数が少なすぎると適合度が小さくなってしまう。このようにパラメータ

数と適合度の間にはトレードオフの関係があり，両者をどのように決定すればよいか問題になる．AICは「当てはまりの良さではなく予測の良さがモデルの良さ」だと考え，この問題に対して一つの解を与える．AICが最小のモデルを選択すれば，多くの場合良いモデルとなる．

AICを求める公式は以下ようになる．

$$AIC = -2 \ln L + 2k$$

ここで L は最大尤度， k は最尤推定したモデルのパラメータ数である．

5.2 AICによるモデル選択

AICに基づいて，これまでに検討した4つのモデルの中から最も良いモデルを選択する．今回のモデルに関しては，予算の影響を考慮しないモデル0のパラメータ数は $k = 2$ となり，考慮するモデル1のパラメータ数は $k = 3$ である．公式にしたがって各モデルのAICを求めると表5のようになる．

表5: AICの比較(平均(分散)で記載．太字はその時間帯での最小値)

	モデル 0-full	モデル 0-quad	モデル 1-full	モデル 1-quad
0～6時	-145.12 (34285.50)	-150.90 (35144.00)	-104.74 (27183.61)	-149.40 (35373.96)
6～12時	-209.03 (74075.40)	-214.26 (74561.02)	-167.02 (63406.55)	-211.16 (74216.33)
12～18時	-363.86 (236341.24)	-367.45 (238325.00)	-348.86 (229967.81)	-357.80 (234788.00)
18～24時	-561.05 (575573.22)	-566.31 (580188.93)	-458.33 (504349.50)	-542.81 (553704.84)
平均	-319.76 (230068.84)	-324.73 (232054.72)	-269.74 (206226.87)	-315.29 (224520.78)

表から，AICの平均はどの時間帯においてもモデル0-quadが最も小さいことが分かる．また図4と同じように，広告枠のモデル毎の内訳を図8に示す．50個の広告枠のうち，モデル1-quadのAICが最小になるものは各時間帯において

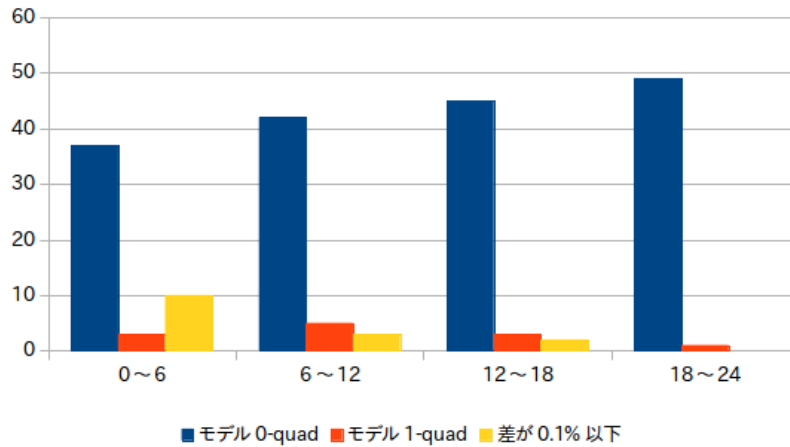


図 8: 50 個のモデルに対するモデルの内訳

それぞれ、3個、5個、3個、1個である。よってどの時間帯においても、95%以上の広告枠はモデル 0-quad が 4 つのモデルの中で 1 番良いモデルとなる。

予算の影響を考慮したモデル 1-full と比べると、モデル 0-quad は AIC 基準にして約 20%の改善となる。この結果から入札者の評価値分布はオークションが開催される時間帯にのみ影響を受け、入札者の予算に影響を受けないことが分かる。なお、予算の影響を受けないとは、あくまで model1 で仮定した形での影響を受けないということである。予算の影響を考慮しないモデルの方が良い結果となった理由としては、予算を間接的に観測したことが原因に挙げられる。他入札者の予算というのは実際に観測することは不可能であり、今回は予算を他入札者の累計勝利回数という形で間接的に観測した。これは勝利回数が多くなればそれだけ予算を消化しているため、回数が増える程単調に予算が減少していく、という仮定の下で立てた仮説であった。Yahoo!オークションなどの「買い手=入札者」が成立するオークションであればこの仮定は正しいが、本研究であつかうディスプレイ広告オークションでは入札者は DSP である。DSP は複数の広告を扱っており、各広告毎に予算が割り当てられているので、広告 A の予算を用いて A を入札し続け、予算が尽きたら広告 B の予算を用いて B の入札を行い... といったように必ずしも予算が単調に減少するわけではない。よって予算 (勝利回数) の影響を考慮したモデルの予測が悪くなってしまったと考えられる。

5.3 時間帯の分割数

ここまでで、予算による影響を考えずに、オークションの時間帯によって評価値分布を変更するモデル 0-quad が最も良い予測が可能であることが分かった。では、時間帯をどれだけ分割する必要があるだろうか。上で議論したモデルでは、1日を0～6時、6～12時、12～18時、18～24時の4つに分割して、それぞれの評価値分布を推定したが、最適な分割数は不明である。分割数を多くすれば、より細かく評価値分布を推定できるため、対数尤度は当然上がっていくと考えられる。しかし、分割数を大きくすると各分割に含まれるデータが少なくなってしまう、十分な推定が行えなくなるし、操作が煩雑になってしまう。またデータを増やすにしても時間的なコストがかかってしまう。

ここでは、参考として1日を均等に1～6分割したモデルの対数尤度を比較し、分割数によってどのように対数尤度が増加していくかを調べる。

1日のオークションデータ 3000 件を実データとして、各モデルの対数尤度を求めたものを、表 6 および図 9 に示す (参考として分割なしのモデル 0-full の尤度も載せておく)。図に示す通り、分割数が大きくなるにつれ対数尤度は増加

表 6: 分割数による対数尤度一覧 (平均 (分散) で記述)

分割なし	2分割	3分割	4分割	5分割	6分割
647.52	652.55	656.44	657.47	661.16	662.73
(571533.17)	(576701.63)	(578170.56)	(578865.96)	(579292.02)	(579563.03)

していくことが確認できる。

第 6 章 入札額および落札額の推定

第 5 章では、最尤法と AIC を用いて、入札者の財に対する評価値分布モデルを推定することができた。他の入札者の評価値分布が得られれば、それを元に解析的に最適な入札戦略が得られると考えるかもしれない。しかし、広告枠間の依存関係や入札要求数自体の変動など、実際のオークションは複雑である。よって、シミュレーションを用いて入札戦略の性能を評価できればよい。シミュレーションを実現するには自己の入札額に対する勝利確率や、勝利時の落札額の期待値、敗北時の 1 位の入札額および落札額の期待値が得られればよい。本章で

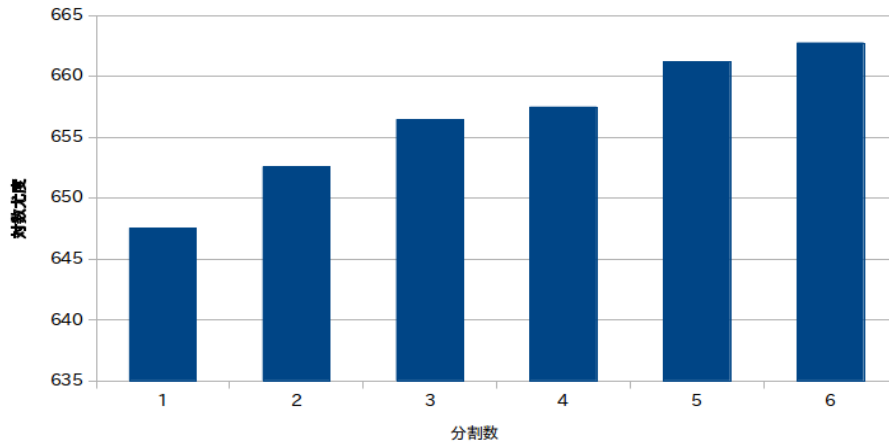


図 9: 分割数による対数尤度の比較

は，それらの導出法を示す．

6.1 オークション勝利時の落札額

各種確率を求める前に条件付き確率の説明を行う．条件付き確率 (Conditional Probability) とは，ある事象 B が起こるという条件の下で別の事象 A の起こる確率をいい， $P(A|B)$ と表し以下のように求めることができる．

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

ここで， $P(A \wedge B)$ は事象 A と B の同時確率 (Joint Probability)， $P(B)$ は事象 B の周辺確率 (Marginal Probability) という．条件付き確率では観測された情報によって確率を改訂することができる．

これまでと同様に，オークションには自己を含め n 人の入札者が参加しており，他入札者は財に対して同一の評価値分布 f を持ち，評価値通りに入札を行うとする．このとき，自己が b の入札を行なって勝利する事象を B，オークションの落札額が x になる事象を A とすると，同時確率 $P(A \wedge B)$ および事象 B の周辺確率 $P(B)$ はそれぞれ

$$\begin{aligned} P(A \wedge B) &= (n-1)f(x)F(x)^{n-2} \\ P(B) &= F(b)^{n-1} \quad (x < b) \end{aligned}$$

と表せる．前者は，自己を除く $(n-1)$ 人のうちの 1 人が x の入札をし，残りの

$(n - 2)$ 人が x より小さい入札をする確率を表し、後者は、自己を除く $(n - 1)$ 人が b より小さい入札をする確率を表す。よって、自己が b の入札を行なって勝利したという条件の下での落札額 x の条件付き確率密度関数 $g_w(x|b)$ は、

$$\begin{aligned} g_w(x|b) &= P(A|B) \\ &= \frac{(n - 1)f(x)F(x)^{n-2}}{F(b)^{n-1}} \quad (x < b) \end{aligned}$$

であり、入札額 b で勝利した時の落札額の期待値 $E_w(b)$ は

$$E_w(b) = \int_0^b x \cdot g_w(x|b) dx$$

となる。例として、広告枠 ID18 を財とするオークションに勝利したときの落札額の期待値を図 10 に示す。

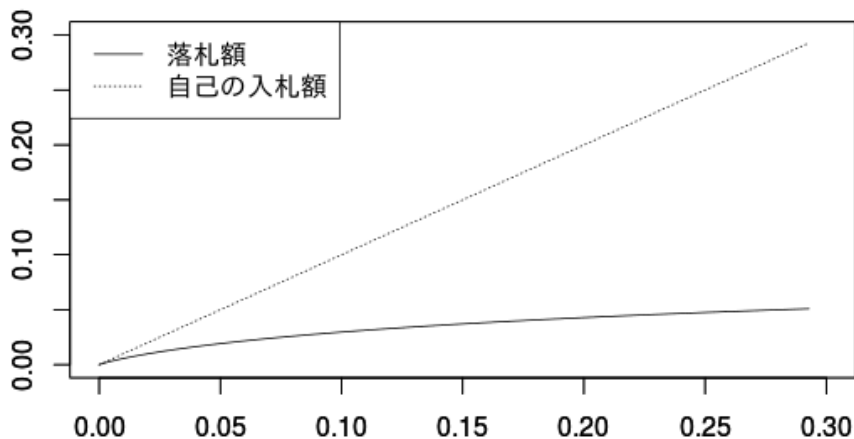


図 10: 広告枠 ID18 で勝利したときの落札額の期待値

6.2 オークション敗北時の 1 位の入札額および落札額

オークションで自己が敗北した時に、1 位の入札額とその時の落札額がいくらになるのかを求める。1 位の入札額を知ることによって自己がいくら入札を行えば勝利することができたのか、また落札額を知ることによって他入札者がいく

ら予算を消化したのかを知ることができる。

自己が b の入札を行なって敗北したという条件の下で 1 位の入札額 x の条件付き確率密度関数 $h_l(x|b)$ は

$$h_l(x|b) = \frac{(n-1)f(x) \cdot F(x)^{n-2}}{1-F(b)^{n-1}}$$

となる。分母は、少なくとも 1 人は b より大きい入札を行う確率を、分子は、 $(n-1)$ 人のうち 1 人が x の入札を行い、残りの $(n-2)$ 人が x 以下の入札を行う確率を求めることで導出できる。

また、同条件下での落札額 x の条件付き確率密度関数 $g_l(x|b)$ は、ディスプレイ広告オークションが第二価格入札を採用していることから、2 位の入札額の確率密度を求めればよい。よって、

$$g_l(x|b) = \begin{cases} \frac{(n-1)f(x) \cdot (n-2)(1-F(x)) \cdot F(x)^{n-3}}{1-F(b)^{n-1}} & (x > b) \\ \frac{(n-1)(1-F(x)) \cdot F(x)^{n-2}}{1-F(b)^{n-1}} & (x = b) \end{cases}$$

と導出できる。落札額は、自己の入札 b が 2 位になっている場合 ($x = b$) とそうでない場合 ($x > b$) でそれぞれ確率密度が異なるので注意する必要がある。以上より、敗北時の 1 位の入札額および落札額 (2 位の入札額) のそれぞれの期待値を求めると、

$$\int_b^\infty x \cdot h_l(x|b) dx$$

$$\int_b^\infty x \cdot g_l(x|b) dx$$

となるが、積分範囲が $[b, \infty]$ となっており、入札者の入札額が現実的でない場合も考慮してしまっている。実際、マイクロアド社から入手した実オークションデータを分析すると、何十円といった入札額はなく、広告枠によって最高入札額にばらつきがあった。そこで、各広告枠に対する最高入札額を求め、それを v_{max} とし、他入札者は $[0, v_{max}]$ の範囲の入札しか行わない、と仮定すると、1 位の入札額および落札額のそれぞれの期待値 $E_{(l,1)}(b)$, $E_{(l,2)}(b)$ は

$$E_{(l,1)}(b) = \frac{\int_b^{v_{max}} x \cdot h_l(x|b) dx}{\int_b^{v_{max}} h_l(x|b) dx}$$

$$E_{(l,2)}(b) = \frac{\int_b^{v_{max}} x \cdot g_l(x|b) dx}{\int_b^{v_{max}} g_l(x|b) dx}$$

となる。 $[b, v_{max}]$ で $g_w(x|b)$ を積分することにより，落札額が $[b, v_{max}]$ の範囲になる全確率を 1 に正規化している。例として，広告枠 ID18 を財とするオークションに敗北したときの，1 位の入札額およびその時の落札額の期待値を図 11 に示す。

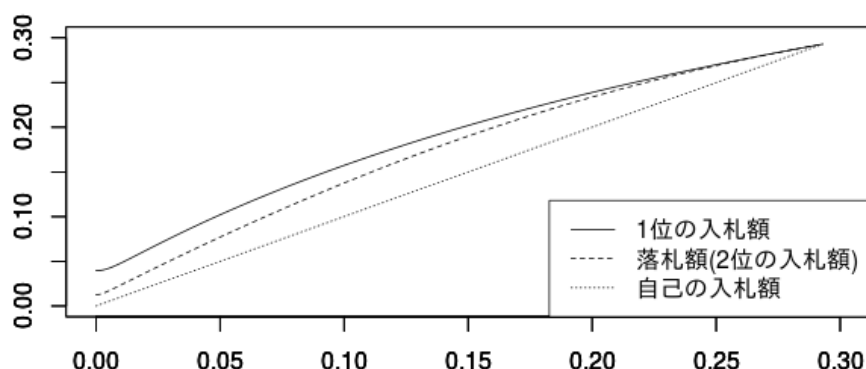


図 11: 広告枠 ID18 で敗北した時の 1 位の入札額，落札額の期待値

第 7 章 おわりに

本研究ではある DSP の立場から観測可能なデータのみを用いて，他 DSP が財に対して持つ評価値分布の推定を行った。まずは予算の影響の有無を考慮したモデルとオークションの開催時間の影響の有無を考慮したモデルを作成し，最尤法を用いてモデル推定を行った。最尤法の際に利用する対数尤度関数については，オークション理論を援用し，自己が勝利した時に観測できる落札額の確率，自己が敗北する確率を求めることで導出した。また，AIC を用いてモデル選択を行い，予算の影響を考慮せずに，時間帯によって異なる評価値分布を利用するモデルが 95% 以上の広告枠に対して最も優れていることを明らかにした。そして，1~6 分割したモデルで対数尤度を求め，分割数に応じて対数尤度が増加していくことを確認した。最後に，推定された評価値分布を用いて，自己の入札額に対する勝利確率や，勝利した時の落札額の期待値，または敗北したと

きの1位の入札額, 落札額の期待値を求める計算式を獲得した.

本研究の貢献は以下の通りである.

1. 部分的なデータのみを用いた評価値分布モデルの獲得

オークション理論に基づき, 他の入札者が財に対して同一の評価値分布を持つという仮定のもとで評価値分布の推定を可能にした. それによって自己の入札額に対する勝利確率や, 敗北した場合の1位の入札額の期待値に関する計算式を導出できた.

2. 経時変化を考慮した評価値分布モデルの獲得

実オークションデータを用いて最尤推定を行うことによって, 予算制約による評価値分布を考慮したモデルよりも, 予算を直接考慮せずに時間帯毎にモデルを作成する方が, より実データに合致する結果が得られることを明らかにした (AIC 基準で約 20%の改善). また時間帯の分割数も複数の場合を検討することで, 分割数に応じて対数尤度は単調に増大していくことを確認した.

謝辞

本研究を行うにあたり, 貴重なデータをご提供いただきました株式会社マイクロアド様に深く感謝申し上げます. そして本研究を行うにあたり熱心なご指導, 御助言をして下さいました松原繁夫准教授に厚くお礼申し上げます. また多くの有益な御助言をして下さいました石田亨教授をはじめ, 石田・松原研究室の皆様方に心より感謝いたします.

参考文献

- [1] 梶井厚志, 松井彰彦: ミクロ経済学戦略的アプローチ, 日本評論社 (2000).
- [2] 渡辺隆裕: ゼミナールゲーム理論入門, 日本経済新聞出版社 (2008).
- [3] 渡邊泰典: ネットオークションにおける評価額分布の構造型実証分析による推定, 赤門マネジメント・レビュー, pp. 1-16 (2008).
- [4] Milgrom, P. R.: *Putting auction theory to work*, Cambridge University Press (2004).
- [5] Paarsch, H. J. and Hong, H.: An introduction to the structural econometrics of auction data, *MIT Press Books*, Vol. 1 (2006).

- [6] Vickrey, W.: Counterspeculation, auctions, and competitive sealed tenders, *The Journal of finance*, Vol. 16, No. 1, pp. 8–37 (1961).
- [7] Klemperer, P.: Auction theory: A guide to the literature, *Journal of economic surveys*, Vol. 13, No. 3, pp. 227–286 (1999).
- [8] Laffont, J.-J., Ossard, H. and Vuong, Q.: Econometrics of first-price auctions, *Econometrica: Journal of the Econometric Society*, pp. 953–980 (1995).
- [9] Akaike, H.: A new look at the statistical model identification, *Automatic Control, IEEE Transactions on*, Vol. 19, No. 6, pp. 716–723 (1974).

付録：推定データ

A.1 モデルの推定データ

本研究で推定された4つのモデルのうち、最も尤度が大きくなったモデル 0-quad の最尤推定量を表 A.1 に示す。

表 A.1: モデル 0-quad(上段: θ_1, θ_2) 下段:対数尤度)

ID	0～6 時	6～12 時	12～18 時	18～24 時
1	(-4.571,0.307) 235.363	(-4.525,0.231) 155.919	(-4.542,0.329) 222.844	(-4.544,0.292) 487.032
2	(-4.143,0.093) 97.371	(-4.188,0.197) 86.399	(-4.124,0.160) 29.511	(-4.116,0.231) 34.737
3	(-2.658,0.271) 47.132	(-2.683,0.454) 35.922	(-2.699,0.436) 107.242	(-2.705,0.408) 256.421
4	(-1.996,0.228) 43.443	(-2.003,0.428) 70.596	(-1.995,0.385) 89.748	(-1.961,0.309) 73.430
5	(-5.196,0.948) 212.438	(-5.158,0.981) 123.663	(-5.225,0.849) 375.145	(-5.473,0.974) 1673.087
6	(-4.198,0.873) 33.555	(-4.254,0.691) 81.607	(-4.447,1.116) 62.641	(-4.534,0.968) 163.005
7	(-4.225,1.348) 40.270	(-4.048,1.408) 136.528	(-4.057,1.601) 203.916	(-4.256,1.706) 201.658
8	(-3.265,1.236) 9.778	(-2.875,1.006) 38.506	(-2.884,1.087) 59.396	(-2.988,1.095) 65.029
9	(-4.477,1.496) 64.758	(-4.548,1.631) 51.425	(-4.583,1.370) 174.414	(-4.734,1.515) 345.916
10	(-2.456,0.053) 57.363	(-2.492,0.164) 169.192	(-2.490,0.110) 459.045	(-2.477,0.080) 470.987
11	(-2.990,1.179) 5.531	(-2.531,1.090) 3.467	(-2.697,1.325) -1.985	(-2.695,1.333) -2.909
12	(-2.324,0.935)	(-1.959,0.987)	(-2.054,1.049)	(-2.202,1.031)

表 A.1: モデル 0-quad(上段: θ_1, θ_2) 下段:対数尤度)

ID	0～6 時	6～12 時	12～18 時	18～24 時
	1.284	-19.865	-27.759	-6.634
13	(-6.296,2.231) 365.870	(-6.213,2.707) 477.986	(-5.986,3.306) 339.439	(-6.463,3.248) 528.610
14	(-2.484,1.091) 1.950	(-2.927,2.088) -14.677	(-2.709,1.849) -11.829	(-2.365,2.258) -12.567
15	(-5.963,1.564) 168.328	(-5.734,1.541) 505.621	(-5.937,1.396) 927.330	(-6.009,1.749) 446.955
16	(-3.720,2.468) -0.517	(-3.477,3.181) -9.871	(-3.495,3.139) -14.128	(-3.841,2.862) 1.434
17	(-4.093,2.641) 6.384	(-3.759,2.653) -0.382	(-4.019,3.390) 9.297	(-3.779,2.653) 2.050
18	(-5.264,1.583) 206.376	(-5.234,1.751) 269.517	(-5.187,2.166) 243.883	(-5.461,1.915) 491.724
19	(-4.166,1.082) 53.378	(-4.158,1.146) 85.558	(-4.286,1.402) 87.452	(-4.293,1.547) 132.392
20	(-4.743,2.232) 30.777	(-4.530,2.238) 193.306	(-4.668,2.284) 234.929	(-4.764,2.223) 85.276
21	(-3.970,1.651) 19.768	(-3.459,2.037) -3.283	(-3.557,1.823) 6.780	(-3.711,1.907) 24.862
22	(-5.277,1.572) 205.877	(-5.299,1.474) 253.587	(-5.321,1.179) 878.781	(-5.407,1.399) 508.866
23	(-3.493,1.035) 24.766	(-3.348,1.045) 34.552	(-3.329,1.106) 70.292	(-3.357,1.191) 39.305
24	(-4.195,1.121) 46.096	(-4.251,1.346) 115.884	(-4.382,1.386) 172.854	(-4.291,1.506) 208.469
25	(-3.609,1.098) 26.118	(-3.543,0.901) 28.764	(-3.207,0.642) 24.887	(-3.655,0.861) 138.333
26	(-3.325,1.298)	(-3.485,1.362)	(-3.331,1.313)	(-3.377,1.368)

表 A.1: モデル 0-quad(上段: θ_1, θ_2) 下段:対数尤度)

ID	0～6 時	6～12 時	12～18 時	18～24 時
	8.268	23.323	8.892	16.885
27	(-3.932,1.165)	(-3.864,1.393)	(-3.786,1.157)	(-3.940,1.371)
	35.749	19.454	48.473	111.295
28	(-3.765,1.328)	(-3.579,1.326)	(-3.620,1.343)	(-3.617,1.258)
	28.146	56.043	84.887	115.631
29	(-2.597,1.390)	(-2.454,1.499)	(-2.500,1.264)	(-2.743,1.389)
	-3.730	-15.158	-7.326	-4.622
30	(-3.284,2.624)	(-3.314,2.853)	(-3.010,1.956)	(-3.296,2.534)
	-5.853	-12.877	-15.735	-20.600
31	(-3.024,1.150)	(-3.113,1.196)	(-3.181,1.320)	(-3.092,1.197)
	4.504	20.636	17.927	7.648
32	(-3.671,1.428)	(-3.232,2.010)	(-3.375,1.492)	(-3.757,1.861)
	31.201	-7.420	22.842	28.370
33	(-4.033,0.963)	(-4.111,1.384)	(-3.859,1.559)	(-4.071,1.174)
	39.281	39.733	13.314	80.299
34	(-5.059,0.715)	(-5.092,0.787)	(-5.198,0.909)	(-5.215,0.905)
	296.096	457.870	398.160	1017.570
35	(-5.516,2.097)	(-5.678,1.856)	(-5.805,1.760)	(-5.659,2.416)
	86.651	492.660	601.254	232.047
36	(-5.866,1.221)	(-5.728,1.174)	(-5.768,1.006)	(-5.715,1.035)
	248.098	287.127	920.434	1172.961
37	(-4.033,1.432)	(-3.815,1.341)	(-3.893,1.597)	(-4.001,1.629)
	28.817	123.691	122.267	153.626
38	(-2.195,0.214)	(-2.212,0.468)	(-2.139,0.443)	(-2.214,0.625)
	27.394	35.018	5.183	5.916
39	(-4.968,1.370)	(-5.000,0.929)	(-5.448,1.047)	(-5.679,1.111)
	53.822	76.486	198.998	580.737
40	(-2.206,0.971)	(-2.700,0.882)	(-3.122,1.107)	(-2.985,1.404)

表 A.1: モデル 0-quad(上段: θ_1, θ_2) 下段:対数尤度)

ID	0～6 時	6～12 時	12～18 時	18～24 時
	-0.933	1.823	6.157	-2.036
41	(-4.868,1.304) 106.304	(-4.676,1.327) 82.316	(-4.818,1.471) 144.653	(-5.003,1.622) 352.677
42	(-4.823,1.678) 72.012	(-4.681,1.681) 256.458	(-4.868,1.747) 396.714	(-4.797,1.810) 398.198
43	(-3.178,1.226) 1.121	(-3.244,1.047) 28.840	(-3.358,1.180) 44.745	(-3.514,1.117) 93.663
44	(-3.883,1.031) 42.795	(-3.926,1.157) 56.581	(-3.934,1.271) 56.881	(-3.755,1.313) 54.136
45	(-2.896,0.679) 27.220	(-2.817,0.613) 28.219	(-2.895,0.679) 58.420	(-2.968,0.652) 81.741
46	(-3.816,1.281) 10.075	(-3.837,0.932) 15.700	(-4.011,1.257) 19.053	(-4.266,1.291) 117.352
47	(-2.613,1.063) 8.486	(-2.626,1.198) 4.270	(-2.732,1.171) 17.094	(-2.802,0.977) 47.334
48	(-6.281,1.142) 253.936	(-6.121,1.224) 221.165	(-6.239,1.340) 372.603	(-6.050,1.164) 1155.098
49	(-5.340,1.129) 235.728	(-5.119,1.129) 131.708	(-5.294,1.035) 440.942	(-5.423,0.937) 1073.276
50	(-5.020,0.904) 233.910	(-4.984,0.911) 162.978	(-5.235,0.844) 585.402	(-5.397,1.031) 1031.169

A.2 推定された評価値分布

モデル 0-quad の評価値分布が経時変化と共にどのように遷移していくかをグラフに示す。50 個の広告枠の評価値分布を調べたところ、評価値分布の遷移は大きく分けて以下の 3 つの型があることが分かった。

単調減少型

評価値が時間と共に単調に低くなっていく型。評価値分布は時間と共に左に移動する。

山型

オークションの開始と終了付近の評価値が低く、中盤に評価値が高くなっていく型。評価値分布はオークションの開始と共に右に移動し、その後左に移動する。

不変型

評価値分布が1日を通して大きく遷移しない型

図 A.1 から図 A.3 はそれぞれ ID が 19, 43, 7 の広告枠に対する評価値分布の遷移を表している。これらのグラフは、それぞれ単調減少型、山型、不変型の評価値分布となっている。

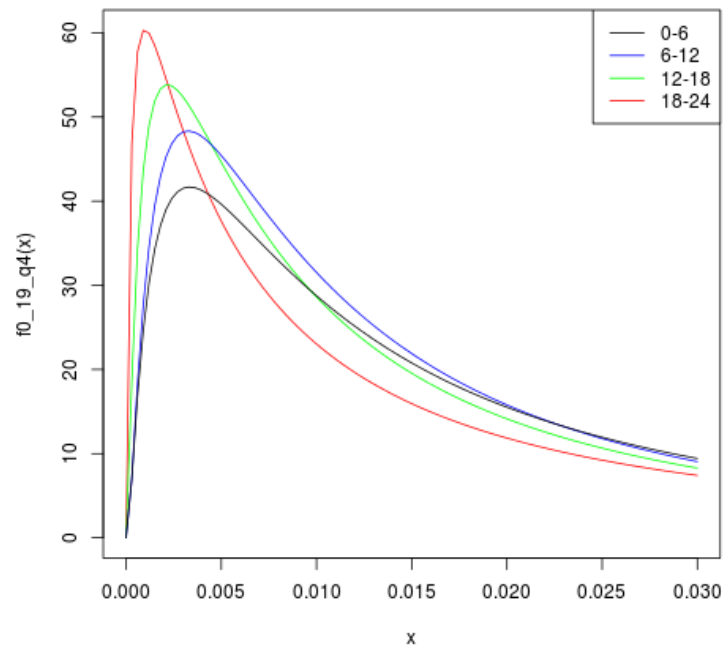


図 A.1: 単調減少型 (ID19) の評価値分布

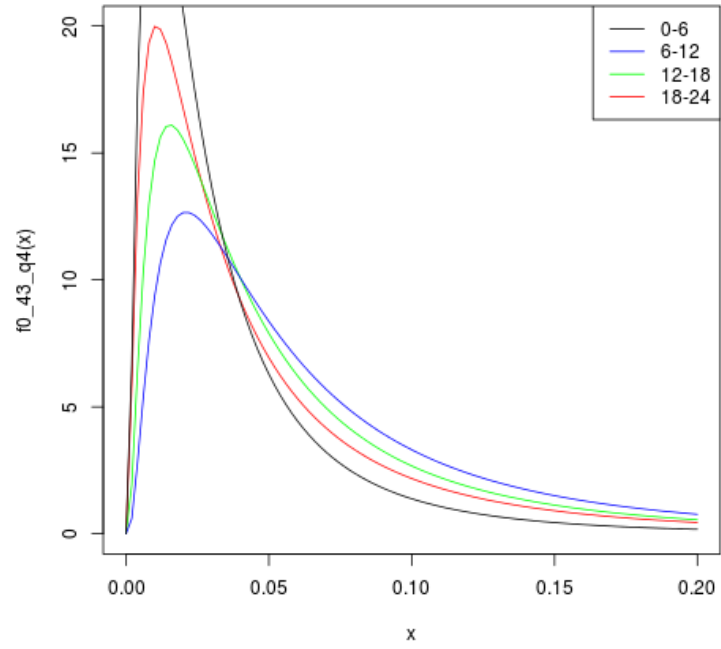


図 A.2: 山型 (ID43) の評価値分布

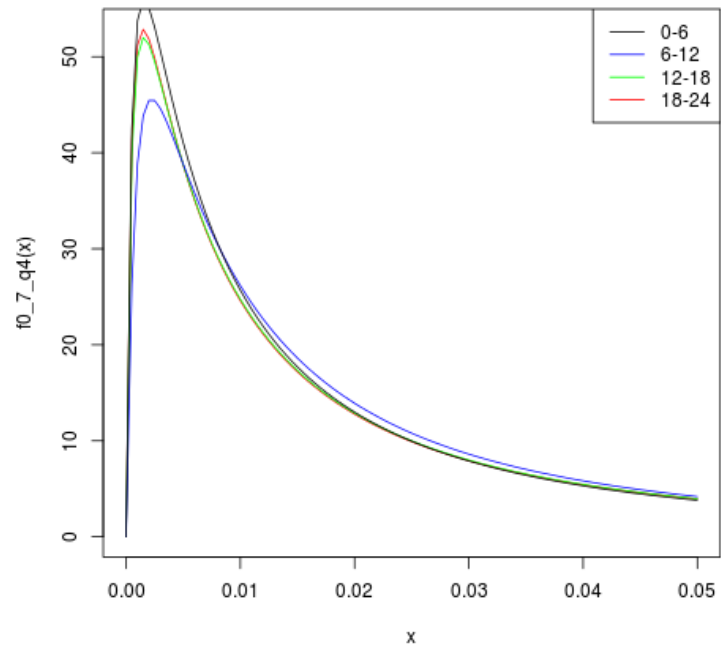


図 A.3: 不変型 (ID7) の評価値分布