Selling Impressions: Efficiency vs Competition

Dirk Bergemann   Tibor Heumann   Stephen Morris
Constantine Sorkin   Eyal Winter

APIOC 2021

Impressions and Digital Advertising

- web content is primarily monetized by ads
- opportunities to show ads to user browsing a website, "impressions", are traded via auctions...
- ... in search and in display advertising
- seller is publisher: website that user is visiting
- bidder is advertiser
- object of auction is viewer ("eyeball", "attention")
- example: publisher is nyt.com, bidder is Bank of America

Efficiency vs Competition in Digital Advertising

- publishers of advertising on the internet face a fundamental economic trade-off in deciding how much information to provide advertisers about viewers:
- more information implies a more efficient match of advertiser and viewer, and so more surplus to split between publisher and advertiser...
- ...but more information gives rise to a thinner market, and so more information rent for the advertiser
- Levin and Milgrom (2011) discuss this as an example of a more general "conflation" question: how to draw boundaries between goods?

Efficiency vs Competition

- equivalently: how much information would the seller (or publisher) like buyers (advertisers) to have about the good they are buying?
- different buyers may be given different information (which viewers are bundled in the market for impressions may differ across advertisers).
- **our question**: How much information would the seller (publisher) like buyers (advertisers) to have about their valuations of a good (an impression) in an auction?
    - a lot, to maximize efficiency?
    - a little, to maximize competition?
    - or something in between?

The (Abstract) Question in More Detail

- consider classic problem of second price auction of single object to buyers with symmetric independent private values.....
- .....but suppose the seller controls how much each buyer knows about his private value (without knowing the private value herself)
- would the seller prefer full information (buyers know their values perfectly), no information (buyers know nothing about their values), or something in between?
- with full information: efficient allocation but information rents - revenue is expectation of second highest value
- with no information: inefficiency but no information rent - revenue is common ex ante expected value

Answer

- optimal information structure is something in between.....
- in particular, low valuation buyers are told their values but high valuation buyers are pooled, i.e., just told that their value exceeds a critical threshold
- in fact, critical quantile where pooling starts depends only on the number of buyers (and is independent of the distribution of values)
- intuition: competition is lowest when there is a high winning value
- this is our main theoretical result and first main contribution

Selling Impressions

- in the market for digital advertising, the object being sold is a viewer impression
- viewers are typically heterogeneous in many dimensions, their demographic characteristics, their preferences, their (past) shopping behavior, their browsing history and many other aspects, observable and unobservable
- advertisers display a corresponding degree of heterogeneity in their willingness to pay for a match between their advertisement and a specific viewer
- today focus on digital advertising, many other applications, e.g. asset design as information design, how to bundle or not to bundle financial claims

Selling Impressions By Algorithms

- information of advertiser and of publisher jointly inform bidding in auction
- two prevalent algorithms of how the joint information enters into the bid formation:
  automated bidding and *manual bidding*
- in automated bidding , or short *autobidding*, seller offers a bidding algorithm that generates optimal bids for the advertisers given the disclosed information
- in manual bidding seller offers disclosure algorithm that generates information about attributes, each bidder then translates manually into bid for impression
- autobidding converts high-dimensional information across millions of impressions into bids with minimal latency

Model

Model (Basic)

- $i = 1, ..., N$ advertisers bid for viewer in second-price auction
- private values v_i symmetrically and independently distributed according to F
- publisher chooses a information structure (signal), symmetrically and independently:

$$s_i : \mathbb{R} \to \Delta\mathbb{R}$$

- generates a distribution G over posterior expectations:

$$w_i \triangleq \mathbb{E}[v_i \mid s_i]$$

Revenue

- objective of the seller is to maximize revenue in a second-price auction
- revenue is equal to second-highest expected valuation across bidders
- k-th highest valuation is denoted by $w_{(k)}$
- objective of seller is to solve:

$$R \triangleq \max_{\{s_i:\mathbb{R}\to\Delta\mathbb{R}\}_{i\in N}} \mathbb{E}[w_{(2)}].$$

Analysis

First Steps of Analysis

- find optimal symmetric information structure
- information structure generates posterior expectation w_i with distribution G:

$$w_i \triangleq \mathbb{E}[v_i \mid s_i]$$

- Blackwell/Strassen/Rothschild-Stiglitz show: there exists a signal s that induces a distribution of expected valuations G from F if and if F is a mean preserving spread of G
- F is a mean preserving spread of G if

$$\int_v^\infty dF(t) \leq \int_v^\infty dG(t), \; \forall v \in \mathbb{R}_+$$

and

$$\int_0^\infty dF(t) = \int_0^\infty dG(t).$$

- if F is a mean preserving spread of G we write $F \prec G$

Revenue

- second-order statistic $w_{(2)}$ of N symmetrically and independently distributed random variables is

$$\mathbb{P}(w_{(2)} \leq t) = NG^{N-1}(t)(1 - G(t)) + G^N(t)$$

- expected revenue of seller:

$$R = \mathbb{E}[w_{(2)}] = \int_0^\infty t d(NG^{N-1}(t)(1 - G(t)) + G^N(t))$$

- maximization problem:

$$R = \max_G \int_0^\infty t d(NG^{N-1}(t)(1 - G(t)) + G^N(t))$$
$$\text{subject to } F \prec G.$$

- non-linear problem in optimization variable G
  neither convex nor concave program

Quantile Representation

- denote by q_i a random variable that is uniformly distributed in $[0,1]$ and

$$F^{-1}(q_i) = v_i.$$

- distribution function of quantile of second-highest valuation:

$$S_N(q) \triangleq N q^{N-1}(1-q) + q^N$$

- quantile distribution S_N is independent of the underlying distribution F or G
- just as quantile of any random variable is uniformly distributed, the quantile of second-order statistic of N random variables is distributed according to S_N for every distribution

Quantile Representation: Change of Variable

- revenue is expectation over quantiles using measure $S_N(q)$
- revenue given quantile of second-order statistic is G^{-1}:

$$\max_{G^{-1}} \int_0^1 S'_N(q) G^{-1}(q)\, dq \qquad \text{subject to } G^{-1} \prec F^{-1} \tag{R}$$

- seller can choose any distribution of expected valuations whose quantile function G^{-1} is a mean-preserving spread of quantile function F^{-1}
- $F \prec G$ if and only if $G^{-1} \prec F^{-1}$
- objective is linear in G^{-1}

Main Result

Proposition (Optimal Information Structure)

Suppose that F is absolutely continuous, then the unique optimal symmetric information structure is given by:

$$S_N(v_i) = \begin{cases} v_i, & \text{if } q_i \leq q^*; \\ \mathbb{E}[v_i \mid F(v_i) \geq q^*], & \text{if } q_i \geq q^*. \end{cases}$$

where $q^* \in [0, 1)$ is independent of F.

- reveal the valuation of all those bidders who have a valuation lower than some threshold determined by a fixed quantile q^*
- otherwise reveal no information beyond the fact that the valuation is above the threshold
- with change of variables, "upper censorship"

Competition through Information

- optimal information structures supports competition at the top of the distribution at the expense of an efficient allocation
- bundles for every bidder all valuations above the threshold $F^{-1}(q^*)$ into a single mass point
- information rent of winning bidder is depressed with corresponding gain in revenue for seller

Intuitive Proof Step 1: Integrate by Parts

- if $\overline{v} = G^{-1}(1)$ is the upper bound on expected value, by integration by parts, revenue is:

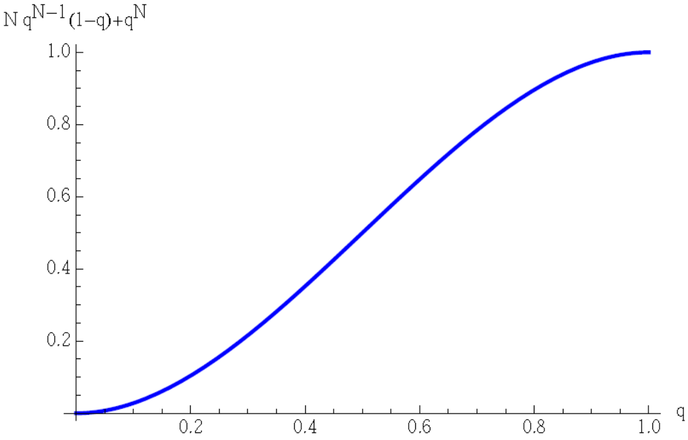
$$\int_0^1 S_N'(q)G^{-1}(q)dq = \overline{v} - \int_0^1 S_N(q)dG^{-1}(q)$$

- so we have minimization problem

$$\min_{G^{-1}} \int_0^1 S_N(q)dG^{-1}(q) \qquad \text{subject to } G^{-1} \prec F^{-1} \tag{R}$$

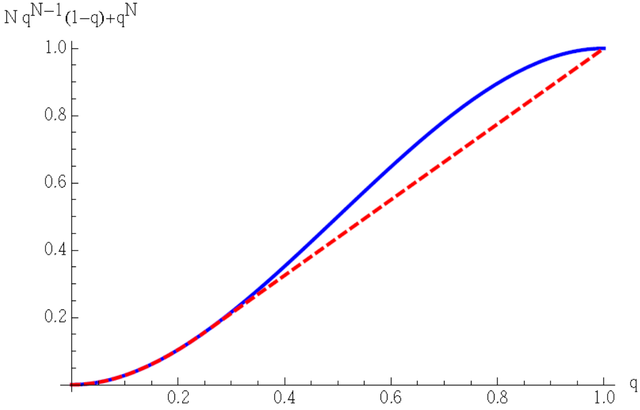
- hint: if $\overline{v} = 1$, G^{-1} is itself a distribution function.

Step 2: Convexification of Second Order Statistic



- graph of $S_N\left(q\right)$ for $N = 3$
- unique inflection point for all N

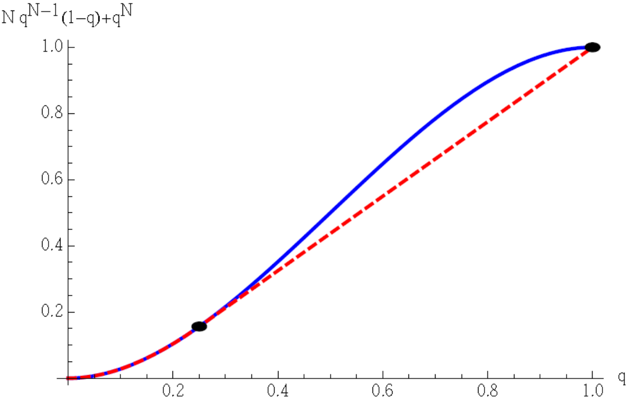
Convex Hull of Quantile Function



- find largest convex function below the original one
- problem reduces to finding q such that:

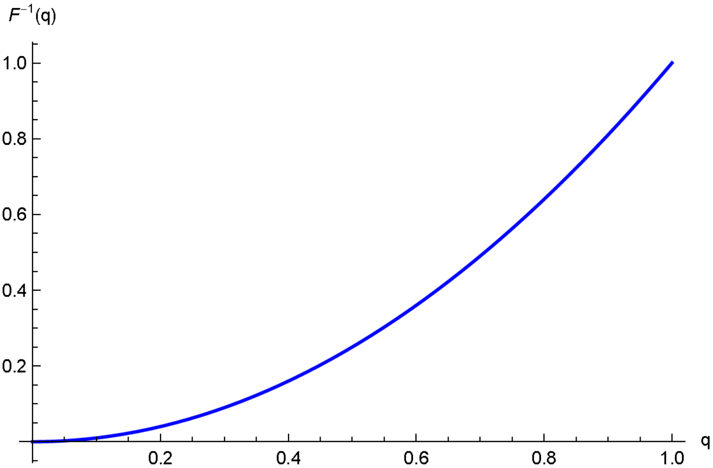
$$S_N(q) + S'_N(q)(1-q) = S_N(1) = 1$$

End Points of Affine Segment



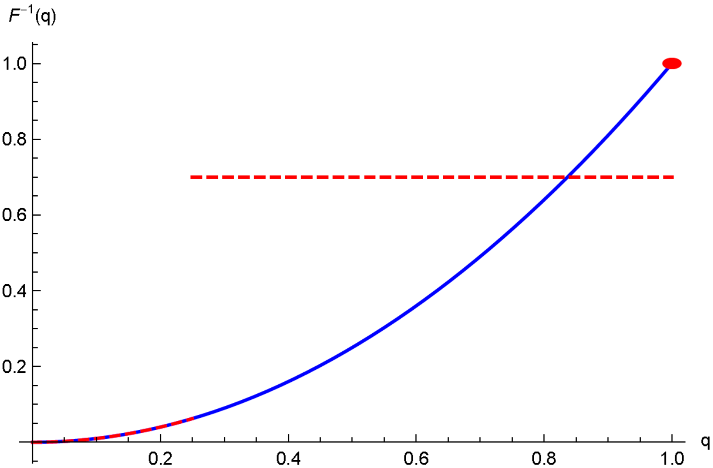
- we take the mass to the extremes of the affine segment
- the mass at each extreme must keep the expected mean of quantile constant

Step 3: Back to Value Distribution



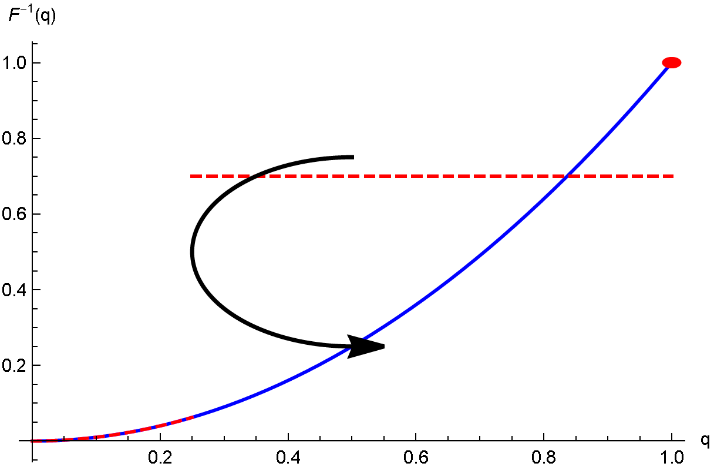
- map back to value distribution of bidder i
- we draw the quantile function for $F(v) = \sqrt{v}$

From Quantile to Convexified Quantile



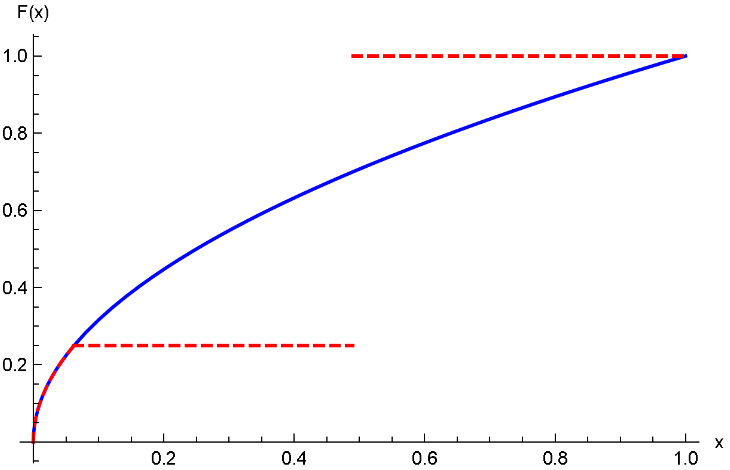
- the mass is moved to the end points
- while keeping expectation of quantile constant

From Convex Quantile to Convex Distribution



- we have been working with the quantile function
- to recover the distribution we rotate

From Convex Distribution to Information Structure



- we now have the distribution F
- there is one step in distribution of expected value

Verification

- this is an example of a problem of characterizing extreme points of monotone functions subject to majorization constraints (Kleiner et al. 2021)

Proposition (Kleiner et al. Proposition 2)

Let G^{-1} be such that for some countable collection of intervals $\{[\underline{x}_i, \bar{x}_i) \mid i \in I\}$,

$$G^{-1}(q) = \begin{cases} F^{-1}(q) & q \notin \cup_{i \in I}[\underline{x}_i, \bar{x}_i) \\ \frac{\int_{\underline{x}_i}^{\bar{x}_i} F^{-1}(t)dt}{\bar{x}_i - \underline{x}_i} & q \in [\underline{x}_i, \bar{x}_i) \end{cases}$$

If $\operatorname{conv} S_N$ is affine on $[\underline{x}_i, \bar{x}_i)$ for each $i \in I$ and if $\operatorname{conv} S_N = S_N$ otherwise, then G solves the maximization problem. Moreover, if F is strictly increasing the converse holds.

What is the Critical Quantile?

Proposition (Critical Quantile)

The quantile $q^*(N) \in [0,1)$ *that determines the optimal information structure is* 0 *if* $N = 2$, *is increasing in* N *and approaches* 1 *as* $N \to \infty$; *for* $N \geq 3$, *it is implicitly defined as the solution of:*

$$S'_N(q)(1-q) = 1 - S_N(q)$$

- this is an Nth degree polynomial in q

Critical Quantiles

N	$q^*(N)$
2	0
3	0.25
4	0.46
5	0.58
10	0.81
100	0.98

- optimal quantile is independent of the distribution and only depends on the number of bidders (optimal information design)
- optimal reserve price is independent of the number of bidders and only depends on the distribution (optimal auction design)
- expected numbers of bidder at top of the distribution

$$N\left(1-q_N^*\right) \in (1.79, 2.25)$$

Variational Intuition

- suppose we initially have quantile threshold q and write $\underline{v} = F^{-1}(q)$ and $\overline{v} = \mathbb{E}_F[v|v \geq \underline{v}]$
- suppose we lower threshold by dq :
- expected gain in approximately by bringing marginal bidder in:

$$\overbrace{dqS_N'(q)}^{\text{marginal gain}} \times \overbrace{\overline{v} - \underline{v}}^{\text{increase in payment}}$$

- expected loss is approximately by lowering price on inframarginal

$$\overbrace{1 - S_N(q)}^{\text{inframarginal loss}} \times \overbrace{\frac{\overline{v} - \underline{v}}{1 - q}}^{\text{decrease in payment}} dq$$

Reserve Price

- second-price auction with reserve price $r > 0$

Proposition (Optimal Information with Reserve r)

Given a reserve price r, an optimal distribution of expected valuations is given by:

$$G^{-1}(q) = \begin{cases} F^{-1}(q) & \textit{if } q \in [0, q_1) \cup (q_2, q_3]; \\ r & \textit{if } q \in (q_1, q_2]; \\ \bar{v} & \textit{if } q \in (q_3, 1]; \end{cases}$$

for some quantiles $q_1 \leq q_2 \leq q_3$ (inequalities are not necessarily strict) and $r < F^{-1}(v_2) < \bar{v}$.

- bundling now occurs twice:
  (i) around the reserve price; (ii) upper censorship

Market for Impressions

Market for Impressions: Qualitative Features

- private information in digital advertising takes a particular distributed form....

1. viewer is object of auction and has many attributes (demographics, past browsing behavior, past purchase behavior, etc.)
2. publisher as seller has private information about attributes of viewer
3. advertiser as bidder has private information about their preference (willingness to pay) for attributes of viewer

- value of the match or impression between advertiser and viewer is jointly determined by these different sources of private information

A Model with Two-Sided Private Information

- viewer has attributes $x \in X$ distributed according to F_x.
- advertiser i has a preference for attributes $y_i \in Y$, distributed according to F_y, identically and independently distributed across i
- an impression is a match between advertiser and viewer...
- the value v_i of a viewer is

$$v_i = u(x, y_i)$$

- there is an induced distribution F over value v_i

Statistical Assumptions

- an advertiser's preference tells them nothing about their or others' valuation of the object (without knowing the attribute)
- a publisher's knowledge of viewer attributes tells them nothing about valuations
- more specifically:

$$(x, v_1, ..., v_N) \text{ and } (y, v_1, ..., v_N)$$

are vectors of independently distributed random variables

Micro Foundation for Statistical Assumptions

- one microfoundation for statistical assumptions:
- each viewer has J (binary) attributes:

$$x_j \in \{-1, +1\}, \quad j = 1, ..., J$$

- each advertiser i has preferences for attributes:

$$y_{ij} \in \{-1, +1\}, \quad j = 1, ..., J$$

- we refer to vectors (x, y) as the *characteristics*
- an impression is a match between advertiser and viewer, *match quality* between advertiser i and viewer:

$$m_i \triangleq \frac{1}{\sqrt{J}} \sum_{j=1}^{J} x_j y_{ij}$$

Match Quality and Value

- an advertiser value v_i of a viewer is determined by a strictly increasing function u of the match quality m_i:

$$u : \mathbb{R} \rightarrow \mathbb{R}_+,$$

such that:

$$v_i \triangleq u(m_i),$$

- refer to u as *valuation function*

A Model of Auto-Bidding

1. publisher commits to signal generated conditional on advertiser's reported preference and viewer's attributes
2. publisher commits to submitting advertiser optimal bid as a function of reported preference and publisher's signal
3. Preferences and attributes are realized, signals and bids are realized and the impression is allocated to the highest bidder at the second highest price

Information Design

- publisher chooses a information structure (signal):

$$s_i : \{-1, 1\}^J \times \{-1, 1\}^J \to \Delta\mathbb{R}$$

  as a function of (reported) preferences and attributes...
- ... or equivalently of (induced) value

$$s_i : \mathbb{R} \to \Delta\mathbb{R}$$

- generates a distribution G over posterior expectations

$$w_i \triangleq \mathbb{E}[v_i \mid s_i(v_i), y_i]$$

Automated Bidding

- advertiser submits his preference
  subject to truthtelling (honesty)
- publisher commits to

1. complement advertiser's preference with attribute information
2. publisher submits bid $b_i : \{-1, 1\}^J \times \mathbb{R} \to \mathbb{R}$:

$$b_i(y_i, s_i) = w_i \triangleq \mathbb{E}[v_i \mid s_i(v_i), y_i]$$

- critical aspect of automated bidding, or auto-bidding is that publisher complement preference with attribute information and establishes subsequent bid

Eliciting Advertisers' Preferences

- examine advertisers' incentives to truthfully report their preferences
- a reporting strategy for bidder i is denoted by:

$$\widetilde{y}_i : \{-1, 1\}^J \to \Delta\{-1, 1\}^J.$$

- given reported preferences, the seller discloses to the bidder a signal $s(\widetilde{v}_i)$, where

$$\widetilde{v}_i \triangleq u\left(\frac{1}{\sqrt{J}} \sum_{j=1}^{J} \widetilde{y}_{ij}(y_{ij}) x_j\right)$$

- since preferences and attributes are symmetrically distributed, a sufficient statistic for the bidder's strategy is the fraction of preferences truthfully reported:

$$t_i \triangleq \sum_{j=1}^{J} \frac{\widetilde{y}_i y_i}{J}$$

Auto-Bidding

Proposition (Truthful Reporting)

Under the optimal information structure, it is a dominant strategy for an advertiser to report truthfully his preferences to the publisher.

- distribution of bids $\tilde{b}_i$ is the same for every reported strategy
- truthtelling generates the highest correlation among all joint distributions (v_i, b_i)

Manual Bidding

- advertiser submits his preference
  subject to truthtelling (honesty)
- publisher commits to

1. complement advertiser's preference with attribute information

- advertiser combines preference and attribute information to set advertiser-optimal bid
  subject to obedience

Manual Bidding

- truthtelling is not an equilibrium for every $N,\ u$
- there is a class of information structures balancing revenue and incentive compatibility with large N
- consider the two-sided pooling structure:

$$s(v_i) = \begin{cases} \mathbb{E}[v_j \mid F(v_j) \leq 1-q] & \text{if } F(v_j) \leq 1-q^* \\ v_j & \text{if } 1-q^* \leq F(v_j) \leq q^* \\ \mathbb{E}[v_j \mid F(v_j) \geq q] & \text{if } F(v_j) \geq q^* \end{cases}$$

- above information structure adds pooling at the bottom to pooling at the top

Truthful Reporting Under Manual Bidding

Proposition (Honesty and Obedience)

Under manual bidding, it is a dominant strategy for the advertiser to report his preference truthfully in the two-sided pooling structure.

Proposition (Approximate Optimality)

Under the two-sided pooling information structure the revenue converges to the one under the optimal information structure when the number of bidders grows large:

$$\lim_{N \to \infty} \left(\mathbb{E}[w_{(2)}] - R \right) = 0.$$

- revenue under two-sided pooling is given by $w_{(2)}$

Comment on Manual Bidding

- suppose that the advertiser chooses his bid after receiver signal from publisher
- advertiser now has the option of double deviation: misreporting preferences to control information and then bidding as a function of true preferences
- analogous to Bayesian persuasion with private information

Large Markets

Large Markets

- large number of (possible) bidders is arguably the prevailing condition in digital advertising how does information respond to random participation of bidders
- revenue performance of auction with optimal information structure when the actual number of participating bidders grows large.

Random Number of Bidder

- with probability p, valuation is equal zero
- with probability $1 - p$,valuation is distributed with F
- limit as $N \to \infty$ and $p \to 1$ while expected number of bidders with positive values constant at:

$$\lambda \triangleq N(1 - p)$$

- critical number ρ of expected bidders

$$\rho \triangleq N(1 - q^*) \tag{1}$$

- as $N \to \infty$, (1) converges in terms of ρ:

$$\rho^2 e^{-\rho} = 1 - e^{-\rho} - \rho e^{-\rho} \Leftrightarrow \rho \approx 1.793$$

Equilibrium Information

- λ is expected number of serious bidders, ρ

Proposition

As $N \to \infty$, $p \to 1$, the optimal information structure is:

1. *If $\lambda \leq \rho$, then bidders observe binary signals and expected value is either 0 or $\mathbb{E}[v_i]\,\lambda/\rho$.*
2. *If $\lambda > \rho$, bidder v_i with $F(v_i) \leq (\lambda - \rho)/\lambda$ learns value, and bidder $v_i \in [F^{-1}((\lambda - \rho)/\lambda), 1]$ is bundled.*

- bundle zero values with positive values ("broad search")
- increase number of bidders even at cost of decreasing expected valuations
- with sufficiently many bidders, we have pooling of high-valuation bidders

Large Number of Bidders with Heavy Tails

- Arnosti, Beck and Milgrom (2016) argued heavy tails distribution prevail in digital advertising.
- F has regularly varying tails with index α, if

$$\lim_{t \to \infty} \frac{1 - F(kt)}{1 - F(t)} = k^{\alpha},$$

- we assume $\alpha < 0$, a $\alpha < -1$ for finite mean
- for example Pareto distribution

Revenue Comparison with Heavy Tails

- expected revenue in second price auction with complete disclosure of information, R_c :

$$R_c \triangleq \mathbb{E}[v_{(2)}].$$

- compare revenue of optimal information structure, R with revenue of complete disclosure, R_c for large N

Proposition (Revenue Ratio with Many Bidders)

As $N \to \infty$, there exists $z \in (1, \infty)$ s.th.:

$$\lim_{N \to \infty} \frac{R}{R_c} = z.$$

Furthermore, in the limit $\alpha \to -1$, $z \to \infty$.

Revenue Gains

- gains from optimal information structure do not vanish
- when the distribution has fat tails, or $\alpha < 0$

$$\mathbb{E}[v_{(1)}] - \mathbb{E}[v_{(2)}] \to \infty, \text{ as } N \to \infty.$$

- optimal information structure thickens the market at the tail of the distribution
- thus provide a revenue improvement even as the numbers of bidders becomes arbitrarily large

Discussion and Conclusion

- auction format (revenue equivalence)
- reserve price and optimal auction
- vertical differentiation of attributes
- correlated values and adverse selection
- privacy policies (targeting negative and positive news)
- asymmetric information across bidders...

Literature

- incentive to generate information in a second price auction in a parametrized model (circle and normal distribution), (Ganuza (2004))
- optimal mechanism and information structure in an independent private value model (Bergemann and Pesendorfer (2007))
- notion of conflation in Levin and Milgrom (2010)
- automated versus manual bidding, Aggarwal et al. (2019), Deng et al. (2020),
- Hartline et al. (2019) (dashboard mechanism) concerned with indirect mechanism without thruthtelling, we are concerned with augmented/additional information