Theoretical Discussion for HT and AL Estimation

Bruce Wen

November 20, 2022

1 Haile-Tamer Extensions

Let H be the result of an empirical distribution function, e.g. $\hat{G}_{i:n}(v)$ as defined in Haile-Tamer (2003). The i^{th} order statistic is well-known to follow the cumulative distribution,

$$F_{i:n}(v) = \sum_{j=i}^{n} \frac{n!}{j!(n-j)!} F(v)^j \left[1 - F(v)\right]^{n-j}$$

From Haile Tamer (2003), define the strictly increasing differentiable function $\phi_{i:n}(H) : [0,1] \to [0,1]$ as the implicit solution to

$$H = \frac{n!}{(n-i)!(i-1)!} \int_0^{\phi} s^{i-1}(1-s)^{n-i} ds$$

We will now prove this is correct.
Proof:
We want to show that, $\sum_{j=i}^{n} \frac{n!}{j!(n-j)!} F(v)^j \left[1 - F(v)\right]^{n-j} = \frac{n!}{(n-i)!(i-1)!} \int_0^{F(v)} s^{i-1}(1-s)^{n-i} ds$. Do repeated integration by parts on the RHS,

$$\begin{aligned}
\int_0^{F(v)} s^{i-1}(1-s)^{n-i} ds &= \int_0^{F(v)} (1-s)^{n-i} s^{i-1} ds \\
&= \left[(1-s)^{n-i}\frac{s^i}{i}\right]_{s=0}^{F(v)} + \int_0^{F(v)} (n-i)(1-s)^{n-i-1}\frac{s^i}{i} ds \\
&= \left[(1-s)^{n-i}\frac{s^i}{i} + (n-i)(1-s)^{n-i-1}\frac{s^{i+1}}{i(i+1)}\right]_{s=0}^{F(v)} + \int_0^{F(v)} (n-i)(n-i-1)(1-s)^{n-i-2}\frac{s^{i+1}}{i(i+1)} ds \\
&= \ldots \\
&= \sum_{j=i}^{n} \left[\frac{(n-i)!}{(n-j)!}(1-s)^{n-j}\frac{s^j}{j!/(i-1)!}\right]_{s=0}^{F(v)} \\
&= \sum_{j=i}^{n} \frac{(n-i)!}{(n-j)!}(1-F(v))^{n-j}\frac{F(v)^j}{j!/(i-1)!}
\end{aligned}$$

Finally, $\frac{n!}{(n-i)!(i-1)!} * \sum_{j=i}^{n} \frac{(n-i)!}{(n-j)!}(1-F(v))^{n-j}\frac{F(v)^j}{j!/(i-1)!} = \sum_{j=i}^{n} \frac{n!}{j!(n-j)!} F(v)^j \left[1 - F(v)\right]^{n-j}$ which completes the proof.

1.1 Estimating the highest and 2nd-highest order statistics

First define the inverse function of ϕ as, $\phi_{i:n}^{-1}(F(v)) : [0,1] \to [0,1]$,

$$\phi_{i:n}^{-1}(F(v)) = \frac{n!}{(n-i)!(i-1)!} \int_0^{F(v)} s^{i-1}(1-s)^{n-i} ds$$

Then the bounds for the second highest order statistic, using only the top 2 bids, are,

$$G_{n:n}(b) \leq F_{n-1:n}(b) \leq \phi^{-1}_{n-1:n}\left(\min_{j\in\{n-1,n\}}\phi_{j:n}(G_{n-1:n})\right)$$

And the bounds for the highest order statistic, using only the top 2 bids, are,

$$\phi_{n-1:n}\left(G_{n:n}(b)\right)^N \leq F_{n:n}(b) \leq \phi^{-1}_{n:n}\left(\min_{j\in\{n-1,n\}}\phi_{j:n}(G_{n-1:n})\right)$$

2 Aradillas-Lopez Derivations

The profit payoff function in Aradillas-Lopez can be re-written in expectation form,

$$\begin{aligned}
\pi_n(r) &= \underbrace{\int_0^\infty max\{r,v\}dF_{n-1:n}(v)}_{\text{Expected Revenue}} - \underbrace{v_0}_{\text{Value to Auctioneer}} - \underbrace{F_{n:n}(r)(r-v_0)}_{\text{If highest valuation}<r\text{, reduce profit to 0.}} \\
&= \int_0^\infty max\{r,v\}f_{n-1:n}(v)d(v) - v_0 - F_{n:n}(r)(r-v_0) \\
&= \mathbb{E}_V\left[max\{r,v\}\right] - v_0 - F_{n:n}(r)(r-v_0) \qquad (i)
\end{aligned}$$

(i) Here we assume the discrete version of expectation, and thus do not require differentiability of the $max()$ function.

where $V \sim F_{n-1:n}$. Now, we can use the law of large numbers and sample from the distribution at different values of v, i.e. calculate,

$$\frac{1}{N}\sum_{i=1}^{N} max\{r,v_i\}$$

which converges to $\mathbb{E}_V\left[max\{r,v\}\right]$ as $N\to\infty$.
To sample v_i, we do the following. First create the random variable, $U \sim Uniform[0,1]$, which is the range of the cumulative distributive function. Then, $V \sim F^{-1}_{n-1:n}(U)$. Note that $F_{n-1:n}(v)$ is weakly increasing in v (see ii), so it is fine to take the inverse. So we draw the required v_i from V.

Claim (ii): $F_{n-1:n}(v)$ is weakly increasing in v.
Proof (ii): Let $v_1, v_2 \in [0,\bar{V}]$ such that $v_1 < v_2$. Since $G_{n:n}(v) \leq F_{n-1:n}(v) \leq G_{n-1:n}(v)$ by construction, it suffices to show that $G_{n-1:n}(v_1) \leq G_{n:n}(v_2)$. Since $G_{i:n}(v) = \frac{1}{T_n}\sum_{t=1}^{T}\mathbb{1}\{n_t = n, b_{i:n_t} \leq v\}$, and $b_{n-1:n} < b_{n:n}$ in the data for number of bidders $n \geq 2$, it must be that $G_{n-1:n}(v_1) \leq G_{n:n}(v_2)$.

The sharp bounds for $F_{n-1:n}(b)$ are,

$$G_{n:n}(b) \leq F_{n-1:n}(b) \leq G_{n-1:n}(b)$$

And the sharp bounds for $F_{n:n}(b)$ are,

$$\phi_{n-1:n}\left(G_{n:n}(b)\right)^n \leq F_{n:n}(b) \leq G_{n:n}(b)$$

So the profit function is bounded by,

$$\int_0^\infty max\{r,v\}dG_{n-1:n}(v) - v_0 - \phi_{n-1:n}\left(G_{n:n}(r)\right)^n(r-v_0) \leq \pi_n(r) \leq \int_0^\infty max\{r,v\}dG_{n:n}(v) - v_0 - G_{n:n}(r)(r-v_0)$$

which can be simplified to be just estimating, for all $0 \leq r \leq \bar{V}$ (pointwise),

$$\frac{1}{N}\sum_{i=1}^{N} max\{r, v_{lb,i}\} - v_0 - G_{n:n}(r)(r - v_0) \leq \pi_n(r) \leq \frac{1}{N}\sum_{i=1}^{N} max\{r, v_{ub,i}\} - v_0 - \phi_{n-1:n}\left(G_{n:n}(r)\right)^n (r - v_0)$$

where $v_{lb,i}$ is drawn from $V_{lb} \sim G^{-1}_{n-1:n}(U)$ and $v_{ub,i}$ is drawn from $V_{ub} \sim G^{-1}_{n:n}(U)$. We can set N to be large, say 100000.

Meanwhile, the Haile-Tamer profit bounds are,

$$\frac{1}{N}\sum_{i=1}^{N} max\{r, v_{lb,i}\} - v_0 - \phi^{-1}_{n:n}\left(\min_{j \in \{n-1,n\}} \phi_{j:n}(G_{n-1:n})\right)(r - v_0) \leq \pi^{HT}_n(r) \leq \frac{1}{N}\sum_{i=1}^{N} max\{r, v_{ub,i}\} - v_0 - \phi_{n-1:n}\left(G_{n:n}(r)\right)^n (r - v_0)$$

where $v_{lb,i}$ is drawn from $H^{-1}(U)$ where $H(U) = \phi_{n-1:n}\left(G_{n:n}(U)\right)^n$ and $v_{ub,i}$ is drawn from $V_{ub} \sim G^{-1}_{n:n}(U)$. So only the lower bounds differ between the Haile-Tamer and Aradillas-Lopez methods.