Order Statistics and Revenue Equivalence

CSCI 1440/2440

2024-01-31

We derive the first-, second-, and kth-order statistics for the uniform distribution on $[0, 1]$. We use these results to prove that the expected revenue of the first- and second-price auctions are equal.

1 *Order Statistics*

Definition 1.1. The **kth-order statistic**, denoted $X_{(k;n)}$, is the the kth-largest value among n i.i.d.draws of a random variable X.

In particular, the first-order statistic is the maximum of n draws, the second-order statistic is the second highest of n draws, and the nth-order statistic is the minimum of n draws.[1]

[1] It is equally legitimate to define order statistics in the reverse order, so that the first-order statistic is the minimum, instead of the maximum, of n draws.

Order statistics are useful in analyzing the outcomes of first- and second-price auctions, as these outcomes depend on the highest and second-highest draws from the distribution of bidders' values.

Consider a random variable X that is uniform on $[0, 1]$. If $n = 1$, then we expect the value of the first[2]-order statistic to be the expected value of X itself, namely $1/2$. If $n = 2$, then we would expect the two order statistics to divide the unit interval into thirds, in which case the expected value of the first-order statistic is $2/3$ and the expected value of the second-order statistic is $1/3$. In general, we would expect n order statistics to divide the unit interval into $n + 1$ regions, so that the expected value of the smallest order statistic is $1/n+1$ and the expected value of the largest, is $n/n+1$.

[2] and only

In the rest of this section, we formalize this intuition. We simply write $X_{(k)}$, when n is clear from context.

1.1 *First-Order Statistic*

Fix a value of n. We are interested in calculating the expected value of $X_{(1)}$, the first-order statistic, when sampling i.i.d. from a uniform distribution, call it U, on $[0, 1]$. That is,

$$\mathbb{E}\left[X_{(1)}\right] = \int_0^1 x f_{X_{(1)}}(x)\ dx.$$

We will proceed by computing the CDF $F_{X_{(1)}}$, which is easy to compute, and then taking derivatives to arrive at the PDF, $f_{X_{(1)}}$.[3]

[3] The CDF is defined as follows:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\ dt.$$

By the Fundamental Thm of Calculus,

$$f_X(x) = \frac{d}{dx} F_X(x).$$

In particular,

$$f_{X_{(1)}}(x) = \frac{d}{dx} F_{X_{(1)}}(x).$$

Observe that the CDF at some value $x \in [0, 1]$ is the probability that all n draws are less than x: i.e.,

$$F_{X_{(1)}}(x) = \Pr(X_{(1)} \leq x)$$

$$
\begin{aligned}
&= \Pr(X_j \leq x, \text{ for all } j \in [n]) \\
&= \prod_n \Pr(X \leq x) \\
&= \prod_n F(x) \\
&= x^n.
\end{aligned}
$$

Now

$$
\begin{aligned}
f_{X_{(1)}}(x) &= \frac{d}{dx} F_{X_{(1)}}(x) \\
&= \frac{d}{dx} x^n \\
&= n x^{n-1}.
\end{aligned}
$$

Therefore,

$$
\mathbb{E}\left[X_{(1)}\right] = \int_0^1 x f_{X_{(1)}}(x)\ dx = \int_0^1 n x^n\ dx = \frac{n}{n+1} x^{n+1}\Big|_0^1 = \frac{n}{n+1}.
$$

1.2 *Second-Order Statistic*

We follow the same steps to compute the second-order statistic (using fewer words).

CDF:

$$
\begin{aligned}
\Pr(X_{(2)} \leq x) &= \Pr(X_{(1)} \leq x) + \sum_{i=1}^{n} \Pr(X_j \leq x, \forall j \neq i \text{ and } X_i > x) \\
&= \Pr(X_{(1)} \leq x) + \sum_{i=1}^{n} \Pr(X_j \leq x, \forall j \neq i) \Pr(X_i > x) \\
&= x^n + n x^{n-1}(1-x).
\end{aligned}
$$

In words, all the draws are less than x, which again happens with probability x^n, *or* only $n-1$ of the draws are less than x, which can happen in n different ways, each with probability $x^{n-1}(1-x)$.

PDF:

$$
f_{X_{(2)}}(x) = n x^{n-1} + n(n-1)x^{n-2}(1-x) - n x^{n-1} = n(n-1)x^{n-2}(1-x).
$$

Expected value of the second-order statistic:

$$
\begin{aligned}
\mathbb{E}\left[X_{(2)}\right] &= \int_0^1 x f_{X_{(2)}}(x) dx \\
&= n(n-1) \int_0^1 \left(x^{n-1} - x^n\right) dx \\
&= n(n-1) \left(\frac{1}{n} - \frac{1}{n+1}\right) \\
&= n(n-1) \frac{1}{n(n+1)} \\
&= \frac{n-1}{n+1}.
\end{aligned}
$$

2 Revenue Equivalence

Theorem 2.1. *If bidder's values are uniform i.i.d., then the expected revenue of the first-price auction is equal to that of the second-price auction, assuming bidders behave according to their respective equilibrium strategies.*

Proof. The support of the uniform distribution does not matter; we choose $[0,1]$ for convenience. Let R_1 and R_2 denote the expected revenue of the first- and second-price auctions, respectively.

In a second-price auction, the bidder with the highest value wins, paying the second-highest bid, which, because the auction is truthful, is in fact the second-highest value. Therefore, the expected revenue is equal to the expected second-highest value, which is precisely the expected value of the second-order statistic: i.e.,

$$R_2 = \frac{n-1}{n+1}.$$

In a first-price auction, the winner pays as she bid. The expected revenue, therefore, is equal to the expected highest bid. The expected value of the first-order statistic, however, is the expected highest *value*, not the expected highest bid. But since the equilibrium bid function $b_i = (n-1/n)\, v_i$ is monotonically non-decreasing in value, the highest bidder has the highest value. Letting $X_{(1)}$ be a random variable denoting this value (i.e., the first-order statistic),

$$\begin{aligned} R_1 &= \mathbb{E}\left[\left(\frac{n-1}{n}\right) X_{(1)}\right] \\ &= \left(\frac{n-1}{n}\right) \mathbb{E}\left[X_{(1)}\right] \\ &= \left(\frac{n-1}{n}\right)\left(\frac{n}{n+1}\right) \\ &= \frac{n-1}{n+1}. \end{aligned}$$

Therefore, $R_1 = R_2$. □

A kth-Order Statistic

Beta Function The Beta function $\mathrm{B}(x,y)$ is by the following integral:

$$\mathrm{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1} dy.$$

When x and y are positive integers, this function simplifies as:

$$\mathrm{B}(x,y) = \frac{(x-1)!(y-1)!}{(x+y-1)!}.$$

We will use the Beta function in (the very last step of) our derivation of the expected value of the kth-order statistic.

To start, let's compute the probability the kth-order statistic lies in some small interval $[x, x + \Delta x] \subset [0, 1]$. When the draws are i.i.d.,

$$\Pr(X_{(k)} \in [x, x+\Delta x]) = n\binom{n-1}{k-1} \Pr(X < x)^{n-k} \Pr(X \in [x, x+\Delta x]) \Pr(X > x+\Delta x)^{k-1} + \mathrm{O}(\Delta x^2).$$

The middle three probabilities are, respectively, the chance of:

- exactly $n - k$ values less than x,
- exactly one value between x and $x + \Delta x$, and
- exactly $k - 1$ values greater than $x + \Delta x$.

This gives the probability of one specific arrangement of this form, so we multiply by the number of possible arrangements. There are n possible agents who could have a value between x and $x + \Delta$, after which there are $\binom{n-1}{k-1}$ possible groups of agents who could have values greater than x, after which the remaining $n - k$ agents are fixed. **N.B.** There is also a chance that multiple values fall between x and $x + \Delta x$. As each such probability will contain a Δx^i term with $i \geq 2$, we include the term $\mathrm{O}(\Delta x^2)$.

The assumption that X is uniformly distributed on $[0, 1]$ yields the following further simplification:

$$\Pr(X_{(k)} \in [x, x+\Delta x]) = n\binom{n-1}{k-1} x^{n-k} \Delta x \ (1 - x - \Delta x)^{k-1} + \mathrm{O}(\Delta x^2).$$

Letting $x_{i+1} = x_i + \Delta x$, we can express the expectation of interest in discretized space as follows:

$$\sum_{i=1}^{m} x_i \ \Pr(X_{(k)} \in [x_i, x_{i+1}]).$$

To calculate the corresponding continuous expectation, we take the limit as $m \to \infty$, so that the Δx terms become arbitrarily small:

$$\begin{aligned}
\mathbb{E}\left[X_{(k)}\right] &= \lim_{m\to\infty} \sum_{i=1}^{m} x_i \ \Pr(X_{(k)} \in [x_i, x_i + \Delta x]) \\
&= n\binom{n-1}{k-1} \left(\lim_{m\to\infty} \sum_{i=1}^{m} x_i^{n-k+1} \ \Delta x \ (1 - x_i - \Delta x)^{k-1} + \mathrm{O}(\Delta x^2) \right) \\
&= n\binom{n-1}{k-1} \int_0^1 x^{n-k+1} (1-x)^{k-1} dx \\
&= n\binom{n-1}{k-1} \mathrm{B}(n-k+2, k) \\
&= \left(\frac{n!}{(k-1)!(n-k)!} \right) \left(\frac{(n-k+1)!(k-1)!}{(n+1)!} \right) \\
&= \frac{n-(k-1)}{n+1}.
\end{aligned}$$