Concavity and Convexity of Order Statistics in Sample Size

Mitchell Watt*

November 7, 2021

Abstract

We show that the expectation of the k^{th}-order statistic of an i.i.d. sample of size n from a monotone reverse hazard rate (MRHR) distribution is convex in n and that the expectation of the $(n-k+1)^{\text{th}}$-order statistic from a monotone hazard rate (MHR) distribution is concave in n for $n \geq k$. We apply this result to the analysis of independent private value auctions in which the auctioneer faces a convex cost of attracting bidders.

1 Preliminaries

Let $X_1, X_2, ..., X_n$ be a sample of size n drawn identically and independently from a continuous distribution with cumulative distribution function F with support in $\mathbb{R}$ and let f be its associated probability density function. Let $X_{(1;n)} \leq X_{(2;n)} \leq ... \leq X_{(n-1;n)} \leq X_{(n;n)}$ be the order statistics (obtained by sorting the sample from smallest to largest) and denote by $\mu_{(k;n)}$ the expected k^{th} order statistic of a sample of size n. Throughout, we assume that F is such that the expectations $\mu_{(k;n)}$ exist (for finite $1 \leq k \leq n$).

The *hazard rate* (or failure rate) of distribution F is defined by

$$h(x) = \frac{f(x)}{1 - F(x)}.$$

*Department of Economics, Stanford University. Email: mwatt@stanford.edu. I gratefully acknowledge the support of the Stanford Graduate Fellowship in Science and Engineering and the Ric Weiland Graduate Fellowship in the Humanities and Sciences.

The *reversed hazard rate* of distribution F is defined by

$$h'(x) = \frac{f(x)}{F(x)}.$$

A distribution has *monotone hazard rate (MHR)* (or increasing failure rate) if the function $h(x)$ is nondecreasing in x. A distribution has *monotone reverse hazard rate (MRHR)* (decreasing reverse failure rate) if $h'(x)$ is nonincreasing in x. A distribution is *log-concave* if $\log(f(x))$ is a concave function of x.

It is well-known (for example, by [3]) that log-concave distributions have MHR and MRHR (but the reverse implication does not hold). Log-concave distributions include the normal distribution, the uniform distributions over a convex set, the extreme value distribution, the gamma distribution (for certain shape parameters), the Weibull distribution (for certain shape parameters) and the exponential distribution (which is the unique family of distributions with a *constant* hazard rate).

2 Main result

We consider the expected k^{th} order statistic and the expected $(n-k+1)^{\text{th}}$ order statistic for fixed k as a function of the sample size $n \geq k$. Informally, we refer to the former as the k^{th} "bottom" order statistic and the latter as the k^{th} "top" order statistic. It is well-known that for *any* distribution function F, the expected minimum observation of the sample, $\mu_{(1;n)}$, is a nonincreasing, convex function of the sample size and the expected maximum observation, $\mu_{(n;n)}$, is a nondecreasing, concave function of the sample size (for completeness, we include the proof of this result below). In Theorem 1, we extend this characterization to the top and bottom order statistics under assumptions on the distribution's hazard rate.

Theorem 1. *Let F be a distribution and consider any fixed $k \geq 1$:*

(a) If F has MRHR, the expected k^{th} order statistic, $\mu_{(k;n)}$, is a nonincreasing, convex function of the sample size n for $n \geq k$.

(b) If F has MHR, the expected $(n-k+1)^{\text{th}}$ order statistic, $\mu_{(n-k+1;n)}$, is a nondecreasing, concave function of the sample size n for $n \geq k$.

In particular, if f is log-concave, the conclusions of both (a) and (b) above hold.

Proof of (a). The bottom order statistics satisfy the identity due to [2]

$$\Delta_{(k;n)} := \mu_{(k;n)} - \mu_{(k;n+1)} = \binom{n}{k-1} \int_{-\infty}^{\infty} F^k(x)(1-F(x))^{n-k+1} \mathrm{d}x \qquad (1)$$

for $n \geq k$. Because the integrand on the right is nonnegative for all x, we see immediately that $\mu_{(k;n)} \geq \mu_{(k;n+1)}$, so that $\mu_{(k;n)}$ is nonincreasing in n. (Note that this conclusion does not depend on F being MRHR.)

To show convexity, we will show that $\Delta_{(k;n+1)} \leq \Delta_{(k;n)}$. By subtracting eq. (1) for n and $n+1$, we obtain

$$\Delta_{(k;n)} - \Delta_{(k;n+1)} = \binom{n}{k-1} \int_{-\infty}^{\infty} F^k(x)(1-F(x))^{n-k+1} \left[1 - \frac{n+1}{n-k+2}(1-F(x))\right] \mathrm{d}x. \tag{2}$$

Define x^* by $F(x^*) = \frac{k-1}{n+1}$. Note that the integrand in eq. (2) is negative for $x < x^*$, zero at $x = x^*$ and positive for $x > x^*$.

To sign the integral in eq. (2), we use an approach similar to [4], which exploits a lemma due to [1], stated below for reference.

Lemma 1. *Let W be a measure on interval (a, b) and g a nonnegative function defined on the same interval.*

(a) If $\int_t^b \mathrm{d}W(x) \geq 0$ for all $t \in (a, b)$ and if g is nondecreasing, then $\int_a^b g(x)\mathrm{d}W(x) \geq 0$.

(b) If $\int_a^t \mathrm{d}W(x) \leq 0$ for all $t \in (a, b)$, and if g is nonincreasing, then $\int_a^b g(x)\mathrm{d}W(x) \leq 0$.

Consider the integral

$$J(t) := \int_t^{\infty} F^{k-1}(x)(1-F(x))^{n-k+1} \left[1 - \frac{n+1}{n-k+2}(1-F(x))\right] f(x)\mathrm{d}x.$$

Note that the integrand of J differs from that of eq. (2) by a multiplicative factor of the reverse hazard rate, $h'(x)$. Note that the integrand of $J(t)$ is negative for $x < x^*$, zero at $x = x^*$ and positive for $x > x^*$. Thus, $J(t)$ must be nonnegative for $t \geq x^*$, and for $t < x^*$, we have that $J(t) \geq J(-\infty)$. So to show $J(t) \geq 0$ for all t, it suffices to show that $J(-\infty) \geq 0$.

We make the change of variable $u = F(x)$ and rewrite the integral $J(-\infty)$ as

$$\int_0^1 u^{k-1}(1-u)^{n-k+1} \left[1 - \frac{n+1}{n-k+2}(1-u)\right] \mathrm{d}u. \tag{3}$$

We use computer algebra software[1] to compute this integral and obtain

$$J(-\infty) = \frac{(k-1)!(n-k+1)!}{(n+2)!}.$$

[1] We used Wolfram Alpha for this calculation.

This is clearly nonnegative so that $J(t) \geq 0$ for all t. So by Lemma 1(a), multiplying the integrand in $J(t)$ through by the inverse reverse hazard rate $1/h'(x)$ (which is nondecreasing by assumption that F is MRHR), we obtain that the integral in eq. (2) is nonnegative as well. This implies $\Delta_{(k;n)} \geq \Delta_{(k;n+1)}$, so that $\mu_{(k;n)}$ is convex in n. □

Remark. Note that for $k = 1$, the integral in eq. (2) is $\binom{n}{k-1} \int_{-\infty}^{\infty} F^{k+1}(x)(1 - F(x))^{n-k+1} \mathrm{d}x$, which is always nonnegative. This shows that the minimal order statistic is always convex in n, without any assumptions on the distribution F.

Proof of (b). We apply another identity due to [2],

$$\mu_{(r;n)} - \mu_{(r-1;n-1)} = \binom{n-1}{r-1} \int_{-\infty}^{\infty} F^{r-1}(x)[1 - F(x)]^{n-r+1} \, \mathrm{d}y.$$

Letting $r = n - k + 1$, we obtain

$$\delta_{(k;n)} := \mu_{(n-k+1;n)} - \mu_{(n-k;n-1)} = \binom{n-1}{n-k} \int_{-\infty}^{\infty} F^{n-k}(x)[1 - F(x)]^{k} \, \mathrm{d}y. \quad (4)$$

The integrand is nonnegative, so that $\mu_{(n-k+1;n)} \geq \mu_{(n-k;n-1)}$, that is, the k^{th} top order statistic is nondecreasing as a function of n. (Note that this conclusion does not depend on F being MHR.)

To show concavity, we will show that $\delta_{(k;n)} \geq \delta_{(k;n+1)}$. By subtracting eq. (4) for n and $n + 1$, we obtain

$$\delta_{(k;n+1)} - \delta_{(k;n)} = \binom{n-1}{n-k} \int_{-\infty}^{\infty} F^{n-k}(x)(1 - F(x))^{k} \left[\frac{n}{n-k+1} F(x) - 1 \right] \mathrm{d}x. \quad (5)$$

Define $x^{\dagger}$ by $F(x^{\dagger}) = \frac{n-k+1}{n}$. The integrand in eq. (5) is negative for $x < x^{\dagger}$, zero at $x = x^{\dagger}$ and positive for $x > x^{\dagger}$. We will show that the integral in eq. (5) is nonpositive.

To do so, consider the integral

$$K(t) = \int_{-\infty}^{t} F^{n-k}(x)(1 - F(x))^{k-1} \left[\frac{n}{n-k+1} F(x) - 1 \right] f(x) \mathrm{d}x,$$

which differs from that of eq. (5). The integrand of $K(t)$ is negative for $x < x^{\dagger}$, zero at $x = x^{\dagger}$ and positive for $x > x^{\dagger}$. Thus $K(t)$ is nonpositive if $t \leq x^{\dagger}$ and $K(t) \leq K(\infty)$ for $t > x^{\dagger}$. Thus, if we show that $K(\infty) \leq 0$, we will have that $K(t) \leq 0$ for all t.

We make the change of variable $u = F(x)$, to rewrite $K(\infty)$ as

$$\int_0^1 u^{n-k}(1-u)^{k-1}\left[\frac{n}{n-k+1}u - 1\right] \mathrm{d}u.$$

We use computer algebra software[2] to compute this integral and obtain

$$K(\infty) = -\frac{(n-k)!(k-1)!}{(n+2)!}.$$

Since the above integral is nonpositive, $K(t) \leq 0$ for all t. So by Lemma 1(b), since $1/h(x)$ is nonincreasing for MHR distributions, we have that the integral in eq. (5) is nonpositive as well. Thus, $\delta_{(k;n+1)} \leq \delta_{(k;n)}$ so that $\mu_{(n-k+1;n)}$ is concave in n.

$\square$

Remark. Note that for $k = 1$, the integral in eq. (5) is $-\binom{n-1}{n-k}\int_{-\infty}^{\infty} F^{n-k}(x)(1 - F(x))^{k+1}\mathrm{d}x$, which is always nonpositive. This shows that the maximal order statistic is always concave in n, without assumptions on the distribution F.

We end this section by noting that in the absence of assumptions on the distribution F that the conclusions of Theorem 1 may not hold. In particular, we consider samples from the Pareto(a, v) distribution for $a, v > 0$, which has probability density function $f(x) = va^v x^{-v-1}\mathbb{1}_{[x \geq a]}$ and cumulative distribution function $F(x) = 1 - (a/x)^v$. This distribution does not have MHR as its hazard rate is a decreasing function. The expected order statistics are calculated in [5] as

$$\mu_{(k;n)} = \frac{n!\Gamma(n-k-1/v+1)}{(n-k)!\Gamma(n-1/v+1)},$$

where $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}\mathrm{d}x$. For the specific example of $a = 1, v = 3/4$, the expected $(n-1)^{\text{th}}$ order statistics are 3, 5.4, 8.1 for $n = 2, 3, 4$ respectively, which violates concavity (in fact, this order statistic is a convex function of n generally). By considering the transformation $X \mapsto -X$ of the same Pareto random variable, we also obtain non-convex 2^{nd} order statistics.

[2] We use Wolfram Alpha.

3 Application

One well-known application of the theory of order statistics is to auction theory. In the independent private values (IPV) model of a standard auction, each potential buyer i's valuation for an item, v_i, is drawn identically and independently from a common knowledge distribution F. The realized valuation is private knowledge to buyer i. By the payoff and revenue equivalence theorem (see, for example, [6]), in any auction with N bidders where the bidder with the highest bid wins (in the Bayes-Nash equilibrium, this will be the buyer for whom $v_i = v_{(N;N)}$), the expected payoff to the auctioneer is $\mu_{(N-1;N)}$.

Suppose that the auctioneer faces a cost associated with attracting bidders to the auction. In particular, suppose that attracting N bids to the auction results in a cost $c(N)$ for the auctioneer, where c is a positive, convex function of N.[3] The cost might capture outlays associated with advertising the auction, organizing the auction or screening potential bidders. Theorem 1 allows us to conclude the following about the number of bidders in the resulting auction.

Proposition 1. *Consider an IPV auction with a convex cost of attracting bidders. If the value distribution F has MHR and compact support, the auctioneer's objective function is concave and has a finite maximizer N^*.*

Proof. The auctioneer's payoff as a function of N is

$$g(N) = \begin{cases} \mu_{(N-1;N)} - c(N) & \text{if } N \geq 2 \\ 0 & \text{otherwise} \end{cases} .$$

Since $\mu_{(N-1;N)}$ is concave by Theorem 1, we have that $g(N)$ is concave. Since $\mu_{(N-1;N)}$ is bounded above by assumption, and $\lim_{N\to\infty} c(N) = \infty$, we have that $\lim_{N\to\infty} g(N) = -\infty$. Since $g(0) = 0$, $g(N)$ must have a finite maximizer N^*. □

Remark. MHR of the value distribution is a common assumption in auction theory as it implies Myerson's "regularity" (see [7]) which results in the optimal mechanism for the seller taking the (simple) form of a posted-price mechanism.

[3] Convexity is a common assumption of cost functions, reflecting (in this case) that it is increasingly difficult to attract marginal participants to the auction.

Concavity of the auctioneer's objective may be desirable in this setting as it allows the auctioneer to safely attract bidders sequentially. That is, the net benefit of attracting some bidder never depends on the ability to attract further bidders, or equivalently, the auctioneer never regrets attracting a bidder because they were not able to attract further bidders.

References

[1] R.E. Barlow and F. Proschan, *Statistical theory of reliability and life testing: Probability models*, Holt, Rinehart and Winston, New York, 1975.

[2] HA David, *Augmented order statistics and the biasing effect of outliers*, Statistics & Probability Letters **36** (1997), no. 2, 199–204.

[3] Ramesh C Gupta and N Balakrishnan, *Log-concavity and monotonicity of hazard and reversed hazard functions of univariate and multivariate skew-normal distributions*, Metrika **75** (2012), no. 2, 181–191.

[4] Xiaohu Li, *A note on expected rent in auction theory*, Operations Research Letters **33** (2005), no. 5, 531–534.

[5] Henrick John Malik, *Exact moments of order statistics from the pareto distribution*, Scandinavian Actuarial Journal **1966** (1966), no. 3-4, 144–157.

[6] Paul Milgrom, *Putting auction theory to work*, Cambridge University Press, 2004.

[7] Roger B Myerson, *Optimal auction design*, Mathematics of operations research **6** (1981), no. 1, 58–73.