Decreasing Returns to Sampling Without Replacement

David P. Myatt
London Business School
www.david-p-myatt.org
dmyatt@london.edu

David Ronayne
ESMT Berlin
www.davidronayne.com
david.ronayne@esmt.org

November 2025.[1]

Abstract. We study sampling from a finite population without replacement when seeking an extreme (lowest or highest) value. An example is a buyer searching for the lowest price. It is well known that there are decreasing returns to sampling from continuous populations: the expected minimum is a decreasing and discretely convex function of the sample size. We show that is true for sampling without replacement from a finite population. We also give a simple sufficient condition on population values for the properties to hold for other order statistics.

Order statistics are relevant to many calculations. A classic case from economics is a buyer searching for the best deal by randomly sampling competing price offers: such a buyer pays the lowest price (that is, the first-order or lowest-order statistic) from a sample of quotations.[2] The expected price paid (the expectation of that lowest-order statistic) clearly falls as the sample size increases. A natural hypothesis is that there are decreasing returns to sampling, so that the expected lowest-order statistic is (discretely) convex in the sample size. This is a known result whenever the sampled population distribution is continuous (so that replacement is immaterial).[3] In fact, a recent and elegant paper by Watt (2025) shows that this is also true for other lower-order statistics when the reverse hazard rate of the (continuous) population distribution satisfies an appropriate monotonicity condition.

Naturally, sampling in some settings is made without replacement from a finite population. For example, in our recent work (Myatt and Ronayne, 2025) we study oligopolistic pricing and search when firms set non-randomized prices for buyers who seek distinct quotations from them, which we treat as sampling without replacement.[4] Our results here show that there are decreasing returns to searching for the best deal: the expected lowest-order statistic is a discretely convex function of the sample size.

A second relevant setting (considered by Watt (2025) when bidders are drawn from a continuous population) is an auctioneer's decision of how many bidders to invite. Under private valuations, an

[1]Ronayne is grateful for financial support from Deutsche Forschungsgemeinschaft via CRC TRR 190 (project 280092119).
[2]Foundational work includes that by Burdett and Judd (1983) and Stigler (1961). Empirical work, such as that by Baye, Morgan, and Scholten (2004), has also relied on the properties of expected order statistics from distributions of prices.
[3]In economic applications the distribution from which draws are made is usually assumed to be continuous, as in auction theory (Krishna, 2013, Appendix C). Similarly, in the canonical model of pricing with search (Burdett and Judd, 1983) a continuum of firms' equilibrium price choices form a continuous distribution, which buyers draw from.
[4]Our solution concept returns distinct prices that are undercut-proof and such that no firm wishes to raise price given that competitors can cut prices in response. The pricing games used in the earlier literature instead involve mixed-strategy solutions which (from a buyer-search perspective) are equivalent to sampling from a continuous population.

Anglo-Japanese auction efficiently generates surplus equal to the highest valuation. Our main results imply that the expectation of this surplus is discretely concave (as well as increasing) in the number of bidders invited. Further, our supplementary results (which apply to other order statistics) show that the expected revenue (equal to the second-highest valuation) is increasing and (subject to a condition on the pattern of valuations) discretely concave in the number of bidders.[5] When that is so, the auctioneer's problem (with any cost that is weakly convex in the number of bidders) is nicely behaved.

There is comprehensive work on order statistics from sampling from finite populations in the statistics literature: see, for example, Nagaraja (1992, Section 7) and Wilks (1962, Section 8.8). However, the properties of the expected values of those statistics appear not to be reported.

The core contribution of this note is to confirm that the result known for continuous populations also holds when random draws are made without replacement from a finite population: the expected lowest (highest) order statistic is decreasing (increasing) and discretely convex (concave) in the sample size. This holds without restrictions on the population values. Pushing further to study other order statistics (the kth lowest or kth highest realization) we identify a simple condition on population values, inspired by Watt (2025), such that the discrete-convexity result holds in our sampling-without-replacement environment. We do not attain an analytic result there for all population sizes n, but do provide a sufficient criterion and numerically verify convexity for all $n \leq 100$ (with code to check for any n).

1. Extreme Order Statistics

Core Result. Consider a finite population of $n \geq 3$ values indexed such that $x_1 \leq x_2 \leq \cdots \leq x_n$. The population is sampled uniformly and without replacement. Let a sample of $q \in \{1 \dots, n\}$ values be drawn, the random variable equal to the lowest population rank in the sample be $R_q \in \{1, \dots, n\}$, its corresponding value be X_q, and the expected value of the lowest-order statistic be $\mu_q \equiv \mathrm{E}[X_q]$.[6]

Proposition 1 (Monotonicity and Convexity of the Sample Minimum). *The expected sample minimum is at least weakly decreasing in the sample size. That is, for $q \in \{1, \dots, n-1\}$,*

$$\mu_q \geq \mu_{q+1}. \tag{1}$$

If the lowest population value is unique ($x_1 < x_2$), then μ_q is strictly decreasing, i.e., eq. (1) is strict.

The expected sample minimum is discretely convex in the sample size. That is, for $q \in \{1, \dots, n-2\}$,

$$\mu_q - \mu_{q+1} \geq \mu_{q+1} - \mu_{q+2}. \tag{2}$$

If the second and third lowest population values are distinct ($x_2 < x_3$), then μ_q is strictly discretely convex so that eq. (2) holds with a strict inequality.

[5] Such results also apply to a uniform-price auction for k objects when the price equals the $(k+1)$st highest valuation.

[6] If there are ties in the population values then (without loss of generality) we maintain distinct population ranks for them.

Illustration. Consider $n = 3$ and $x_1 < x_2 < x_3$. With $q = 1$, each value is equally likely, so that $\mu_1 = (x_1 + x_2 + x_3)/3$. If $q = 2$, then the three possible (and equally likely) samples are $\{x_1, x_2\}$, $\{x_1, x_3\}$, and $\{x_2, x_3\}$. In the first two cases the minimum is x_1 whereas in the third case it is x_2 so $\mu_2 = (2x_1 + x_2)/3$. Finally, if $q = 3$, then the lowest value is always found and so $\mu_3 = x_1$. We have

$$\mu_1 - \mu_2 = \frac{x_3 - x_1}{3} \quad \text{and} \quad \mu_2 - \mu_3 = \frac{x_2 - x_1}{3}. \tag{3}$$

By inspection, the expected minimum strictly falls as the sample size increases and there are (strictly) decreasing returns (that is, $\mu_1 - \mu_2 > \mu_2 - \mu_3$) to such an increase.

Proof of Proposition 1. Defining $x_0 = 0$, the expectation of the sample minimum satisfies

$$\mu_q = \sum_{j=1}^{n} x_j \Pr[R_q = j] = \sum_{j=1}^{n} (x_j - x_{j-1}) \Pr[R_q \geq j]. \tag{4}$$

$\Pr[R_q \geq j]$ is the probability that the entire sample belongs to the $n - j + 1$ highest ranks. It satisfies

$$\Pr[R_q \geq j] = \binom{n - j + 1}{q} \Big/ \binom{n}{q} = \Pr[R_{q-1} \geq j] \times \frac{n - j - q + 2}{n - q + 1}, \tag{5}$$

where the second equality applies for $q \geq 2$ and follows from considering the factorial terms in the component binomial terms; this generates a convenient recursive form for the probabilities. We note (of course) that $\Pr[R_q \geq j] = 0$ for $j > n - q + 1$ because a sample (without replacement) of size q must include at least one observation from outside the highest $q - 1$ population values.

Differencing with respect to q, and for sample sizes satisfying $q \in \{1, \ldots, n - 1\}$,

$$\Delta\mu_q \equiv \mu_{q+1} - \mu_q = \sum_{j=1}^{n} (x_j - x_{j-1}) \left(\Pr[R_{q+1} \geq j] - \Pr[R_q \geq j] \right) \tag{6}$$

$$= -\sum_{j=1}^{n} (x_j - x_{j-1}) \Pr[R_q \geq j] \left(\frac{j - 1}{n - q} \right), \tag{7}$$

following from eq. (5). Each summand is non-negative, which gives the first result of the proposition. If $x_1 < x_2$, then the summand for $j = 2$ is strictly positive and the proposition's second claim follows.

We now turn to matters of discrete convexity. For $q \in \{1, \ldots, n - 2\}$,

$$\Delta\mu_{q+1} \equiv \mu_{q+2} - \mu_{q+1} = -\sum_{j=1}^{n} (x_j - x_{j-1}) \Pr[R_{q+1} \geq j] \left(\frac{j - 1}{n - q - 1} \right) \tag{8}$$

$$= -\sum_{j=1}^{n} (x_j - x_{j-1}) \Pr[R_q \geq j] \left(\frac{n - j - q + 1}{n - q} \right) \left(\frac{j - 1}{n - q - 1} \right). \tag{9}$$

Differencing again with respect to q, where once again $q \in \{1, \ldots, n-2\}$,

$$\Delta^2 \mu_q \equiv \Delta(\Delta\mu_q) = \Delta\mu_{q+1} - \Delta\mu_q = \sum_{j=2}^{n} (x_j - x_{j-1}) \Pr[R_q \geq j] \left(\frac{j-1}{n-q}\right)\left(\frac{j-2}{n-q-1}\right). \quad (10)$$

Each summand is non-negative, which gives the third claim of the proposition. If $x_2 < x_3$ then the summand for $j = 3$ is strictly positive and therefore $\Delta^2\mu_q > 0$, which is our final claim. □

Ties for the Population Minimum. We now consider ties for the lowest value in the population. Let the lowest $t \in \{2, \ldots, n-1\}$ population values be tied, so that $x_1 = \cdots = x_t < x_{t+1} \leq \cdots \leq x_n$. If the sample size satisfies $q > n - t$ then that sample is guaranteed to find the lowest population value, and so any further increase in the sample size has no effect. For smaller sample sizes, inspection of the relevant terms in proof of Proposition 1 provide the implications, which we state below.

Proposition 2 (The Effect of Tied Lowest Values). *Suppose that there are $t > 1$ tied minimum population values so that $x_1 = \cdots = x_t < x_{t+1} \leq \cdots \leq x_n$.*

For smaller sample sizes $q \in \{1, \ldots, n-t\}$ equations (1) and (2) hold strictly so that μ_q is strictly decreasing and strictly discretely convex over such q.

For larger sample sizes $q \in \{n-t+1, \ldots, n\}$ the lowest value is drawn with certainty so that $\mu_q = x_1$.

The Highest Order Statistic. Proposition 1 also tells us that the expected value of the sample maximum is (weakly) increasing in the sample size, and strictly so if the highest population value is unique. Similarly, the expected sample maximum is discretely concave in the sample size, and strictly so if the second and third highest population values are distinct. A result analogous to Proposition 2 also follows when there are tied maximum population values so that the statistic is strictly increasing and strictly discretely concave for smaller sample sizes $q \in \{1, \ldots, n-t\}$ and flat thereafter.

2. Other Order Statistics

We now turn to the expectation of the kth lowest ranked member of a sample with $k \in \{1, \ldots, n\}$. Consider samples with a size $q \in \{k, \ldots, n\}$ and (re)define R_q to be the population rank of the kth lowest observation drawn, where X_q and μ_q are its value and expected value, respectively.

An initial result is the monotonic effect of an increase in the sample size on the order statistics.

Proposition 3 (Monotonicity of Lower-Order Statistics). *The distribution of kth lowest order statistic X_q undergoes at least a weak first-order stochastic shift downwards following an increase in the sample size q, and so its expectation μ_q is a weakly decreasing function of q. If the population values are each unique so that $x_1 < x_2 < \cdots < x_n$, then the claims hold strictly.*

Proof. We couple the random variables R_q and R_{q+1} on the same probability space: take a random permutation of the n population ranks, and then obtain R_q and R_{q+1} from the first q and $q+1$ members of that permutation. The addition of one more rank cannot raise the kth lowest rank, and so $R_{q+1} \leq R_q$. We conclude that R_q first-order stochastically dominates R_{q+1} and that μ_q is non-increasing in q. If the population values are all unique (no ties) then there is positive probability that the added rank results in $X_{q+1} < X_q$. This implies a strict first-order domination and that μ_q strictly decreases in q. □

Proposition 3 also tells us that the distribution of the kth highest order statistic is stochastically ordered so that its expectation is increasing in q, and strictly so when population values are distinct.

We now consider the (discrete) convexity of μ_q. $\Pr[R_q = j]$ takes a hypergeometric form:[7]

$$\Pr[R_q = j] = \binom{j-1}{k-1}\binom{n-j}{q-k}\Big/\binom{n}{q} = \frac{k}{j}\left[\binom{j}{k}\binom{n-j}{q-k}\Big/\binom{n}{q}\right], \tag{11}$$

which applies to $j \in \{k, \ldots, n-q+k\}$. This probability is zero if $j < k$ (k draws within the sample rank weakly below j) or if $j > n-q+k$ ($q-k$ draws rank strictly above j).

Watt (2025) provided a condition for (discrete) convexity when samples are made from a continuous distribution $F(\cdot)$ with density $f(\cdot)$. His sufficient condition is a monotonic reverse hazard rate (MRHR) so that $f(x)/F(x)$ is decreasing (Watt, 2025, Theorem 1).

A sufficient condition for MRHR is that the density $f(x)$ is decreasing, so that $F(x)$ is concave. This, in turn, is equivalent to convexity of the inverse $F^{-1}(z)$. For a finite population, the analog of this is that the sequence of population values is discretely convex, so that $x_i - x_{i-1}$ is increasing in i.

Definition. *The population values have* increasing spacings *if they form a discretely convex function of the index i. This holds if $y_i \equiv x_{i+1} - x_i$ is increasing in i, so that $y_{i+1} \geq y_i$ for $i \in \{1, \ldots, n-2\}$.*

This condition allows us to extend our earlier result to non-extreme order statistics. We begin by defining $S_{nq} \equiv 0$ and $S_{jq} \equiv \sum_{\ell=j+1}^{n} \Pr[R_q \geq \ell]$ for $j \in \{1, \ldots, n-1\}$. This also satisfies

$$S_{jq} \equiv \sum_{\ell=j+1}^{n} \Pr[R_q \geq \ell] = \sum_{\ell=j+1}^{n} (\ell - j) \Pr[R_q = \ell] = \mathrm{E}\left[\max\left\{(R_q - j), 0\right\}\right]. \tag{12}$$

This expression is useful because (as we show below) we can, when population values have increasing spacings, write the expected kth order statistic as a weighted sum of such terms which are (at least for many parameter values) discretely convex in q. To see why (at least heuristically) we note that it is the expectation of the hinge function $\max\{(R_q - j), 0\}$. An increase in the sample size pushes the distribution of R_q downward, from where further increases in q are less effective.

[7]Of course $\Pr[R_q = j] = 0$ if either $j < k$ or if $n - j < q - k$, given that either inequality implies that the lowest k members of the sample must include population values ranked strictly above j.

Proposition 4 (Convexity of Lower-Order Statistics). *If the population values have increasing spacings, so that they form a discretely convex sequence, and if S_{jq} is discretely convex in q, then the expectation of the kth lowest-ranked observation is discretely convex in the sample size q.*

Proof. Developing further the expression from eq. (4) in the proof of Proposition 1,

$$\mu_q = \sum_{j=k}^{n} x_j \Pr[R_q = j] \quad \text{(where we begin at } k \text{ because } \Pr[R_q = j] = 0 \text{ for } j < k) \tag{13}$$

$$= x_k + \sum_{j=k+1}^{n} y_{j-1} \Pr[R_q \geq j] \quad \text{(recalling that we defined } y_j \equiv x_{j+1} - x_j) \tag{14}$$

$$= x_k + \sum_{j=k+1}^{n} y_{j-1}(S_{(j-1)q} - S_{jq}) \quad \left(\text{recalling that we defined } S_{jq} \equiv \sum_{\ell=j+1}^{n} \Pr[R_q \geq \ell]\right) \tag{15}$$

$$= x_k + y_k S_{kq} + \sum_{j=k+1}^{n-1} (y_j - y_{j-1}) S_{jq} \quad \text{(where we eliminated } y_{n-1}S_{nq} \text{ because } S_{nq} = 0) \tag{16}$$

If population values have increasing spacings then $y_j - y_{j-1} \geq 0$ for all j and so μ_q is (given the discrete convexity of each S_{jq}) the sum of discretely convex functions and is discretely convex. $\square$

We now evaluate the discrete convexity of S_{jq}, which corresponds to $S_{jq} - S_{j(q+1)} \geq S_{j(q+1)} - S_{j(q+2)}$, or equivalently $S_{jq} - 2S_{j(q+1)} + S_{j(q+2)} \geq 0$, which is explicitly

$$\sum_{\ell=j+1}^{n} (\ell - j) \left(\Pr[R_q = \ell] - 2\Pr[R_{q+1} = \ell] + \Pr[R_{q+2} = \ell]\right) \geq 0. \tag{17}$$

We can evaluate this criterion numerically for any relevant n, k, j, and q. Numerical calculations show that this condition is satisfied for all relevant parameters for $n \leq 100$.

Corollary. *If the population values have increasing spacings, and $n \leq 100$, then (from numerical verification) eq.* (17) *is satisfied so that the expected kth lowest observation is discretely convex in q.*

In our appendix we provide Python code which checks for the discrete-convexity of S_{jq} for any n.

AI use declaration

We used ChatGPT 5.1 for the initial conversion of MATLAB code to a Python script, which we then checked and developed manually, resulting in the version now in our Appendix. The AI tool Refine.ink was used to check the final version of the manuscript for consistency and clarity. We subsequently edited as needed and take full responsibility for the content of the published article.

References

Baye, M. R., J. Morgan, and P. Scholten (2004): "Price Dispersion in the Small and in the Large: Evidence from an Internet Price Comparison Site," *Journal of Industrial Economics*, 52(4), 463–496.

Burdett, K., and K. L. Judd (1983): "Equilibrium Price Dispersion," *Econometrica*, 51(4), 955–969.

Krishna, V. (2013): *Auction Theory*. Academic Press, MA, USA.

Myatt, D. P., and D. Ronayne (2025): "Finding a Good Deal: Stable Prices, Costly Search, and the Effect of Entry," *Working Paper including supplements and extensions*.

Nagaraja, H. (1992): "Order Statistics from Discrete Distributions," *Statistics*, 23(3), 189–216.

Stigler, G. J. (1961): "The Economics of Information," *Journal of Political Economy*, 69(3), 213–225.

Watt, M. (2025): "Concavity of Order Statistics in Sample Size: Implications for Auction Design," *Economics Letters*, 256(112628).

Wilks, S. S. (1962): *Mathematical Statistics*. New York: John Wiley.

Appendix A. Python code for discrete convexity checks

```
#!/usr/bin/env python3
"""
Discrete Convexity Checker for S_{jq}
- --nmin N          : minimum n to check (default 3)
- --nmax N          : maximum n to check (default 100)
- --csv PATH        : write any violations to a CSV file
- --early-stop      : stop on first violation
- --quiet           : suppress progress messages
"""

from __future__ import annotations
from math import comb
from fractions import Fraction
from dataclasses import dataclass
import argparse
import csv
from typing import List, Iterable

# ---------- Core combinatorics ----------
def pmf_Rq(n: int, q: int, k: int, ell: int) -> Fraction:
    if ell < k or ell > n - q + k:
        return Fraction(0)
    num = comb(ell - 1, k - 1) * comb(n - ell, q - k)
    den = comb(n, q)
    return Fraction(num, den)

def S_fraction(n: int, q: int, k: int, j: int) -> Fraction:
    if j == n:
        return Fraction(0)
    lo = max(j + 1, k)
    hi = n - q + k
    if lo > hi:
        return Fraction(0)
    total = Fraction(0)
    for ell in range(lo, hi + 1):
        total += (ell - j) * pmf_Rq(n, q, k, ell)
    return total

def delta2_S(n: int, k: int, j: int, q: int) -> Fraction:
    return (S_fraction(n, q,   k, j)
            - 2 * S_fraction(n, q+1, k, j)
            +     S_fraction(n, q+2, k, j))

```

```python
# ---------- Scanning & reporting ----------
@dataclass
class Violation:
    n: int
    k: int
    j: int
    q: int
    value: Fraction  # negative value of Delta^2 S

def admissible_tuples(n: int) -> Iterable[tuple[int, int, int, int]]:
    for k in range(1, n + 1):
        q_min, q_max = k, n - 2
        if q_min > q_max:
            continue
        for j in range(1, n):
            for q in range(q_min, q_max + 1):
                yield (k, j, q)

def check_range(nmin: int, nmax: int, early_stop: bool = False, quiet: bool =
    False) -> List[Violation]:
    violations: List[Violation] = []

    for n in range(max(3, nmin), nmax + 1):
        if not quiet:
            print(f"[n={n}] checking ...")
        for (k, j, q) in admissible_tuples(n):
            d2 = delta2_S(n, k, j, q)
            if d2 < 0:
                v = Violation(n=n, k=k, j=j, q=q, value=d2)
                violations.append(v)
                if not quiet:
                    print(f"  VIOLATION at (n={n}, k={k}, j={j}, q={q}): Delta2 =
                        {d2} \approx {float(d2):.12g}")
                if early_stop:
                    return violations
    return violations

def write_csv(path: str, violations: List[Violation]) -> None:
    with open(path, "w", newline="") as f:
        w = csv.writer(f)
        w.writerow(["n", "k", "j", "q", "delta2_fraction", "delta2_float"])
        for v in violations:
            w.writerow([v.n, v.k, v.j, v.q,
                        f"{v.value.numerator}/{v.value.denominator}",
                        float(v.value)])
```

```python

# ---------- CLI ----------
def main() -> None:
    ap = argparse.ArgumentParser(description="Check discrete convexity of S_{jq} in q.")
    ap.add_argument("--nmin", type=int, default=3, help="Min n to check (default: 3)")
    ap.add_argument("--nmax", type=int, default=100, help="Max n to check (default: 100)")
    ap.add_argument("--csv", type=str, default=None, help="Optional path to write violations as CSV")
    ap.add_argument("--early-stop", action="store_true", help="Stop on first violation")
    ap.add_argument("--quiet", action="store_true", help="Suppress progress messages")
    args = ap.parse_args()

    if args.nmax < args.nmin:
        print("Error: nmax must be \geq nmin.")
        return
    if args.nmax < 3:
        print("Nothing to check: need n \geq 3 so that q \leq n-2 exists.")
        return

    violations = check_range(args.nmin, args.nmax,
                             early_stop=args.early_stop, quiet=args.quiet)

    if not violations:
        print(f"No violations found: Delta^2_q S_jq \geq 0 for all {args.nmin} \leq n \eq {args.nmax} "
              f"and all admissible k, j, q.")
    else:
        print(f"Found {len(violations)} violation(s) in {args.nmin} \leq n \leq {args.nmax}.")
        for v in violations[:10]:
            print(f"  (n={v.n}, k={v.k}, j={v.j}, q={v.q})  Delta2={v.value} \approx {float(v.value):.12g}")
        if args.csv:
            write_csv(args.csv, violations)
            print(f"Wrote full list of violations to: {args.csv}")

if __name__ == "__main__":
    main()
```

LISTING 1. Discrete Convexity Checker for S_{jq}