Optimal information extractable in a measurement

H. M. Bharath\textsuperscript{1,2} and Saikat Ghosh\textsuperscript{1}

\textsuperscript{1}Department of Physics, Indian Institute of Technology - Kanpur, UP-208016, India
\textsuperscript{2}School of Physics, Georgia Institute of Technology, Atlanta, Georgia-30332, U.S.A

(Dated: December 1, 2014)

Abstract

In any measurement of a parameter, effect of finite resource is manifested in a finite number \((M)\) of probable values for the estimate. Here it is shown that this count of the discrete number of measurement outcomes leads to a pair of equivalent bounds: \(\delta \geq 1/M\) on precision(\(\delta\)) and \(I \leq \log(M)\) on mutual information \((I)\) between the parameter and its estimate. Furthermore, the probe particle symmetry is found to play a significant role in limiting \(M\) and thereby, in limiting extractable information in a measurement. In particular, for \(N\) indistinguishable spin-half probe particles, it is shown that \(M\) is tightly bounded to \(O(N^2)\). The corresponding operator that saturates this bound represents a non interferometric measurement. While inherently Heisenberg limited, this measurement is precise up to \(\delta \sim 1/N^2\). These results can be used to benchmark any precision measurement thereby establishing a particular link between information theory and metrology.
In measurement of a continuous parameter \((x)\), a range for the "true value" is inferred from the measurement outcome or the estimate \((z)\). Half-width \((\delta)\) of the inferred range, \(x \in [z - \delta, z + \delta]\), sets the measurement precision \([1, 2]\). The central theme for precision metrology has therefore been to minimize precision \([2-7]\). One can note that precise measurement is also equivalent to maximizing extractable information of the true value \([10]\). Here we explore this viewpoint and its possible implications on bounds on measurement precision.

Traditionally, one chooses an appropriate statistical “width” estimator, for example, the root-mean-square-error (RMSE \(\equiv \sigma(x)\)) to infer a range for the true value. Accordingly, \(\delta\) is estimated from a range: \(|x - z| \leq \sigma(x)\). While this procedure has wide applicability \([1, 2]\), here we observe that it does not always reflect optimal precision. In contrast, we propose an information theoretic measure for precision which is we use to analyze existing precision measurement models and propose measurement operators for extracting optimal quantum advantage in a measurement with \(N\) spin-half particles.

Limitations on resource (for example number of probe particles, measurement time or more generally, limitations on available energy) has a straightforward consequence on a model describing a measurement: it predicts a finite, distinct set of probable estimates \(z \in \{z_1, z_2 \cdots z_M\}\) with corresponding probabilities or likelihood functions \(\{P_1(x), P_2(x) \cdots P_M(x)\}\). Here each \(z\) corresponds to a final experimental outcome, that might include any number of repetitions of a basic measurement.

For example, one can consider a canonical experiment for estimating an unknown phase \((\phi)\), inferred from a measurement of \(x = \cos(\phi)\) in a Mach-Zender interferometer \([5\,\tilde{7}]\). For a single-photon incident on one of its input ports, the output evolves to a state \(|\psi_{out}\rangle_\phi\) with a parametric dependence on a phase-element inserted in one of the interferometer arms. The measured observable \((\hat{O})\) usually corresponds to the difference counts at the output ports with an expected value: \(\langle \psi_{out}|\hat{O}|\psi_{out}\rangle = \cos(\phi)\) (Fig.1b).

In this single shot experiment, the possible values of the estimate \(z\) are two distinct numbers \(z \in \{-1, 1\}\) i.e., \(M = 2\), with associated probabilities \(P_{-1}(\phi) = \sin^2(\phi/2)\) and \(P_{+1}(\phi) = \cos^2(\phi/2)\). While the numbers are the eigenvalues of \(\hat{O}\), the probabilities reflect a parametric dependence on \(\phi\) (Fig.1c).

\(N\) repetitions with \(N\) uncorrelated photons, increase the number of probable values \((M)\) of the outcome. This can be seen by considering the form of the measured operator \(\hat{O}\),
Any general measurement model of a continuous parameter \( x \) with finite resource has a discrete set of probable distinct estimated values or outcomes \( z \in \{ z_1, z_2 \cdots z_M \} \) with corresponding probability distributions or likelihood functions \( \{ P_1(x), P_2(x) \cdots P_M(x) \} \).

(b) As an example, one can consider a Mach-Zender interferometer to estimate an unknown phase \( \phi \) is. One version of the setup measures the difference counts in the two detectors \( D_1 \) and \( D_2 \) thereby estimating \( x = \cos(\phi) \). The incident single photon in port \( a \), corresponding to a state \( |\psi\rangle = |1\rangle_a |0\rangle_b \), evolves to \( |\psi_{\text{out}}\rangle = e^{i\phi/2}[\cos(\phi/2)|\psi\rangle - i\sin(\phi/2)|\tilde{\psi}\rangle] \), with \( |\tilde{\psi}\rangle = |0\rangle_a |1\rangle_b \). The corresponding measured operator in this case can then be expressed as:

\[
\hat{O} = |\psi\rangle \langle \psi| - |\tilde{\psi}\rangle \langle \tilde{\psi}|.
\]

(c) In a single shot measurement, there are two probable values \( z \in \{-1, 1\} \) with associated probabilities \( P_{-1}(\phi) = \sin^2(\phi/2) \) and \( P_{+1}(\phi) = \cos^2(\phi/2) \). For this case, \( M = 2 \).

(d) N-repetitions of this experiment with \( N \) uncorrelated single photons corresponds to \( M = N + 1 \) probable values: \( z \in \{-1, -1 + \frac{2}{N} \cdots - 1 + \frac{2k}{N} \cdots 1\} \) with associated probabilities distributed binomially. A single measurement outcome after \( N \)-repetitions return one of the \( z \)-s, with corresponding precision \( \delta_r \) calculated from the RMSE:

\[
\sigma(\phi) = \frac{\sin(\phi)}{\sqrt{N}}.
\]
which is now described in a $2^N$ dimensional Hilbert space as:

$$\hat{O} = \frac{1}{N}[(\hat{O} \otimes 1 \otimes \cdots \otimes 1) + \cdots + (1 \otimes \cdots \otimes \hat{O})].$$

It has $M = N + 1$ distinct eigenvalues while the associated probabilities are distributed binomially (Fig.1d). However, the expected value remains $\langle \hat{O} \rangle = \cos(\phi)$.

Following the traditional approach, for an experiment with an outcome $z_r$ (after $N$-repetitions), the true value can now be inferred from the estimate by inverting the relation $|\cos(\phi) - z_r| \leq \sigma(\phi)$ i.e.,

$$\cos(\phi) = \frac{Nz_r}{N+1} \pm \sqrt{\frac{1}{N+1} - \frac{Nz_r^2}{(N+1)^2}},$$

which asymptotically, for large $N$, reduces to $\cos(\phi) \in \left[z_r - \frac{\sqrt{1-z_r^2}}{\sqrt{N}}, z_r + \frac{\sqrt{1-z_r^2}}{\sqrt{N}}\right]$. It can be noted that the precision depends on the specific estimate $z_r$. However, in the allowed range of $x \in [-1, 1]$, the maximum it can take sets the precision for the model. Equivalently, $\delta = \delta_{\text{max}} \sim \frac{1}{\sqrt{N}}$. This asymptotic, $N^{-1/2}$ scaling of precision, with $N$ repetitions, is usually termed as the Standard Quantum Limit (SQL) [3–7, 9].

Here $\delta$ is estimated from the inferred range set by the RMSE i.e., $\delta \sim \text{RMSE}$. There has therefore been extensive search for optimal bounds on RMSE leading to generalized versions of Fisher Information and Cramer-Rao bound [2, 11]. However, for any measurement with arbitrary output probability distributions, this procedure might not be well motivated as is found here for Quantum Phase Estimation Algorithms (QPEA)(inset of Fig.2c and Table I).

**Results:**

**Limiting distributions**

Nevertheless, a bound on $\delta$ can be evaluated from a simple argument, independent of $\sigma(x)$. We next argue that for $M$ probable values of the estimate, the precision can at best be $\frac{1}{M}$.

As was the case in the above example of SQL, though $\delta_r$ depends on a specific measurement outcome $z_r$, the overall precision for a model is set by the worst it can do over the possible range of $x$ (without loss of generality, here and for all subsequent cases, the range of $x$ is scaled to $\pm 1$). This is equivalent to finding the maximum ($\delta_{\text{max}}$) over all $\delta_r$. Furthermore, one can note that $\delta_s$ together cover the probable range of $x$, or equivalently,
\[ \sum_{r=1}^{M} \delta_r \geq 1. \] Consequently, \( \delta_{\text{max}} \) satisfies the following inequality:

\[
\delta_{\text{max}} \geq \frac{1}{M} \sum_{r=1}^{M} \delta_r \geq \frac{1}{M} \tag{1}
\]

which sets a lower bound for precision: \( \delta \geq 1/M. \)

It needs to be established that this bound is achievable. One can however note that the first inequality is saturated when all \( \delta_r \)'s are equal i.e., \( z_r \)'s are equispaced. We next demonstrate that the second inequality is tight by explicitly constructing a distribution that saturates it.

One can consider a distribution with \( P_r(x) = 1 \) for the estimate \( z_r \) that is closest to the natural value \( x \) and zero for all others (Fig.2a). It corresponds to a measurement that always returns the same value \( z_r \). The corresponding inferred range for the true value is \( x \in \left[ \frac{z_r-1+z_r}{2}, \frac{z_r+z_r+1}{2} \right] \). Since \( x \) is unknown, it is straightforward to note that precision is optimized when all \( z_r \)'s are equispaced, saturating the bound \( \delta = 1/M \).

Clearly, this distribution represents a biased estimator [1]. We refer to it as biased Minimum-Error-Distribution or MED (biased) for reasons stated later.

One can probably appreciate the argument above to be information theoretic in nature. We next show that indeed the mutual information \( (I) \) [15,18] between \( x \) and \( z \) is bounded:

\[ I \leq \log(M), \tag{2} \]

with the bound saturated for MED (biased). This implies that the true value, can be estimated up to \( \log(M) \) bits or equivalently, to a precision of \( 1/M \) from the estimate \( z \). Accordingly, a reasonable expression to evaluate precision for a general distribution is \( \delta = e^{-I} \) which is independent of RMSE or any other statistical “width” functions (see Table I for a comparison of several standard cases).

To prove the inequality (2), one can note that the joint probability distribution of \( x \) and an estimate \( z_r \) is \( P(x,z_r) = P(x)P_r(x) \). Here \( P(x) \) is the probability for the parameter in its probable range. Since \( x \) is unknown, it is equally likely (no prior) to be anywhere in the (scaled) range \([-1,1]\) with a uniform distribution of \( P(x) = 1/2 \). The mutual information is then, up to an additive constant:

\[
I = \sum_r \int dx P(x,z_r) \log(P(x,z_r)) - \sum_r \left( \int dx P(x,z_r) \right) \log \left( \int dx P(x,z_r) \right).
\]
**Figure 2: Distribution of estimates:** (a) A biased distribution that minimizes precision $\delta$ in estimation of a single parameter $x$. (b) Corresponding unbiased distribution. These two distributions also uniquely minimizes RMSE, while the unbiased estimator saturates the C-R bound, thereby optimizing Fisher information for any measurement with finite resource i.e., with $M$ discrete outcomes. They are therefore termed here as Minimum-Error-Distribution or MED (biased) and MED (unbiased), respectively. (c) A schematic of the circuit implementing Quantum Phase Estimation Algorithm(QPEA) for $N = 4$ (d) Corresponding distribution of probable values with a precision $\delta = 1/2^4$, which is very close to MED(unbiased). However, the RMSE for this distribution is ill-defined, rapidly oscillating with varying $\phi$ (or $N$), as shown in the inset.

The first term on the right is bounded above by 0 while the second is bounded below by $-\log(M)$ implying $I \leq \log(M)$.

It is instructive also to consider unbiased estimators with an additional constraint: the expected value to be equal to $x$. The corresponding limiting probability distribution that minimizes $\delta$ has zero probabilities for all but $z_r \leq x \leq z_{r+1}$. The non-zero $P_r(x)$ and $P_{r+1}(x)$ can be trivially evaluated from the unbiased condition and that the probabilities add up to unity (Fig.2b). A probable estimate $z_r$ then implies a range $[z_{r-1}, z_{r+1}]$ for the true value.
| Measurement          | RMSE       | Precision($\delta$) | Mut. Info.(I) | $e^{-I}$ | # of probes(N) |
|----------------------|------------|---------------------|---------------|----------|----------------|
| SQL                  | $\frac{1}{\sqrt{M}}$ | $\frac{1}{\sqrt{M}}$ | $\frac{1}{2} \log(M/(2\pi e))$ | $\frac{\sqrt{2\pi e}}{\sqrt{M}}$ | $M = N + 1$ |
| MED(Biased)          | $\frac{1}{12M}$ | $\frac{1}{M}$ | $\log(M)$ | $\frac{1}{M}$ | – |
| MED(Unbiased)        | $\frac{1}{M}$ | $\frac{2}{M}$ | $\log(M) - 1/2$ | $\frac{\sqrt{\pi}}{M}$ | – |
| QPEA [17]            | $\sqrt{\frac{8 \log(2)}{\pi N}}$ | $\frac{1}{M}$ | $\log(M) - 2(1 - \gamma)$ | $\sqrt{\frac{1 - \gamma}{M}}$ | $M = 2^N$ |
| Q-Metrology [22]     | $\sim \frac{1}{10\sqrt{N}}$ | $\frac{1}{\sqrt{M}}$ | $\sim \frac{1}{2} \log(M)$ | $\sim \frac{1}{\sqrt{M}}$ | $M = N^{\log(2\nu+1)}$ |

Table I: A comparison of precision and RMSE for a few standard cases (as discussed in the text). Mut. Info. represents Mutual Information. Approximate expressions in the asymptotic regime are indicated, the rest are exact (see Supplementary Information for details). Usual symbols for standard constants are used, $\gamma$ being the Euler-Mascheroni constant. The last column lists $M$ for each case (where applicable) in terms of the number of probe particles ($N$). For quantum metrology (Q-Metrology), $\nu$ is the number of classical repetitions as discussed in the text.

The corresponding precision is again optimized for equispaced $z_i$ and yields $\delta = \frac{2}{M}$. This distribution will be referred to as MED (unbiased).

It is interesting to note that the above two limiting distributions also uniquely minimize the RMSE for their respective cases [11, 12], hence the terminology Minimum-Error-Distribution (MED). Furthermore, MED (unbiased) saturates the Cramer-Rao (CR) bound with optimal Fisher information corresponding to any discrete distribution with fixed $M$. One can therefore conclude that any measurement model that saturates the bound on RMSE will have to yield to the MEDs [See supplementary information for details on MEDs].

"Precision-Capacity"

The above results suggest $M$ to be the only quantity limiting attainable precision. It is therefore defined hereafter as precision-capacity, which is the number of distinct probable outcomes in any measurement model of a classical parameter (including all repetitions in a measurement). Since precision capacity is independent of dynamical details (classical or quantum), it sets a universal bound on precision and can act as a basis for comparing a wide range of models [13, 14, 16].

Measurement models with quantum dynamics
For measurement models incorporating quantum-mechanical dynamics, *precision-capacity* is the number of distinct eigenvalues of the measured observable ($\hat{O}$). To optimize precision, a reasonable strategy is therefore to choose an observable with the largest possible precision-capacity corresponding to chosen probe states.

In absence of any additional constraints on the model, the dimensionality of the Hilbert space $\text{dim}\{\mathcal{H}\}$ sets an obvious bound i.e.,

$$M \leq \text{dim}\{\mathcal{H}\}.$$  \hspace{1cm} (3)

This inequality is saturated for any non-degenerate observable. Typically, $\text{dim}\{\mathcal{H}\}$ is exponential with the number of probe particles ($N$) used, suggesting a possibility of exponential scaling for the corresponding precision with $N$ \[14\].

Quantum phase estimation Algorithm (QPEA) \[17\] due to Kitaev is a classic example that reaches $M = \text{dim}\{\mathcal{H}\}$ attaining a precision that almost saturates the corresponding bound: $\delta \geq \frac{1}{M}$ (see Table I). An important building block for quantum computers \[18\], it has also been used in interferometric measurements for estimating an unknown phase ($\phi$) \[19\] (Fig.2c).

QPEA uses $N$ spin-half probe particles particles, while the measured observable can be identified as \[17, 18\]:

$$\mathcal{O}_{\text{QPEA}} = \sum_{k=0}^{2^N - 1} \frac{2\pi k}{2^N} |k\rangle \langle k|,$$  \hspace{1cm} (4)

with $\{|k\rangle\}$ representing the computational basis. The corresponding distribution of probable estimates is a close approximation of MED (unbiased) (See Fig.2d) with a precision $\delta \sim 1/2^N$ (see Table I).

Interestingly, though the precision is well defined for QPEA, the corresponding RMSE is not. It rapidly oscillates with varying $\phi$ (inset of Fig.2d) reaching a maxima of the order $\sim \frac{1}{\sqrt{2^N}}$ (see Table I). RMSE therefore completely fails to capture the error estimate for QPEA.

One can note that this maximum precision-capacity ($M = 2^N$) can be achieved for a measurement using $N$ classical (i.e., non-quantum encoded) bits. Therefore quantum bits show no particular precision enhancement. This is expected as a consequence of Holevo’s theorem \[20\]. Equivalently, complete *distinguishability* of states is utilized to saturate the precision capacity in QPEA.
We next show if one imposes a constraint of particle \textit{indistinguishability}, quantum bits show a significant enhancement in precision-capacity, thereby establishing a true quantum advantage in precision metrology.

For \( N \) identical classical bits, the precision-capacity is trivially bounded as \( M = N + 1 \). For \( N \) indistinguishable quantum bits, we show that the precision-capacity is tightly bounded to \( O(N^2) \) by proving the following theorem (for a comparison of all cases based on probe particle symmetry, see Table II):

\textbf{Theorem:} For a system of \( N \) indistinguishable spin-half particles, precision-capacity \( M \) is tightly bounded by the following inequalities

\[
M \leq \frac{(N+2)^2}{4} \quad \text{for even } N
\]
\[
M \leq \frac{(N+1)(N+3)}{4} \quad \text{for odd } N.
\]

The inequality is saturated for the observable, \( \hat{O} = J_N^2 + \epsilon J_N^{(z)} \), where \( J_N \) is the total spin operator for \( N \) particles while \( J_N^{(z)} \) is its \( z \) component and \( \epsilon \) is a number, small compared to the eigenvalues of \( J_N^2 \), such that simultaneous eigenspaces of \( J_N^2 \) and \( J_N^{(z)} \) have distinct eigenvalues.

The symmetric group \( S_N \) (permutation group of \( N \) particles) has a unitary representation in the corresponding Hilbert space, \( H^{\otimes N} \). For an observable \( \hat{O} \) to respect particle indistinguishability, each of its eigenspaces should be invariant under the unitaries representing \( S_N \). The number of distinct eigenvalues, \( M \), is then the number of sub-representations in the group. \( M \) is therefore bounded above by the number of irreducible representations (counting isomorphisms) in an irreducible decomposition. This number is independent of the decomposition itself \cite{21}. Therefore, the above theorem may be restated as:

\textbf{Theorem (restated):} The common eigenspaces of \( J_N^2 \) and \( J_N^{(z)} \) are irreducible under \( S_N \).

The details of the proof is outlined in \textit{Methods} section below. The precision in a measurement that uses this operator as the observable is tightly bounded by \( \delta \geq \frac{1}{N^2} \).

\textbf{Discussions:}

It may be observed that, often a combination of distinguishable and indistinguishable probes (or groups of them) are used to optimize estimation of an unknown phase. Furthermore, the measured operator is usually limited to that of a \textit{presumed} interferometric
### Table II: Upper bounds on precision-capacity based on symmetry of $N$ spin-half probe particles for classical (non-superimposable) and quantum bits.

| Symmetry       | Classical | Quantum |
|----------------|-----------|---------|
| Indistinguishable | $2^N$     | $2^N$   |
| Distinguishable  | $\sim N$  | $\sim N^2$ |

scheme with an ability to project an $N$ particle entangled output state to the corresponding entangled input. We end with an analysis of a representative model for such schemes.

Quantum Metrology (Giovanetti et. al.) [22] models estimate an unknown phase($\phi$) to a precision of $1/N$, saturating the Heisenberg limit [5, 7]. However, we find that $M \sim N^2$ in this case failing to saturate precision-capacity.

At its simplest, these models use a maximally entangled input state $|\Psi_{in}\rangle = \frac{1}{\sqrt{2}}[|\psi\rangle^{\otimes N} + |\tilde{\psi}\rangle^{\otimes N}]$ (here, $|\psi\rangle$ and $|\tilde{\psi}\rangle$ are as defined in the caption of Fig.1). The corresponding measured observable is: $O' = |\psi\rangle \langle \psi|^{\otimes N} - |\tilde{\psi}\rangle \tilde{\langle \psi|}^{\otimes N}$ with an expected value: $\langle O' \rangle = \cos(N\phi)$, thereby estimating $N\phi$ instead of $\phi$. This observable has three eigenvalues: $\{-1, 0, +1\}$ which are the probable values of the estimate, with $M = 3$. Repeating this experiment $\nu$ times yields a precision of $\frac{1}{\sqrt{\nu}}$ for $N\phi$ while $\phi$ is determined only upto one significant digit [9, 23].

However, as pointed out by Giovannetti et. al., with a choice of $N = 10^j/2\pi$, one obtains the $j^{th}$ decimal place for $\phi$ with a precision of $1/\sqrt{\nu}$. The estimation will be exact for choice of $\nu \geq 100$. Therefore, a precision $1/N$ can be recovered by combining sequentially a series of runs with varying $N$, with a choice of $j = 1, 2, 3 \cdots \log(N)$ [22].

The model predicts $M = N^{\log(2\nu+1)}$ probable values of the estimates or outcomes. The run for the $j^{th}$ decimal place requires $\nu$ repetitions of an experiment each with three outcomes $\{-1, 0, 1\}$ and therefore, has overall $2\nu + 1$ outcomes. Since each decimal place is independently estimated, the total number of outcomes is

$$M = (2\nu + 1) \times \cdots \times (2\nu + 1) = (2\nu + 1)^{\log(N)} = N^{\log(2\nu+1)},$$

which reduces to $M \sim N^2$ for $\nu = 100$.

It can be noted that in this model, while $\nu$ particles used in estimating a particular decimal place are indistinguishable, the batches used for different decimal places are not.

Here we have presented an information theoretic approach to precision metrology. This
is in contrast to the usual analysis of RMSE and its scaling with a count for resource, which is not always transparent [9]. Table I summarizes a comparison of RMSE and an independently evaluated precision ($\delta$) for a few standard measurement schemes discussed in this work. There is a strong disagreement for QPEA indicating RMSE is not always an effective measure of precision. It is also found that a measure set by mutual information $I$ between the true value of a continuous parameter and its probable discrete estimates, is a reasonable and universal measure to characterize and design optimal precision measurements in general, as confirmed by Table I. It can be noted that internal dynamical details in a specific model can nevertheless set a tighter bound, as is the case for classical repetitions or quantum metrology, with bounds set by SQL and Heisenberg limit, respectively.

However, it is also found that for $N$ spin-half probe particles, particle symmetry plays a key role in limiting precision-capacity, a quantity defined here as a count of probable discrete outcomes in any measurement with finite resource (Table II summarizes these results). Accordingly, for $N$ completely indistinguishable particles, a class of operators $\hat{O} = J^2 + \epsilon J_z$ is proposed here that optimizes precision capacity, leading to a bound, $\delta \sim \frac{1}{N^2}$, saturating both the Heisenberg limit and the information theoretic bound. This therefore exploits optimally, quantum-advantage in such measurements. It can finally be noted that usually in quantum metrology, interferometric projective measurements on input-states are usually envisioned as measurement operators. Nevertheless, experimental realization of such measuring optical elements with many probe photons have remained elusive [5, 7, 16]. Instead, estimation of a classical parameter in a “natural” system of ensemble of dilute (indistinguishable) cold atoms in the regime of Dicke super-radiance [3, 24] might provide a probable route of realizing the measurement operator proposed here, in future.

Methods:
Proof of the theorem:

We denote the common eigenspace of $J_N^2$ and $J_N^{(z)}$ by $\mathcal{V}_{Njm}$, with eigenvalues $j(j+1)$ and $m$ respectively. Since $J_N^2$ and $J_N^{(z)}$ are invariant under $S_N$, these spaces too are invariant under $S_N$. In Theorem-1, we prove that they are also irreducible under $S_N$. This theorem is proved by induction on $N$. It is easily verified to be true for $N = 2$, i.e., $\mathcal{V}_{2jm}$ are irreducible under $S_2$. Assuming that it is true for $2, 3 \cdots N-1$, we prove that it is true for $N$, by showing that $\mathcal{V}_{Njm}$ has no proper subspace invariant under $S_N$. To begin, we decompose $\mathcal{V}_{Njm}$ in to subspaces invariant under $S_{N-1}$ (the subgroup of $S_N$ consisting of...
permutations of the first \(N-1\) particles.) The operator \(J_{N-1}^2 \otimes I\) is invariant under \(S_{N-1}\). Therefore, its two eigen spaces within \(V_{Njm}\) are invariant under \(S_{N-1}\). We refer to them by: \(\mathcal{U}_{j^t_{\frac{1}{2}}}\). They have a natural basis:

\[
\mathcal{U}_{j^t_{\frac{1}{2}}} : C_{\pm 1} | j \pm \frac{1}{2}, m - \frac{1}{2} \rangle \otimes | \frac{1}{2}, \frac{1}{2} \rangle \\
+ C_{\pm 2} | j \pm \frac{1}{2}, m + \frac{1}{2} \rangle \otimes | \frac{1}{2}, -\frac{1}{2} \rangle
\]

(6)

here, \(C_{\pm i}\) are Clebsch-Gordon coefficients given by, \(C_{\pm 1} = C_{m - \frac{1}{2}, j^t_{\frac{1}{2}}, m}^{j^t_{\frac{1}{2}}, j^t_{\frac{1}{2}}}, m \pm \frac{1}{2}\). These two spaces are not only invariant, but they are indeed irreducible under \(S_{N-1}\). We use the induction hypothesis to prove this.

**Lemma 1:** The spaces \(\mathcal{U}_{j^t_{\frac{1}{2}}}\) are irreducible under \(S_{N-1}\)

**Proof:** Consider the map \((J_{N-1}^- \otimes \sigma_+) : \mathcal{U}_{j^t_{\frac{1}{2}}} \rightarrow V_{N-1, j^t_{\frac{1}{2}}, m - \frac{1}{2}} \otimes | \frac{1}{2}, \frac{1}{2} \rangle\). This map is invariant under \(S_{N-1}\). Let \(W \subset \mathcal{U}_{j^t_{\frac{1}{2}}}\) be a subspace invariant under \(S_{N-1}\). We have, \((J_{N-1}^- \otimes \sigma_+) W \subset V_{N-1, j^t_{\frac{1}{2}}, m - \frac{1}{2}} \otimes | \frac{1}{2}, \frac{1}{2} \rangle\) is also invariant under \(S_{N-1}\). Noting that, from the induction hypothesis, \(V_{N-1, j^t_{\frac{1}{2}}, m - \frac{1}{2}}\) is irreducible under \(S_{N-1}\) and \(\text{Dim}(V_{N-1, j^t_{\frac{1}{2}}, m - \frac{1}{2}}) = \text{Dim}(\mathcal{U}_{j^t_{\frac{1}{2}}})\), we conclude that \(W = \mathcal{U}_{j^t_{\frac{1}{2}}}\) or \(W = 0\), and therefore, \(\mathcal{U}_{j^t_{\frac{1}{2}}}\) are irreducible under \(S_{N-1}\).

Returning to the main proof, we consider the following cases:

**Case-1:** \(\text{Dim}(\mathcal{U}_{j^t_{\frac{1}{2}}}) \neq \text{Dim}(\mathcal{U}_{j^t_{-\frac{1}{2}}})\)

Let \(W\) be an invariant subspace of \(V_{Njm}\), under \(S_N\). Clearly, \(W\) is also invariant under \(S_{N-1}\), since \(S_{N-1} \leq S_N\). By Schur’s lemma, \(W = \mathcal{U}_{j^t_{\frac{1}{2}}}\) or, \(W\) is the full space \(V_{Njm}\). However, \(\mathcal{U}_{j^t_{\frac{1}{2}}}\) are not invariant under \(S_N\). Therefore, \(V_{Njm}\) is irreducible under \(S_N\).

**Case-2:** \(\text{Dim}(\mathcal{U}_{j^t_{\frac{1}{2}}}) = \text{Dim}(\mathcal{U}_{j^t_{-\frac{1}{2}}})\)

By Schur’s lemma, if \(W \subset V_{Njm}\), is invariant under \(S_N\), it is either the full space \(V_{Njm}\) or it is also irreducible under \(S_{N-1}\), i.e, \(\text{Dim}(W) = \text{Dim}(\mathcal{U}_{j^t_{\frac{1}{2}}} )\). We shall show that the latter case leads to a contradiction. Let us define the projections maps from \(\mathcal{U}_{j^t_{\frac{1}{2}}}\) to \(W\) and vice versa.

**Def:** \(\pi : \mathcal{U}_{j^t_{\frac{1}{2}}} \rightarrow W\), and \(\pi' : W \rightarrow \mathcal{U}_{j^t_{\frac{1}{2}}}\)

The operators \(\pi \pi' : \mathcal{U}_{j^t_{\frac{1}{2}}} \rightarrow \mathcal{U}_{j^t_{\frac{1}{2}}}\) and \(\pi' \pi : W \rightarrow W\) are invariant under the unitary representations of \(S_{N-1}\), when \(\text{Dim}(W) = \text{Dim}(\mathcal{U}_{j^t_{\frac{1}{2}}} )\). Therefore, by Schur’s lemma, \(\pi \pi' = \lambda I\) in \(\mathcal{U}_{j^t_{\frac{1}{2}}}\) and \(\pi' \pi = \lambda I\) in \(W\). If \(\lambda = 0\), \(W\) is orthogonal to \(\mathcal{U}_{j^t_{\frac{1}{2}}}\) and therefore contained in \(\mathcal{U}_{j^t_{-\frac{1}{2}}}\). However, since \(\mathcal{U}_{j^t_{-\frac{1}{2}}}\) is irreducible, \(W = \mathcal{U}_{j^t_{-\frac{1}{2}}}\), which is contradiction because the latter is not invariant under \(S_N\).
If $\lambda \neq 0$, $\pi'\pi$ is invariant under the unitary representations of $S_N$, since $W$ is an invariant subspace. Also, $\pi$ is invariant under $S_N$. Therefore, if $U_\sigma$ is a unitary matrix representing $\sigma \in S_N$, $\lambda I = U_\sigma \pi'\pi U_\sigma^\dagger = U_\sigma \pi' U_\sigma^\dagger \pi \implies \pi' = U_\sigma \pi' U_\sigma^\dagger$. Thus, $\pi'$ is also invariant under $S_N$. This means, $U_{j+\frac{1}{2}}$ is invariant under $S_N$, a contradiction.

**Acknowledgments:**

We acknowledge J. Kolodynski and V. Ravishankar for fruitful discussions.

**Author Information:**

Author emails: bharath.hm@gatech.edu(H. M. Bharath) and gsaikat@iitk.ac.in(Saikat Ghosh)

[1] JCGM, *Evaluation of measurement data: A guide to the expression of uncertainty in measurements*, (2008).

[2] Helstrom, C. W. *Quantum Detection and Estimation Theory* (Academic, 1976).

[3] Wineland, D. J., Bollinger, I. J., Itano, W. M. and Moore, F. L. Spin squeezing and reduced quantum noise in spectroscopy. *Phys. Rev. A* **46**, R 6797 (1992).

[4] Braunstein, S. L. Quantum limits on precision measurement in phase. *Phys. Rev. Lett.* **69**, 3598 (1992).

[5] Giovannetti, V., Lloyd, S. and Maccone, L., Quantum-enhanced measurements: Beating the standard quantum limit. *Science* **306**, 1330 (2004).

[6] Wiseman, H. M. and Milburn, G. J. *Quantum Measurement and Control* (Cambridge University Press, 2010).

[7] Giovannetti, V., Lloyd, S. and Maccone, L., Advances in quantum metrology, Nature Photonics **5**, 222 (2011).

[8] Childs, A. M., Preskills, J. and Renes, J. Quantum information and precision measurement, J. Mod. Optic, **47**, 155 (2000).

[9] Zwierz, M. C.A. Pérez-Delgado and P. Kok, General Optimality of the Heisenberg Limit for Quantum Metrology, *Phys. Rev. Lett.* **105**, 180402 (2010).

[10] Childs, A. M., Preskills, J. and Renes, J. Quantum information and precision measurement, J. Mod. Optic, **47**, 155 (2000).

[11] Cramer, H. *Mathematical Methods of Statistics*, (Princeton University Press, 1946). C. R. Rao,
Information and Accuracy Attainable in the Estimation of Statistical Parameters, Bull. Calcutta Math. Soc. (1945).

[12] Bestekas, D. P., *Nonlinear Programming: 2nd Ed.* (Athena Scientific, 1999).

[13] Boixo, S., Flammia, S. T., Caves, C. M. and Geremia, J. M. Generalized limits for single-parameter quantum estimation. Phys. Rev. Lett., 98, 090401 (2007).

[14] Roy, S. M., and Braunstein, S. L. Exponentially enhanced quantum metrology. Phys. Rev. Lett. 100 220501 (2008).

[15] Shannon, C. E., A Mathematical Theory of Communication, The Bell System Technical Journal, 27 379 (1948).

[16] Demkowicz-Dobrzanski, R., Kolodynski, J. and Guta, M. The elusive Heisenberg limit in quantum-enhanced metrology. Nature Comm., 3, 1063 (2012).

[17] Kitaev, A. Yu. Quantum measurements and the Abelian stabilizer problem. arXiv preprint quant-ph/9511026 (1995).

[18] Niesen, M. A. and Chuang, I. L., *Quantum Computation and Quantum Information*, (Cambridge University Press, 2000)

[19] Higgins, Brendon L., et al. Entanglement-free Heisenberg-limited phase estimation. Nature 450 393 (2007).

[20] Holevo, A. S. Bounds for the quantity of information transmitted by a quantum communication channel. Problems of Information Transmission 9 177 (1973).

[21] Serre, J. P., *Linear Representations of Finite Groups*, (Springer, 1977).

[22] Giovannetti, V., Lloyd, S. and Maccone, L., Quantum metrology, Phys. Rev. Lett. 96 (2006) 010401.

[23] Xiang, Guo-Yong, et al. Entanglement-enhanced measurement of a completely unknown optical phase. Nature Photonics 5 43 (2011).

[24] Dicke, R. H., Coherence in spontaneous radiation processes., Phys. Rev. 93, 99, (1954).
SUPPLEMENTARY INFORMATION

I. ENTRIES OF TABLE-I

In this section we outline the calculations for the RMSE and Mutual Information(I) of Table-I in the main text. We start with a brief discussion on the definition of mutual information for our case:

**Mutual Information**

Mutual information between \(x\) and \(z\) is defined in terms of the joint probability distribution \(P(x, z_r)\). One can note that here while \(x\) is a continuous variable, \(z \in \{z_1 \cdots z_M\}\) is discrete. Accordingly, sum over \(x\) is an integral in the range \(x : [-1, 1]\) and for \(z\), it is a simple summation, from \(r = 1\) to \(r = M\). The mutual information is then:

\[
I = \sum_r \int dx P(x, z_r) \log(P(x, z_r))
- \sum_r \left( \int dx P(x, z_r) \right) \log \left( \int dx P(x, z_r) \right) - \int dx \left( \sum_r P(x, z_r) \right) \log \left( \sum_r P(x, z_r) \right).
\]

For a measurement of an unknown parameter, the “true value” \(x\) is equally likely to be anywhere in \([-1, 1]\). Therefore, the joint probability distribution is \(\frac{1}{2} P_r(x)\). Substituting, one obtains a simpler form:

\[
I = \log(M) + \frac{1}{2} \sum_r \int dx P_r(x) \log(P_r(x))
\]

(8)

Here we have used

\[
P(x, z_r) = \frac{1}{2} P_r(x)
\]

\[
\int dx P(x, z_r) = 1/M
\]

\[
\sum_r P(x, z_r) = 1/2
\]

(9)

These expressions follow again from the fact that \(x\) is equally likely to be anywhere in the interval \([-1, 1]\). We now take up each of the examples and calculate the mutual information and RMSE for each:
A. SQL(Binomial)

1. Mutual Information

For the binomial distribution considered in the very first example, we have, \( M = N + 1 \) and \( P_r(x) = \binom{N}{r} \left( \frac{1+x}{2} \right)^{N-r} \left( \frac{1-x}{2} \right)^r \). Using Stirling’s approximation, we have:

\[
\log(n!) = n \log(n) - n + \frac{1}{2} \log(2\pi n) + \epsilon(n)
\]

(10)

The remainder \( \epsilon(n) \sim \frac{1}{12n} \) can be ignored, since it is found to contribute nothing to the leading order. With this approximation, we have:

\[
\log(P_r(x)) = N \log(N) - N \log(2) + \frac{1}{2} \log(N) - \frac{1}{2} \log(2) - \frac{1}{2} \log(N - r) - r \log(r) - (N - r) \log(N - r) + (N - r) \log(1 - x) + r \log(1 + x)
\]

(11)

Evaluating the integral and sum for the constant terms in the above equation (the first line on the RHS), we obtain:

\[
2N \log(N) - 2N \log(2) + \log(N) - \log(2\pi).
\]

(12)

After performing the integral over \( x \), we are left with a sum over \( r \) for the terms that depend only on \( r \) (second line in the RHS): 

\[
\frac{2}{N+1} \sum_r (1 + 2r) \log(r)
\]

(13)

This can be evaluated by approximating with an integral. Accordingly, we obtain:

\[
\sum_r (2r + 1) \log(r) = 2N \log(N) + 2 \log(N) - N + 1
\]

(14)

The last line of \( (11) \) consists of terms that depend both on \( r \) and \( x \). The two terms are identical after the sum and integral, by symmetry. Observing that \( \sum_r r P_r(x) = N \frac{1-x}{2} \), we obtain

\[
N \int (1 + x) \log(1 + x) dx = 2N \log(2) - N
\]

(15)

Substituting \( (12), (14) \) and \( (15) \) in \( (8) \) we obtain,

\[
I = \log(N) + \frac{1}{2} \left( - \log(N) - \log(2\pi) \right) = \frac{1}{2} \log(N) - \log(\sqrt{2\pi e})
\]

(16)
2. RMSE

By law off large numbers, the RMSE is,

\[ \sigma = \sqrt{\frac{1 - x^2}{N}} \]  

(17)

B. MED(Biased)

Every probability in for this distribution is zero, except for one. Therefore, the second term in (8) vanishes. Thus,

\[ I = \log(M) \]  

(18)

C. MED(Unbiased)

Formally, the MED(Unbiased) can be stated as:

\[ P_{k}^{MED}(x) = \begin{cases} \frac{x-z_{k-1}}{z_{k-1}} & z_{k-1} \leq x \leq z_{k} \\ \frac{z_{k+1}-x}{z_{k+1}} & z_{k} \leq x \leq z_{k+1} \\ 0 & \text{otherwise} \end{cases} \]  

(19)

1. Mutual Information

We first perform the integral in (8) for a fixed \( r \). Observing that \( P_{r}(x) \) is non zero only in the two bins adjoining the point \( x_{r} \), and the length of each bin is \( \Delta = \frac{2}{M} \), we obtain the second term in (8):

\[ \int dx P_{r}(x) \log(P_{r}(x)) = \int_{-\Delta}^{\Delta} \frac{|x|}{\Delta} \log\left(\frac{|x|}{\Delta}\right)dx = -\frac{\Delta}{2} \]  

(20)

Summing over \( r \), we obtain,

\[ I = \log(M) - \frac{1}{2} \]  

(21)

2. RMSE

A straightforward estimation gives the RMSE for this distribution as:

\[ \sigma = \sqrt{(x-z_{r})(z_{r+1}-x)} \approx \frac{1}{M} \]  

(22)
D. Quantum Phase estimation

1. Mutual Information

For this case, $M = 2^N$, where $N$ is the number of particles. The probability distribution in terms of the unknown phase $\phi$ is given by

$$P_r(\phi) = \frac{1}{M^2} \left| \frac{1 - z_r^M}{1 - z_r} \right|^2$$

with $z_r = e^{i(\phi - \frac{2\pi r}{M})}$. Note that $\phi$ belongs to a range of $[0, 2\pi]$. Therefore, the factor of $\frac{1}{2}$ in (8) is replaced by $\frac{1}{2\pi}$. Taking logarithm of the probability, we obtain:

$$\log(P_r(\phi)) = -2 \log(M)$$

$$\log(1 - z_r^M) + \log(1 - z_r^{-M}) - \log(1 - z_r) - \log(1 - z_r^{-1})$$

The first line in (24) is a constant. Therefore, sum over $r$ and integral over $\phi$ yields $-4\pi \log(M)$. The second line consists of terms independent of $r$, since $z_r^M = e^{iM\phi}$. Therefore, we may sum over $r$ and use a Taylor expansion to integrate over $\phi$, since $|z_r| = 1$. We thus have,

$$\int \log(1 - e^{iM\phi}) = -\sum_{k=1}^{\infty} \int \frac{e^{iMk\phi}}{k} d\phi = 0$$

Thus, the second line does not contribute anything. The third line in (24) consists of terms that depend on both $r$ and $\phi$. After a Taylor expansion, one can express:

$$\log(1 - z_r) = -\sum_{k=1}^{\infty} \frac{z_r^k}{k}$$

The probability distribution (23) can also be rewritten as:

$$P_r(\phi) = \frac{1}{M^2} \sum_{i,j=0}^{M-1} z_r^{i-j}$$

Performing the integral over $\phi$, one then obtains:

$$\int d\phi P_r(\phi) \log(1 - z_r) = -\frac{1}{M^2} \sum_{k=1}^{\infty} \sum_{i,j=0}^{M-1} \int \frac{z_r^{k+i-j}}{k} d\phi = -\frac{2\pi}{M^2} \sum_{k=1}^{\infty} \sum_{i,j=0}^{M-1} \frac{\delta_{k+i-j}}{k} = -\frac{2\pi}{M^2} \sum_{k=1}^{M} \frac{M - k}{k}$$

(28)
Using harmonic sums, the sum on the RHS can be written as:

\[-\frac{2\pi}{M^2} \sum_{k=1}^{M} \frac{M - k}{k} = -\frac{2\pi}{M} (\log(M) + \gamma) + \frac{2\pi}{M} + o\left(\frac{1}{M^2}\right) \tag{29}\]

Summing over \(r\) and noting that the two terms in the last line are identical after the sum and integral, we obtain,

\[-4\pi \log(M) + 4\pi(1 - \gamma) \tag{30}\]

Thus, the mutual information is,

\[I = \log(M) - 2(1 - \gamma) \tag{31}\]

2. RMSE

Let us consider the case where \(\phi = \pi - \frac{\pi}{M}\), since that is where RMSE attains its maxima. The probabilities in (23) are:

\[P_r = \frac{1}{M^2} \sec^2 \left(\frac{(2r - 1)\pi}{2M}\right) \tag{32}\]

Observing that \(\sec(\pi - x) = -\sec(x) \implies P_{M+1-r} = P_r\), the expected value is given by

\[\sum_{r=1}^{M} \frac{2\pi(r - 1)}{M} \sec^2 \left(\frac{(2r - 1)\pi}{2M}\right) = \sum_{r=1}^{M/2} \frac{2\pi(M - 1)}{M} \sec^2 \left(\frac{(2r - 1)\pi}{2M}\right) = \pi - \frac{\pi}{M} \tag{33}\]

We now show that the RMSE is \(\sigma \sim \frac{1}{\sqrt{M}}\), to the leading order by evaluating it. The square of the RMSE, by definition, is given by

\[\sigma^2 = \sum_{r=1}^{M} P_r \left(\frac{2\pi}{M}\right)^2 \left(\frac{M + 1}{2} - r\right)^2 \tag{34}\]

The mean is near \(r = \frac{M+1}{2}\). Therefore, changing the variable to \(\frac{M+1}{2} - r\), we obtain a sum starting at the peak of the distribution:

\[\sigma^2 = \frac{8\pi^2}{M^4} \sum_{r=1}^{M} r^2 \csc^2 \left(\frac{r\pi}{M}\right) \tag{35}\]

The leading order in this sum is easily evaluated using an integral:

\[\frac{8\pi^2}{M^4} \sum_{r=1}^{M} r^2 \csc^2 \left(\frac{r\pi}{M}\right) \approx \frac{8}{M\pi} \int_{0}^{\frac{\pi}{2}} x^2 \csc^2(x)dx = \frac{8\log(2)}{M} \tag{36}\]

Therefore, we have shown that,

\[\sigma = \sqrt{\frac{8\log(2)}{M}} \approx \frac{2.35}{\sqrt{M}} \tag{37}\]
E. Q-Metrology

In this strategy, a batch of $\nu$ identical repetitions is used to estimate every decimal place. Estimation with a precision of $1/N$ therefore requires $\log(N)$ batches of $\nu$ particles each. The precision capacity is $M = N^{\log(2\nu+1)}$.

1. Mutual Information

The estimation of each decimal place is independent. Therefore, since mutual information is additive, the total mutual information is the sum of the mutual informations evaluated for each batch. The distribution within a batch is binomial and therefore, we have:

\[
I = \log(N) \left( \frac{1}{2} \log(\nu) \right) \approx \frac{1}{2} \log(M)
\]  
(38)

2. RMSE

The RMSE for the $i^{th}$ decimal place is $\frac{1}{\sqrt{\nu}}$. Therefore, the overall RMSE is given by

\[
\sigma = \sqrt{\sum_{i=1}^{\log(N)} \frac{10^{-2i}}{\nu}} \approx \frac{1}{10\sqrt{\nu}}
\]  
(39)

II. COMPARISON OF THE ERROR BOUNDS WITH THE CRAMER-RAO BOUND: MED AND THE BINOMIAL

For a given random variable $\{z_r\}$ and a given distribution $P_r(x)$, the Cramer-Rao bound sets a lower bound on the error in estimating the parameter $x$. This bound is expressed in terms of the score $S$, defined as $S = \frac{\partial}{\partial x} \log (P_r(x))$. The square-error, $\sigma^2(x)$ is bounded below by the inverse of the Fisher Information, $I(x) = \langle S^2 \rangle$.

One can note that, the Cramer-Rao bound does not provide an absolute bound on the attainable precision in a general estimation; it provides a bound for estimations using a particular distribution. This means, the bound for $P_r(x)$ may be surpassed by using a different distribution $P'_r(x)$ for estimating the same parameter $x$.

One can further note that, the Cramer-Rao bound does not guarantee attainability. Though it sets a lower bound on the error, an estimator which saturates this bound for
the given distribution might not exist. The bound is saturated only for certain distributions. In this section we demonstrate that the binomial distribution and the MED are two such distributions.

(i) Binomial Distribution: The binomial distribution is given by
\[
P_k(x) = \frac{N!}{k!(N-k)!} \left( \frac{1+x}{2} \right)^{N-k} \left( \frac{1-x}{2} \right)^k.
\]
Therefore, the score \( S_k \) is,
\[
S_k(x) = \frac{\partial}{\partial x} \log \left( \frac{N!}{k!(N-k)!} \left( \frac{1+x}{2} \right)^{N-k} \left( \frac{1-x}{2} \right)^k \right) = \frac{N-k}{1+x} - \frac{k}{1-x}
\]
(40)

After evaluating, we have \( \langle S^2 \rangle = \langle \left( \frac{N-k}{1+x} - \frac{k}{1-x} \right)^2 \rangle = \frac{N}{1-x^2} \). Thus, the error bounds are given by \( \sigma \geq \sqrt{\frac{1-x^2}{N}} \), as we have shown, the equality holds in this case.

(ii) MED

The score is given by
\[
S_k(\theta_0) = \begin{cases} 
\frac{1}{x-z_{k-1}} & z_{k-1} \leq x \leq z_k \\
\frac{1}{x_{k+1}-x} & z_k \leq x \leq z_{k+1} \\
0 & z > z_{k+1} \text{ or } x < z_{k-1}
\end{cases}
\]
(41)

Therefore, the Fisher information is \( I_{MED}(x) = \langle S^2 \rangle = \frac{1}{(x-z_{r-1})(z_r-x)} \) for \( z_{r-1} \leq x \leq z_r \).

Thus, the error bound is \( (x - z_{r-1})(z_r - x) \). As seen in the main text, the RMSE of the MED is exactly equal to this expression.

III. MED(UNBIASED) HAS THE LEAST ERROR: DETAILED PROOF

In this section, we provide a detailed proof of the result used in the main text: Among all distributions on a finite set of random variables \( \{z_1 \cdots z_M\} \) with an expected value \( x \), the MED has the least second moment.

Proof:

The MED, \( P_k^{MED}(x) \) is zero at all values of \( k \) except at two consecutive values, \( r \) and \( r+1 \).

Therefore, intuitively, it has the least possible width(Fig. 2 of main text). We prove this result by showing that any other arbitrary distribution \( q_1 \cdots q_M \), with the same expected value \( x \) has a higher second moment than that of the MED.

The second moment, \( u_{MED} \) of the MED is given by \( u_{MED} = x(z_r + z_{r+1}) - z_r z_{r+1} \) where \( x \in [z_r, z_{r+1}] \). We are to show that \( u_q = z_1^2 q_1 + x_1^2 q_1 + \cdots + z_M^2 q_M \) is always larger than
We first define an auxiliary distribution on an auxiliary set of random variables. Let $P_a = q_1 + \cdots + q_r$ and $P_b = q_{r+1} + q_{r+2} + \cdots + q_M$ be a distribution over two random variables $a, b$ defined as

\[
a = \frac{z_1q_1 + \cdots + z_rq_r}{q_1 + \cdots + q_M}
\]

\[
b = \frac{z_{r+1}q_{r+1} + z_{r+2}q_{r+2} + \cdots + z_Mq_M}{q_{r+1} + q_{r+2} + \cdots + q_M}
\]

Clearly, $P_a + P_b = x$. This is a third distribution with the same expected value, and a second moment $u_A = \theta_0(a + b) - ab$. These three distributions are plotted graphically in Fig. 2 of main text.

We complete the proof by showing that $u_q \geq u_A \geq u_{MED}$. The first inequality follows by adding the two cauchy schwarz inequalities:

\[
P_a a^2 \leq z_1^2 q_1 + \cdots z_r^2 q_r \quad \text{&} \quad P_b b^2 \leq z_{r+1}^2 q_{r+1} + \cdots z_M^2 q_M
\]

To show the second inequality, note that $a \leq z_r$ and $b \geq z_{r+1}$ and therefore, $z_r = a + \epsilon$ and $z_{r+1} = b - \delta$ for non negative $\epsilon$ and $\delta$. Using $a \leq x \leq z$, we obtain, $u_A - u_P = (x - a)\delta + (b - x)\epsilon \geq 0$. This completes the proof.

Note that, the equality holds in the second inequality iff $\epsilon = \delta = 0$ or, $x = z_r = a$ or, $x = b = z_{r+1}$. In either cases, the distribution $\{q_1 \cdots q_M\}$ is identical to MED, thus, the MED is unique.