Optimal Choice of Granularity In Commonsense Estimation: Why Half-Orders of Magnitude

Jerry R. Hobbs[1] and Vladik Kreinovich[2]

[1]USC/ISI
4676 Admiralty Way
Marina del Rey, CA 90292
hobbs@isi.edu

[2]Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
vladik@utep.edu

Abstract

It has been observed that when people make crude estimates, they feel comfortable choosing between alternatives which differ by a half-order of magnitude (e.g., were there 100, 300, or 1,000 people in the crowd), and less comfortable making a choice on a more detailed scale, with finer granules, or on a coarser scale (like 100 or 1,000). In this paper, we describe two models of choosing granularity in commonsense estimates, and we show that for both models, in the optimal granularity, the next estimate is 3-4 times larger than the previous one. Thus, these two optimization results explain the commonsense granularity.

1 Introduction

People often need to make crude estimates of a quantity, e.g., estimating the size of a crowd or someone's salary. When people make these crude estimates, they usually feel reasonably comfortable choosing between alternatives which differ by a half order of magnitude (HOM). For example, a person can reasonably estimate whether the size of a crowd was closer to 100, or to 300, or to 1000.

If we ask for an estimate on a more refined scale, e.g., 300 or 350, people will generally be unable to make it. If we ask for an estimate on a coarser scale, e.g., 100 or 1000, people may be able to answer, but they will feel their answer is uninformative.

A particularly striking case of the utility of HOMs is presented by coinage and currency. Most countries have, in addition to denominations for the powers of ten, one or two coins or bills between every two powers of ten. Thus, in the United States, in addition to coins or bills for \$.01, \$.10, \$1.00, \$10.00, and \$100.00, there are also coins or bills in common use for \$.05, \$.25, \$5.00, \$20,00, and \$50.00. These latter provide rough HOM measures for monetary amounts.

It is natural that people should categorize the sizes of physical objects in terms of how they must interact with them. When two objects are roughly of the same size, we manipulate them or navigate about them in roughly the same way. But when one object is about three times larger in linear dimension than another, it must be handled in a different manner. Thus, an orange can be held in one hand, whereas a basketball is more easily held with two, A carton is held in our arms rather than our hands, and carrying a table often requires a second person. For further arguments along these lines, see [1, 2].

These observations lead naturally to the following question: If we are to have a rough logarithmic classification scheme for quantities, what is the *optimal* granularity for commonsense estimates?

There are three requirements we would like the classification scheme to have.

- The categories should be small enough that the types of our interactions with objects are predictable from their category; that HOMs accomplish this is argued above and in Hobbs [1].
- The categories should be large enough that ordinary variation among objects in a class do not usually cross category boundaries; we show that HOMs satisfy this requirement in Section 2.
- The categories should be large enough that aggregation operations have reasonably predictable results; we argue for this claim for HOMs in Section 3.

Thus in this paper we describe two different models for commonsense estimation and show that in both models the optimal granularity is in good accordance with observations about the utility of HOMs. We thus provide a theoretical explanation for the importance of half orders of magnitude in commonsense reasoning.

2 Gaussian Model of Commonsense Estimation

2.1 Main Idea Behind Gaussian Model

We are interested in the situation where we estimate a quantity which can only take non-negative values. To estimate the values of this quantity, we select a sequence of positive numbers $\ldots < e_0 < e_1 < e_2 < \ldots$ (e.g., 1, 3, 10, etc.), and every actual value x of the estimated quantity is then estimated by one of these numbers. Each estimate is approximate: when the estimate is equal to e_i, the actual value x of the estimated quantity may differ from e_i; in other words, there may be an estimation error $\Delta x = e_i - x \neq 0$.

What is the probability distribution of this estimation error? This error is caused by many different factors. It is known that under certain reasonable conditions, an error caused by many different factors is distributed according to Gaussian (normal) distribution (see, e.g., [3]; this fact – called *central limit theorem* – is one of the reasons for the widespread use of Gaussian distribution in science and engineering applications). It is therefore reasonable to assume that Δx is normally distributed.

It is known that a normal distribution is uniquely determined by its two parameters: its average a and its standard deviation σ. Let us denote the average of the error Δx by Δe_i, and its standard deviation by σ_i. Thus, when the estimate is e_i, the actual value $x = e_i - \Delta x$ is distributed according to Gaussian distribution, with an average $e_i - \Delta e_i$ (which we will denote by $\widetilde{e}_i$), and the standard deviation σ_i.

For a Gaussian distribution with given a and σ, the probability density is everywhere positive, so theoretically, we can have values which are as far away from the average a as possible. In practice, however, the probabilities of large deviations from a are so small that the possibility of such deviations can be safely neglected. For example, it is known that the probability of having the value outside the "three sigma" interval $[a-3\sigma, a+3\sigma]$ is $\approx 0.1\%$ and therefore, in most engineering applications, it is assumed that values outside this interval are impossible.

There are some applications where we cannot make this assumption. For example, in designing computer chips, when we have millions of elements on the chip, allowing 0.1% of these elements to malfunction would mean that at any given time, thousands of elements malfunction and thus, the chip would malfunction as well. For such critical applications, we want the probability of deviation to be much smaller than 0.1%, e.g., $\leq 10^{-8}$. Such small probabilities (which practically exclude any possibility of an error) can be guaranteed if we use a "six sigma" interval $[a-6\sigma, a+6\sigma]$. For this interval, the probability for a normally distributed variable to be outside it is indeed $\approx 10^{-8}$.

Within this Gaussian model, what is the optimal granularity?

2.2 Optimal Granularity: Informal Explanation

In accordance with the above idea, for each e_i, if the actual value x is within the "three sigma" range $I_i = [\widetilde{e}_i - 3\sigma_i, \widetilde{e}_i + 3\sigma_i]$, then it is reasonable to take e_i as the corresponding estimate.

We want a granulation which would cover all possible values, so each positive real number must be covered by one of these intervals. In other words, we want the union of all these intervals to coincide with the set of all positive real numbers.

We also want to makes sure that all values that we are covering are indeed non-negative, i.e., that for every i, even the extended "six sigma" interval $[\widetilde{e}_i - 6\sigma_i, \widetilde{e}_i + 6\sigma_i]$ only contains non-negative values.

Finally, since one of the main purposes of granularity is to decrease the number of "labels" that we use to describe different quantities, we want to consider optimal (minimal) sets of intervals. Formally, we can interpret "minimal" in the sense that whichever finite subset we pick, we cannot enlarge their overall coverage by modifying one or several of these intervals. Let us formalize these ideas.

2.3 Optimal Granularity: Formal Description

In the following definitions, we will use the fact that an arbitrary interval $[a^-, a^+]$ can be represented in the Gaussian-type form $[a - 3\sigma, a + 3\sigma]$: it is sufficient to take $a = (a^- + a^+)/2$ and $\sigma = (a^+ - a^-)/6$.

Definition 1.

- *We say that an interval $I = [a - 3\sigma, a + 3\sigma]$ is reliably non-negative if every real number from the interval $[a - 6\sigma, a + 6\sigma]$ is non-negative.*
- *A set $\{I_i\}$, $i = 1, 2, \ldots$, of reliably non-negative intervals I_i is called a granulation if every positive real number belongs to one of the intervals I_i.*
- *We say that a granulation can be improved if, for some finite set $\{i_1, \ldots, i_k\}$, we can replace intervals I_{i_j} with some other intervals I'_{i_j} for which*

$$\bigcup_{j=1}^{k} I_{i_j} \subset \bigcup_{j=1}^{k} I'_{i_j} \qquad \bigcup_{j=1}^{k} I_{i_j} \neq \bigcup_{j=1}^{k} I'_{i_j},$$

  and still get a granulation.
- *A granulation is called optimal if it cannot be improved.*

Theorem 1. *In an optimal granulation, $I_i = [a_i, a_{i+1}]$, where $a_{i+1} = 3a_i$.*

(For reader's convenience, all the proofs are placed in the last section.)

So, half-orders of magnitude are indeed optimal.

3 Uniform Model of Commonsense Estimation

3.1 Motivations

In the Gaussian model, we started with a 3σ bound, and we ended up with a sequence of granules $[a_i, a_{i+1}]$ in which the boundary points a_i form an arithmetic progression: $a_{i+1} = q \cdot a_i$ and $a_i = a_0 \cdot q^i$, with $q = 3$. We could start with a bound of 2.5σ, then we would have got a geometric progression with a different q. Which value of q is indeed optimal?

To find out, let us take into consideration the fact that a granulation is not just for *storing* values, it is also for *processing* these values. Of course, when we replace the actual value by the granule to which it belongs, we lose some information. The idea is to choose the q for which this loss is the smallest.

To estimate the loss, we will consider the simplest data processing operation possible: addition. If we know the exact values of two quantities A and B, then we can compute the exact value of their sum $A + B$. In the granulated case, we do not know the exact values of A and B, we only know the *granules* to which A and B belong, and we want to find out to which of the granules the sum belongs. For example, in the above half-order granulation, we know that the first room has about 10 books, the second about 30, and we want to express the total number of books in the two rooms in similar terms.

The trouble with this problem is that the sum may belong to two different granules. Let us take an example in which we use granules $[1, 3]$, $[3, 9]$, $[9, 27]$, etc. Let us assume that all we know about the first quantity A is that $A \in [1, 3]$, and all we know about the second quantity B is that $B \in [3, 9]$. In this case, the smallest possible values of $A + B$ is $1 + 3 = 4$, and the largest possible value of $A + B$ is $3 + 9 = 12$. In general, the sum $A + B$ can thus take any value from the interval $[4, 12]$. So, it could happen that the sum is in the granule $[3, 9]$, but it could also happen that the sum is in the granule $[9, 27]$.

If we want the granulation to be useful, we must assign a certain granule to the sum $A + B$. Since in reality, the value $A + B$ may belong to two different granules, no matter which of the two granules we assign, there is always a probability that this assignment is erroneous. We would like to select q for which this error probability is the smallest possible.

In order to formulate this question in precise terms, we must describe the corresponding probabilities. A natural way to describe them is as follows: If all we know about A is that A belongs to a granule $\mathbf{a}_i = [a_i, a_{i+1}]$, then it is reasonable to consider all the values from this granule to be equally probable, i.e., to assume that we have a *uniform* distribution on the interval $\mathbf{a}_i = [a_i, a_{i+1}]$. Similarly, If all we know about B is that B belongs to a granule $\mathbf{a}_j = [a_j, a_{j+1}]$, then it is reasonable to consider all the values from this granule to be equally probable, i.e., to assume that we have a *uniform* distribution on the interval $\mathbf{a}_j = [a_j, a_{j+1}]$. Since we have no information about the possible dependence between A and B, it is natural to assume that A and B are independent random

variables. We are now ready for the formal definitions.

3.2 Formalization and the main result

Let $a_0 > 0$ and $q \geq 2$ be real numbers, and let $a_k \stackrel{\text{def}}{=} a_0 \cdot q^k$ and $\mathbf{a}_i \stackrel{\text{def}}{=} [a_i, a_{i+1}]$.

Definition 2. *For every three integers i, j, and k, we can define $P(\mathbf{a}_i + \mathbf{a}_j \in \mathbf{a}_k)$ as the probability that $A_i + A_j \in \mathbf{a}_k$, where A_i is uniformly distributed on the interval $\mathbf{a}_i$, A_j is uniformly distributed on the interval $\mathbf{a}_j$, and A_i and A_j are independent.*

If, as a result of adding $\mathbf{a}_i$ and $\mathbf{a}_j$, we select the granule $\mathbf{a}_k$, then the probability that this assignment is erroneous (i.e., that the actual value of $A_i + A_j$ is not in $\mathbf{a}_k$) is equal to $1 - P(\mathbf{a}_i + \mathbf{a}_j \in \mathbf{a}_k)$. For every i and j, we want to minimize this error, so we select the value k for which this error probability is the smallest:

Definition 3. *For every two integers i and j, we define the sum $\mathbf{a}_i + \mathbf{a}_j$ of granules $\mathbf{a}_i$ and $\mathbf{a}_j$ as a granule $\mathbf{a}_k$ for which the error probability $1 - P(\mathbf{a}_i + \mathbf{a}_j \in \mathbf{a}_k)$ is the smallest possible. The error probability E_{ij} related to this addition is then defined as this smallest probability, i.e., as*

$$E_{ij} \stackrel{\text{def}}{=} \min_k (1 - P(\mathbf{a}_i + \mathbf{a}_j \in \mathbf{a}_k)).$$

Theorem 2.

- *When $q \geq \sqrt{2} + 1 (\approx 2.41)$, then:*
  - $\mathbf{a}_i + \mathbf{a}_i = \mathbf{a}_{i+1}$, *and*
  - $\mathbf{a}_i + \mathbf{a}_j = \mathbf{a}_{\max(i,j)}$ *for $i \neq j$.*
- *When $2 \leq q < \sqrt{2} + 1$, then:*
  - $\mathbf{a}_i + \mathbf{a}_i = \mathbf{a}_{i+1}$;
  - $\mathbf{a}_i + \mathbf{a}_{i+1} = \mathbf{a}_{i+1} + \mathbf{a}_i = \mathbf{a}_{i+2}$; *and*
  - $\mathbf{a}_i + \mathbf{a}_j = \mathbf{a}_{\max(i,j)}$ *for $|i - j| \geq 2$.*

It is worth mentioning that for every q, thus defined addition of granules is commutative but not associative. Indeed:

- For $q \geq \sqrt{2} + 1$, we have:
  - $(\mathbf{a}_0 + \mathbf{a}_0) + \mathbf{a}_1 = \mathbf{a}_1 + \mathbf{a}_1 = \mathbf{a}_2$, while
  - $\mathbf{a}_0 + (\mathbf{a}_0 + \mathbf{a}_1) = \mathbf{a}_0 + \mathbf{a}_1 = \mathbf{a}_1 \neq \mathbf{a}_2$.
- For $q < \sqrt{2} + 1$, we have:

- $(\mathbf{a}_0 + \mathbf{a}_0) + \mathbf{a}_2 = \mathbf{a}_1 + \mathbf{a}_2 = \mathbf{a}_3$, while
- $\mathbf{a}_0 + (\mathbf{a}_0 + \mathbf{a}_2) = \mathbf{a}_0 + \mathbf{a}_2 = \mathbf{a}_2 \neq \mathbf{a}_3$.

Which q is the best? As a measure of quality of a given granulation, it is natural to take the *worst-case* error probability, i.e., the error probability corresponding to the worst-case pair (i, j) (i.e., to the pair with the largest E_{ij}):

Definition 4.

- *By an error probability of a granulation, we mean the value*

$$E(q) \stackrel{\text{def}}{=} \max_{i,j} E_{ij}.$$

- *The granulation with the smallest possible error probability is called optimal.*

Theorem 3. *The granulation is optimal when q satisfies the equation*

$$q^3 - 5q^2 + 4q + 1 = 0,$$

i.e., when $q \approx 3.9$.

In short, here also $q \in [3, 4]$ is the optimal granulation scale.

4 Proofs

4.1 Proof of Theorem 1

1. Let us first prove that for every interval $I_i = [a_i - 3\sigma_i, a_i + 3\sigma_i]$ from an optimal granulation, $a_i = 6\sigma_i$.

Indeed, since all the intervals I_i must be reliably non-negative, we can conclude that $a_i - 6\sigma_i \geq 0$, hence $a_i \geq 6\sigma_i$. So, to complete this part of the proof, it is sufficient to show that we cannot have $a_i > 6\sigma_i$. We will prove this by showing that if $a_i > 6\sigma_i$, then the corresponding granulation can be improved.

Indeed, in this case, we can take $\sigma_i' = a_i/6 > \sigma_i$, and consider a wider interval $I_i' = [a_i - 3\sigma_i', a_i + 3\sigma_i'] \supset I_i$. Due to our choice of σ_i', this new interval is also reliably non-negative. Therefore, if we replace the interval I_i by I_i', we still get a granulation, and $I_i \subset I_i'$, $I_i \neq I_i'$. Thus, the original granulation can be improved.

So, if the granulation is optimal (i.e., cannot be improved), we have $a_i = 6\sigma_i$.

2. Let us now prove that for every interval $I_i = [a_i^-, a_i^+]$ from an optimal granulation, $a_i^+ = 3a_i^-$.

Indeed, from Part 1 of this proof, we can conclude that for an arbitrary interval $I_i = [a_i^-, a_i^+] = [a_i - 3\sigma_i, a_i + 3\sigma_i]$ from the optimal granulation, we have

$3\sigma_i = 0.5 \cdot a_i$, hence $a_i^- = a_i - 3\sigma_i = 0.5 \cdot a_i$ and $a_i^+ = a_i + 3\sigma_i = 1.5 \cdot a_i$. Thus, $a_i^+ = 3a_i^-$.

3. Let us now show that if two intervals from an optimal granulation intersect, then this intersection can only consist of a single point.

To prove this, we will show that if two intervals $I_i = [a_i^-, a_i^+]$ and $I_j = [a_j^-, a_j^+]$ have a more extensive intersection, then the granulation can be improved. Without losing generality, we can assume that $a_i^- \le a_j^-$.

We already know that since both I_i and I_j are intervals from an optimal granulation, we have $a_i^+ = 3a_i^-$ and $a_j^+ = 3a_j^-$. Since $a^- \le a_j^-$, we thus conclude that $a_i^+ = 3a_i^- \le 3a_j^- = a_j^+$.

The fact that the intervals $I_i = [a_i^-, 3a_i^-]$ and $I_j = [a_j^-, 3a_j^-]$ have an intersection means that $a_j^- \le 3a^-$; the fact that this intersection is not simply a single point means that $a_j^- < 3a_i^-$. In this case, $I_i \cup I_j = [a_i^-, 3a_j^-]$.

Let us show that we can improve the granulation if we replace I_i by itself $I_i' = I_i$ and I_j by $I_j' = [3a_i^-, 9a_i^-]$. Indeed, both new intervals are reliably non-negative, and the new union $I_i' \cup I_j' = [a_i^-, 9a_i^-]$ is a strict superset of the old one – because $a_j^- < 3a_i^-$ hence $3a_j^- < 9a_i^-$.

4. So, in an optimal granulation, every interval must be of the type $[a, 3a]$, these intervals must cover the entire real axis, and they cannot intersect in more than one point. Thus, right after each interval $[a_i, 3a_i]$, there should be the next interval $[a_{i+1}, 3a_{i+1}]$, so we should have $a_{i+1} = 3a_i$.

Thus, we get the description from the formulation of the theorem.

5. One can also easily prove that the granulation in which $I_i = [a_i, a_{i+1}]$ with $a_{i+1} = 3a_i$ cannot be improved and is thus optimal. The theorem is proven.

4.2 Proof of Theorem 2

In this proof, we will find the values of $P(\mathbf{a}_i + \mathbf{a}_j \in \mathbf{a}_k)$ for different i, j, and k, and then find, for each i and j, the value k for which the corresponding error $1 - P(\mathbf{a}_i + \mathbf{a}_j \in \mathbf{a}_k)$ is the smallest possible.

1. Let us start by computing the probabilities $P(\mathbf{a}_i + \mathbf{a}_i \in \mathbf{a}_k)$ corresponding to $i = j$.

In this case, both A_i and A_j are uniformly distributed on the interval $[a_0 \cdot q^i, a_0 \cdot q^{i+1}]$. Possible pairs (A_i, A_j) are therefore uniformly distributed on the square $[a_0 \cdot q^i, a_0 \cdot q^{i+1}] \times [a_0 \cdot q^i, a_0 \cdot q^{i+1}]$.

- The smallest possible value of $A_i + A_j$ is $2a_0 \cdot q^i$ which (due to the assumption $q \ge 2$) belongs to the granule $[a_0 \cdot q^i, a_0 \cdot q^{i+1}]$.
- The largest possible value of $A_i + A_j$ is equal to $2a_0 \cdot q^{i+1}$. Due to $q \ge 2$, we have $a_0 \cdot q^{i+1} < 2a_0 \cdot q^{i+1} \le a_0 \cdot q^{i+2}$, hence this value belongs to the next granule $[a_0 \cdot q^{i+1}, a_0 \cdot q^{i+2}]$.

So, the sum $A_i + A_i$ spreads over two different granules: $\mathbf{a}_i$ and $\mathbf{a}_{i+1}$. Thus, of all possible probabilities $P(\mathbf{a}_i + \mathbf{a}_i \in \mathbf{a}_k)$, only the values corresponding to $k = i$ and $k = i + 1$ are different from 0.

Since the pair (A_i, A_j) is uniformly distributed on the square, the probability $P(\mathbf{a}_i + \mathbf{a}_i \in \mathbf{a}_i)$ is equal to the ratio between the area covered by the points for which the sum belongs to $\mathbf{a}_i$ and the area of the entire square. This ratio does not change when we "rescale" the units by dividing both coordinates by a constant $a_0 \cdot q^i$. After this division, the square takes the form $[1, q] \times [1, q]$, and the pairs (x, y) in which we are interested are determined by the inequality $x + y \le q$. Thus, $P(\mathbf{a}_i + \mathbf{a}_i \in \mathbf{a}_i)$ is equal to the ratio S_i/S, where:

- $S = (q-1)^2$ is the area of the square, and
- S_i is the area of the set of all points from this square for which $x + y \le q$.

One can easily see that the set S_i is a right triangle with sides $[1, q-1]$ of length $q - 2$. Thus, the area S_i of this triangle is equal to $(q-2)^2/2$, and the desired probability is equal to

$$P(\mathbf{a}_i + \mathbf{a}_i \in \mathbf{a}_i) = \frac{S_i}{S} = \frac{(q-2)^2}{2(q-1)^2.} \tag{1}$$

Since the sum $A_i + A_i$ can only belong to two possible granules $\mathbf{a}_i$ and $\mathbf{a}_{i+1}$, we have $P(\mathbf{a}_i + \mathbf{a}_i \in \mathbf{a}_i) + P(\mathbf{a}_i + \mathbf{a}_i \in \mathbf{a}_{i+1}) = 1$, so $P(\mathbf{a}_i + \mathbf{a}_i \in \mathbf{a}_{i+1}) = 1 - P(\mathbf{a}_i + \mathbf{a}_i \in \mathbf{a}_i)$. Therefore:

- For $k = i$, the probability of error is equal to $1 - P(\mathbf{a}_i + \mathbf{a}_i \in \mathbf{a}_i)$.
- For $k = i + 1$, the probability of error is equal to $1 - P(\mathbf{a}_i + \mathbf{a}_i \in \mathbf{a}_{i+1}) = P(\mathbf{a}_i + \mathbf{a}_i \in \mathbf{a}_i)$.

The triangle S_i is smaller than the half of the square, hence $P(\mathbf{a}_i + \mathbf{a}_i \in \mathbf{a}_i) < 1/2$, and so $P(\mathbf{a}_i + \mathbf{a}_i \in \mathbf{a}_{i+1}) = 1 - P(\mathbf{a}_i + \mathbf{a}_i \in \mathbf{a}_i) > 1/2 > P(\mathbf{a}_i + \mathbf{a}_i \in \mathbf{a}_i)$. So, for $i = j$, the error $1 - P(\mathbf{a}_i + \mathbf{a}_i \in \mathbf{a}_k)$ is the smallest when $k = i + 1$. For this k, the error probability is equal to

$$E_{ii} = P(\mathbf{a}_i + \mathbf{a}_i \in \mathbf{a}_i) = \frac{(q-2)^2}{2(q-1)^2}. \tag{2}$$

2. Let us now find the values $P(\mathbf{a}_i + \mathbf{a}_j \in \mathbf{a}_k)$ for $i \ne j$.

Without losing generality, it is sufficient to consider only the case when $i < j$. Here:

- $A_i \in [a_0 \cdot q^i, a_0 \cdot q^{i+1}]$,
- $A_j \in [a_0 \cdot q^j, a_0 \cdot q^{j+1}]$, and

- the sum $A_i + A_j$ takes values from $a_0 \cdot (q^i + q^j)$ to $a_0 \cdot (q^{i+1} + q^{j+1})$.

The smallest possible value $a_0 \cdot (q^i + q^j)$ of the sum $A_i + A_j$ is greater than $a_0 \cdot q^j$ but is (due to $i \le j$) smaller than $2a_0 \cdot q^j$ and hence (due to $q \ge 2$) smaller than $a_0 \cdot q^{j+1}$. Thus, this value belongs to the granule $\mathbf{a}_j = [a_0 \cdot q^j, a_0 \cdot q^{j+1}]$.

Similarly, the largest possible value $a_0 \cdot (q^{i+1} + q^{j+1})$ of the sum $A_i + A_j$ belongs to the next granule $\mathbf{a}_{j+1} = [a_0 \cdot q^{j+1}, a_0 \cdot q^{j+2}]$. Thus, out of all possible probabilities $P(\mathbf{a}_i + \mathbf{a}_j \in \mathbf{a}_k)$, only the values corresponding to $k = j$ and $k = j + 1$ are different from 0.

Similarly to the case $i = j$, the probability $P(\mathbf{a}_i + \mathbf{a}_j \in \mathbf{a}_j)$ is equal to the ratio of the two areas. Again, we can "rescale" both coordinates by dividing them by $a_0 \cdot q^i$, after which we get the following simplified geometric representation of the ratio $P(\mathbf{a}_i + \mathbf{a}_j \in \mathbf{a}_j)$ as S_j/S, where:

- S is the total area of the rectangle $[1, q] \times [q^d, q^{d+1}]$ (where $d \stackrel{\text{def}}{=} j - i$), i.e., the value $S = (q-1) \cdot (q^{d+1} - q^d) = q^d \cdot (q-1)^2$; and
- S_j is the total area of the set $\mathcal{S}_j$ of all the points (x, y) from this rectangle for which $x + y \le q^{d+1}$.

The set $\mathcal{S}_j$ is bounded, from above, for $x = 1$, by the point $y = q^{d+1} - 1$, and for $x = q$, by the point $y = q^{d+1} - q$. Thus, this set $\mathcal{S}_j$ consists of two subsets:

- a rectangle $[1, q] \times [q^d, q^{d+1} - q]$ of area $(q-1) \cdot (q^{d+1} - q^d - q)$ and
- a right triangle with sides $[1, q]$ and $[q^{d+1} - 1, q^{d+1} - q]$ of lengths equal to $q - 1$ – whose area is equal to $(q-1)^2/2$.

Hence,

$$S_j = (q-1) \cdot (q^{d+1} - q^d - q) + \frac{(q-1)^2}{2} = q^q \cdot (q-1)^2 - (q-1) \cdot \frac{q+1}{2}. \quad (3)$$

3. We are now going to show that for $d \ge 2$, the area S_j is larger than a half of S; then, similarly to the case $i = j$, we will be able to conclude that $\mathbf{a}_i + \mathbf{a}_j = \mathbf{a}_j$.

To prove this, it is sufficient to prove that the area of the complement $\mathcal{S}_{j+1}$ to $\mathcal{S}_j$ is not larger than one half of area of the the entire rectangle. The area S_{j+1} of this complement is equal to

$$S_{j+1} = (q-1) \cdot \frac{q+1}{2},$$

so this area does not exceed the half of $S = q^d \cdot (q-1)^2$ if and only if

$$(q-1) \cdot \frac{q+1}{2} \le q^d \cdot \frac{(q-1)^2}{2}.$$

Dividing both sides of this inequality by $(q-1)/2$, we get a (simplified) equivalent inequality

$$q + 1 \le q^d \cdot (q - 1). \tag{4}$$

Since $q \ge 2$ and $d \ge 2$, we have $1 \le q^{d-1}$, and $q \le q^{d-1}$, hence $1 + q \le 2q^{d-1} \le q \cdot q^{d-1} = q^d$. We also have $q - 1 \ge 1$, hence $q + 1 \le q^d \le q^d \cdot (q - 1)$, i.e., we have the desired inequality (4). Thus, for $|j - i| = d \ge 2$, we have $k = \max(i, j)$. For this case,

$$E_{ij} = \frac{S_{j+1}}{S} = \frac{q+1}{2q^d \cdot (q-1)}. \tag{5}$$

4. For $j - i = 1$, the inequality $S_{j+1} \le S/2$ is equivalent to

$$(q-1) \cdot \frac{q+1}{2} \le q \cdot \frac{(q-1)^2}{2},$$

i.e., to $q + 1 \le q^2 - q$, or to $q^2 - 2q - 1 \ge 0$. The corresponding equation $q^2 - 2q - 1 = 0$ has two roots $q = 1 \pm \sqrt{2}$, so for $q \ge 2$, the desired inequality is equivalent to $q \ge \sqrt{2} + 1$. So:

- For $q \ge \sqrt{2}+1$, the sum $\mathbf{a}_i + \mathbf{a}_{i+1}$ is equal to $\mathbf{a}_{i+1}$, and the corresponding probability error is equal to

$$E_{i,i+1} = \frac{q+1}{2q \cdot (q-1)}. \tag{6}$$

- For $q < \sqrt{2}+1$, the sum $\mathbf{a}_i + \mathbf{a}_{i+1}$ is equal to $\mathbf{a}_{i+2}$, and the corresponding probability error is equal to

$$E_{i,i+1} = 1 - \frac{q+1}{2q \cdot (q-1)}. \tag{7}$$

The theorem is proven.

4.3 Proof of Theorem 3

In the proof of Theorem 2, we have already found the error probabilities E_{ij} corresponding to different values i, j, and q. Let us use these probabilities to find the optimal q. Since the formulas for E_{ij} were different for $q \ge \sqrt{2}+1$ and for $q < \sqrt{2}+1$, we will consider these two cases separately.

Let us start with $q \ge \sqrt{2}+1$. In this case, due to the formulas (2), (5), and (6), the error probability is equal to:

$$E(q) = \max\left(\frac{(q-2)^2}{2(q-1)^2}, \frac{q+1}{2q \cdot (q-1)}, \frac{q+1}{2q^2 \cdot (q-1)}, \ldots, \frac{q+1}{2q^d \cdot (q-1)}, \ldots\right). \tag{8}$$

Since $q \geq 1$, we have

$$\frac{q+1}{2q \cdot (q-1)} \geq \frac{q+1}{2q^2 \cdot (q-1)} \geq \cdots \geq \frac{q+1}{2q^d \cdot (q-1)} \geq \cdots,$$

and thus, to find the maximum, it is sufficient to consider only the first two terms:

$$E(q) = \max(E_1(q), E_2(q)), \tag{9a}$$

where

$$E_1(q) \stackrel{\text{def}}{=} \frac{(q-2)^2}{2(q-1)^2}; \tag{9b}$$

$$E_2(q) \stackrel{\text{def}}{=} \frac{q+1}{2q \cdot (q-1)}. \tag{9c}$$

Let us show that $E_1(q)$ is an increasing function of q, while $E_2(q)$ is decreasing.

- The expression $E_1(q)$ can be represented as

  $$E_1(q) = \frac{1}{2} \cdot \left(1 - \frac{1}{q-1}\right)^2.$$

  When q increases ($q \uparrow$), $q-1$ also increases ($q-1 \uparrow$), hence

  $$\frac{1}{q-1} \downarrow, \quad 1 - \frac{1}{q-1} \uparrow, \quad E_1(q) = \frac{1}{2} \cdot \left(1 - \frac{1}{q-1}\right)^2 \uparrow.$$

- Similarly, if we divide both numerator and denominator of the expression $E_2(q)$ by q, we get the equivalent expression

  $$E_2(q) = \frac{1 + 1/q}{2(q-1)}.$$

  When $q \uparrow$, we have $1/q \downarrow$, hence $1 + 1/q \downarrow$. On the other hand, we have $2(q-1) \uparrow$, so $E_2(q)$ is the result of dividing a decreasing function by an increasing one – hence itself decreasing.

Since $E_1(q) \uparrow$ and $E_2(q) \downarrow$, we have $E_1(q) - E_2(q) \uparrow$.

- For $q = \sqrt{2} + 1$, we have

  $$E_1(q) = \frac{3}{4} - \frac{1}{2} \cdot \sqrt{2}, \quad E_2(q) = \frac{3}{2} + \sqrt{2},$$

  so $E_1(q) - E_2(q) < 0$.

- When $q \to \infty$, we have $E_1(q) \to 1/2$, while $E_2(q) \to 0$, hence for sufficiently large q, we get $E_1(q) - E_2(q) > 0$.

A continuous increasing function $E_1(q) - E_2(q)$ goes from a negative to a positive value, so it must attain the value 0 at some point q_0. Let us show that on the set $[\sqrt{2}+1, \infty)$, the minimum of $E(q)$ is attained for $q = q_0$. Indeed:

- For $q = q_0$, we have $E_1(q_0) = E_2(q_0)$ hence
$$E(q_0) = \max(E_1(q_0), E_2(q_0)) = E_1(q_0) = E_2(q_0).$$
- For $q < q_0$, we have $E_1(q) < E_2(q)$, hence
$$E(q) = \max(E_1(q), E_2(q)) = E_2(q).$$
Since the function $E_2(q)$ is decreasing and $q < q_0$, we have $E_2(q) > E_2(q_0)$, so $E(q) > E(q_0)$.
- Similarly, for $q < q_0$, we have $E_1(q) > E_2(q)$, hence
$$E(q) = \max(E_1(q), E_2(q)) = E_1(q).$$
Since the function $E_1(q)$ is increasing and $q > q_0$, we have $E_1(q) > E_1(q_0)$, so $E(q) > E(q_0)$.

Thus, among all values $q \in [\sqrt{2}+1, \infty]$, the smallest possible value $E(q)$ is attained when $E_1(q) = E_2(q)$, i.e., when

$$\frac{(q-2)^2}{2(q-1)^2} = \frac{q+1}{2q \cdot (q-1)}.$$

Multiplying both sides of this equation by $2q \cdot (q-1)^2$, we get the equivalent equation $q \cdot (q-2)^2 = q^2 - 1$. Opening the parentheses and moving all the terms to one side, we get the desired equation from the formulation of Theorem 3.

For this value q, we have $E(q) \approx 0.21$. To complete the proof, we must show that for $q \in [2, \sqrt{2}+1]$, the value of $E(q)$ is larger than 0.21. Indeed, for these q, we have

$$E(q) = \max\left(\frac{(q-2)^2}{2(q-1)^2}, 1 - \frac{q+1}{2q \cdot (q-1)}, \frac{q+1}{2q^2 \cdot (q-1)}, \ldots, \frac{q+1}{2q^d \cdot (q-1)}, \ldots\right),$$

hence

$$E(q) \geq 1 - \frac{q+1}{2q \cdot (q-1)}. \tag{10}$$

We already know that when $q \uparrow$, we have

$$\frac{q+1}{2q \cdot (q-1)} \downarrow, \text{ hence } 1 - \frac{q+1}{2q \cdot (q-1)} \uparrow$$

thence, for $q \geq 2$:

$$1 - \frac{q+1}{2q \cdot (q-1)} \geq 1 - \frac{2+1}{2 \cdot 2 \cdot (2-1)} = 0.25.$$

So, from (10), we conclude that $E(q) \geq 0.25 > 0.21$. The theorem is proven.

Conclusions

When people make crude estimates, they feel comfortable choosing between alternatives which differ by a half-order of magnitude (e.g., were there 100, 300, or 1,000 people in the crowd), and less comfortable making a choice on a more detailed scale (like 100 or 110 or 120) or on a coarser scale (like 100 or 1,000). We have shown that for two natural models of choosing granularity in commonsense estimates, in the optimal granularity, the next estimate is 3-4 times larger than the previous one. Thus, we have explained the commonsense HOM granularity.

Acknowledgments

This work was partially supported by NASA under cooperative agreement NCC5-209, by the United Space Alliance, grant No. NAS 9-20000 (PWO C0C67713A6), by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grants number F49620-95-1-0518 and F49620-00-1-0365, by Grant No. W-00016 from the U.S.-Czech Science and Technology Joint Fund, by the National Science Foundation under Grants Number IRI-9619126 (Multimodal Access to Spatial Data), EAR-0225670, and DMS-0532645, by the Defense Advanced Research Projects Agency under Air Force Research Laboratory contract F30602-00-C-0168, by the Army Research Lab grant DATM-05-02-C-0046, by the Star Award from the University of Texas System, and by the Texas Department of Transportation grant No. 0-5453.

References

[1] J. R. Hobbs, "Half orders of magnitude", In: L. Obrst and I. Mani (eds.), *Proceeding of the Workshop on Semantic Approximation, Granularity, and Vagueness, A Workshop of the Seventh International Conference on Principles of Knowledge Representation and Reasoning KR'2000*, Breckenridge, Colorado, April 11, 2000, pp. 28–38.

[2] J. R. Hobbs and V. Kreinovich, "Optimal Choice of Granularity In Commonsense Estimation: Why Half-Orders of Magnitude", *Proceedings of the Joint 9th World Congress of the International Fuzzy Systems Association and 20th International Conference of the North American Fuzzy Information Processing Society IFSA/NAFIPS 2001*, Vancouver, Canada, July 25–28, 2001, pp. 1343–1348.

[3] H. M. Wadsworth, Jr. (editor), *Handbook of statistical methods for engineers and scientists*, McGraw-Hill Publishing Co., N.Y., 1990.