Why Superellipsoids: A Probability-Based Explanation

Pedro Barragan Olague and Vladik Kreinovich

Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
pabarraganolague@miners.utep.edu
vladik@utep.edu
1. Outline

- In many practical situations, possible values of the deviation vector form (approximately) a super-ellipsoid.
- In this talk, we provide a theoretical explanation for this empirical fact.
- This explanation is based on the natural notion of scale-invariance.
2. Need to Describe Uncertainty Domains

- The intent of mass production is to produce gadgets with same values \((x_1, \ldots, x_n)\) of the characteristics \(x_i\).

- In reality, different gadgets have slightly different values \(\tilde{x}_i\) of these characteristics: \(\Delta x_i \overset{\text{def}}{=} \tilde{x}_i - x_i \neq 0\).

- For each of these characteristics \(x_i\), we usually have a tolerance bound \(\Delta_i\) for which \(|\Delta x_i| \leq \Delta_i\).

- Possible values of \(\Delta x_i\) form an interval \([-\Delta_i, \Delta_i]\).

- Thus, possible values of the deviation vector \(\Delta x = (\Delta x_1, \ldots, \Delta x_n)\) are in the box

  \([-\Delta_1, \Delta_1] \times \ldots \times [-\Delta_n, \Delta_n]\).

- In practice, not all \(\Delta x\) from this box are possible.

- It is desirable to describe the set \(S\) of all possible deviation vectors \(\Delta x\); \(S\) is called uncertainty domain.
3. Shall Not We Also Determine Probabilities?

- At first glance, it seems that we should be interested:
  - not only in finding out which deviation vectors $\Delta x$ are possible and which are not,
  - but also in how frequent different possible vectors are.
- In other words, we should be interested in the probability distribution on this domain.
- In reality, however, it is not possible to find these probabilities.
- Indeed, the manufacturing process may slightly change (and often does change).
- After each such change, the tolerance intervals and the uncertainty domain remain largely unchanged.
- However, the probabilities change (often drastically).
4. **Empirical Shapes of Uncertainty Domains**

- In many practical cases, the uncertainty domain can be well approximated by a *super-ellipsoid*:
  \[
  \sum_{i=1}^{n} \left( \frac{|\Delta x_i|}{\sigma_i} \right)^p \leq C.
  \]

- Their approximation accuracy is higher than for other families with the same number of parameters.

- Super-ellipsoids are also actively used in image processing, to describe different components of an image.

- In this talk, we provide a theoretical explanation for this empirical phenomenon.
5. First Idea: Probabilistic Approach

- In reality, there is some probability distribution $\rho_i(\Delta x_i)$ for each of the random variables $\Delta x_i$.
- We have no reason to assume that positive or negative values of $\Delta x_i$ are more probable.
- So, it makes sense to assume that they are equally probable.
- So, each distribution $\rho_i(\Delta x_i)$ is symmetric: $\rho_i(\Delta x_i) = \rho_i(|\Delta x_i|)$.
- Similarly, we have no reasons to believe that different deviations are statistically dependent.
- So, it makes sense to assume that random variables $\Delta x_i$ are independent: $\rho(\Delta x) = \prod_{i=1}^{n} \rho_i(|\Delta x_i|)$.
- Usually, we consider a deviation vector possible if its probability exceed some $t$: $S_t \overset{\text{def}}{=} \{ \Delta x : \rho(\Delta x) \geq t \}$. 
6. Second Idea: Scale Invariance

- Numerical values of the deviations $\Delta x_i$ depend on the choice of a measuring unit.
- If we replace the original unit by a $\lambda$ times smaller one, we get new numerical values $\Delta x'_i = \lambda \cdot \Delta x_i$.
- Since the physics remains the same, it makes sense to require that the uncertainty domains do not change.
- To be more precise, the pdf threshold $t$ may change, but the family of such sets should remain unchanged.
- So, we require that $\{S'_t\}_t = \{S_t\}_t$, where $S'_t$ corresponds to the re-scaled pdf $\rho'(\Delta x) = \text{const} \cdot \rho(\lambda \cdot \Delta)$.
- We will prove that under this scale-invariance, the corresponding sets $S_t$ are exactly super-ellipsoids.
- Thus, we will get the desired explanation.
7. Definitions and the Main Result

- Let $n > 1$, and let $\rho(y) = (\rho_1(y_1), \ldots, \rho_n(y_n))$ be a tuple of positive symmetric smooth functions.

- For every $t > 0$, let us denote
  \[ S_t(\rho) \overset{\text{def}}{=} \left\{ (y_1, \ldots, y_n) : \prod_{i=1}^n \rho_i(y_i) \geq t \right\}. \]

- We say that a tuple $\rho(y)$ is bounded if the set $S_t(\rho)$ is bounded for every $t$.

- For every $\lambda > 0$, by a $\lambda$-re-scaling of the tuple $\rho(x)$, we mean a tuple $\rho_\lambda(y)$, for which $\rho_{\lambda,i}(y_i) \overset{\text{def}}{=} \frac{1}{\lambda} \cdot \rho_i(\lambda \cdot y_i)$.

- We say that $\rho(y)$ is scale-invariant if for every $\lambda > 0$, re-scaling does not change $\{S_t\}_t$: $\{S_t(\rho)\}_t = \{S_t(\rho_\lambda)\}_t$.

- **Main Result.** *If the tuple $\rho(y)$ is bounded and scale-invariant, then each set $S_t(\rho)$ is a super-ellipsoid.*
8. Discussion

- Vice versa, it is easy to prove that:
  - each super-ellipsoid
    \[
    \left\{ y : \sum_{i=1}^{n} \left(\frac{|y_i|}{\sigma_i}\right)^p \leq C \right\}
    \]
  - can be represented as a set $S_t$ for some bounded and scale-invariant distributions.
  - Namely, we can take $\rho_i(y_i) = \text{const} \cdot \exp \left( -\frac{|y_i|^p}{\sigma_i^p} \right)$.
  - Such probability distributions indeed occur: e.g., as probability distributions of measuring errors.
9. Proof

- For convenience, let us consider logarithms
  \[ \psi_i(y_i) \overset{\text{def}}{=} - \log(\rho_i(y_i)) \].

- Let us take the negative logarithm of both sides of the inequality \( \prod_{i=1}^{n} \rho_i(y_i) \geq t \) that describes the set \( S_t(\rho) \).

- We then get an equivalent description \( \sum_{i=1}^{n} \psi_i(y_i) \leq c \), where we denoted \( c \overset{\text{def}}{=} - \log(t) \).

- In these terms, scale-invariance means that the corresponding family of sets is the same for all \( c \).

- In terms of the new functions \( \psi_i(y_i) \), scaling means
  \[ \psi_{\lambda,i}(y_i) = - \ln(\rho_{\lambda,i}(y_i)) = - \log \left( \frac{1}{\lambda} \cdot \rho_i(\lambda \cdot y_i) \right) = \log(\lambda) - \log(\rho_i(\lambda \cdot y_i)) = \psi_i(\lambda \cdot y_i) + \log(\lambda) . \]
10. Proof (cont-d)

- So, scaling has the form \( \psi_{\lambda,i}(y_i) = \psi_i(\lambda \cdot y_i) + \log(\lambda) \).
- In these terms, the fact that scaling does not change the family of sets \( S_t \) implies that
  - if two tuples \((y_1, \ldots, y_n)\) and \((z_1, \ldots, z_n)\) always belong or not belong to the same sets \( S_t \),
  - i.e., if \( \sum_{i=1}^{n} \psi_i(y_i) = \sum_{i=1}^{n} \psi_i(z_i) \),
  - then the re-scaled functions should also have the same value of the sum: \( \sum_{i=1}^{n} \psi_{\lambda,i}(y_i) = \sum_{i=1}^{n} \psi_{\lambda,i}(z_i) \).
- Substituting \( \psi_{\lambda,i}(y_i) \) into this formula, we get
  \[
  \sum_{i=1}^{n} (\psi_i(\lambda \cdot y_i) + \log(\lambda)) = \sum_{i=1}^{n} (\psi_i(\lambda \cdot z_i) + \log(\lambda)), \text{ hence}
  \]
  \[
  \sum_{i=1}^{n} \psi_i(\lambda \cdot y_i) = \sum_{i=1}^{n} \psi_i(\lambda \cdot z_i).
  \]
11. Proof (cont-d)

- Thus, we have the following property:
  - if \( \sum_{i=1}^{n} \psi_i(y_i) = \sum_{i=1}^{n} \psi_i(z_i) \),
  - then \( \sum_{i=1}^{n} \psi_i(\lambda \cdot y_i) = \sum_{i=1}^{n} \psi_i(\lambda \cdot z_i) \).

- In particular, this property holds if we perform very small changes to only two \( y_i \)'s:
  \[
  y_a \rightarrow z_a = y_a + \delta_a, \quad y_b \rightarrow z_b = y_b + \delta_b.
  \]

- Here, \( \psi_a(y_a + \delta_a) = \psi_a(y_a) + \psi'_a(y_a) \cdot \delta_a + o(\delta) \).

- Similarly, \( \psi_b(y_b + \delta_b) = \psi_b(y_b) + \psi'_b(y_b) \cdot \delta_b + o(\delta) \).

- Thus, \( \sum_{i=1}^{n} \psi_i(z_i) = \sum_{i=1}^{n} \psi_i(y_i) + \psi'_a(y_a) \cdot \delta_a + \psi'_b(y_b) \cdot \delta_b + o(\delta) \).

- So, the original equality \( \sum_{i=1}^{n} \psi_i(y_i) = \sum_{i=1}^{n} \psi_i(z_i) \) takes the form
  \[
  \psi'_a(y_a) \cdot \delta_a + \psi'_b(y_b) \cdot \delta_b + o(\delta) = 0.
  \]
12. Proof (cont-d)

- Similarly, the equality \( \sum_{i=1}^{n} \psi_i(\lambda \cdot y_i) = \sum_{i=1}^{n} \psi_i(\lambda \cdot z_i) \) takes the form \( \psi'_a(\lambda \cdot y_a) \cdot \delta_a + \psi'_b(\lambda \cdot y_b) \cdot \delta_b + o(\delta) = 0 \).

- So, the scale-invariance condition takes the form:
  - if \( \psi'_a(y_a) \cdot \delta_a + \psi'_b(y_b) \cdot \delta_b + o(\delta) = 0 \),
  - then \( \psi'_a(\lambda \cdot y_a) \cdot \delta_a + \psi'_b(\lambda \cdot y_b) \cdot \delta_b + o(\delta) = 0 \).

- The 1st condition \( \iff -\frac{\delta_b}{\delta_a} = \frac{\psi'_a(y_a)}{\psi'_b(y_b)} + o(\delta) \).

- The 2nd condition \( \iff -\frac{\delta_b}{\delta_a} = \frac{\psi'_a(\lambda \cdot y_a)}{\psi'_b(\lambda \cdot y_b)} + o(\delta) \).

- So, \( \frac{\psi'_a(\lambda \cdot y_a)}{\psi'_b(\lambda \cdot y_b)} = \frac{\psi'_a(y_a)}{\psi'_b(y_b)} \Rightarrow \frac{\psi'_a(\lambda \cdot y_a)}{\psi'_a(y_a)} = \frac{\psi'_b(\lambda \cdot y_b)}{\psi'_b(y_b)} \).

- The left-hand side of this equality doesn’t depend on \( y_b \); thus, the right-hand side doesn’t depend on \( y_b \).
13. Proof (cont-d)

• Hence, this ratio depends only on $\lambda$. Let us denote this common ratio by $r(\lambda)$: $\psi'_a(\lambda \cdot y_a) = r(\lambda) \cdot \psi'_a(y_a)$.

• The derivative of a smooth function is always measurable.

• Thus, the function $r(\lambda)$ is also measurable, as a ratio of two measurable functions.

• Now, let us take arbitrary values $\lambda_1 > 0$ and $\lambda_2 > 0$.

• Then, we can re-scale first by $\lambda_1$, then by $\lambda_1$, or we can right away re-scale by $\lambda = \lambda_1 \cdot \lambda_2$.

• In the first case,

$$\psi'(\lambda_1 \cdot \lambda_2 \cdot y_a) = r(\lambda_1) \cdot \psi'(\lambda_2 \cdot y_a) = r(\lambda_1) \cdot r(\lambda_2) \cdot \psi'_a(y_a).$$

• In the 2nd case, $\psi'(\lambda_1 \cdot \lambda_2 \cdot y_a) = r(\lambda_1 \lambda_2) \cdot \psi'_a(y_a)$.

• So, we must have $r(\lambda_1 \cdot \lambda_2) = r(\lambda_1) \cdot r(\lambda_2)$. 
14. Proof (cont-d)

- It is known that all measurable functions satisfying this property have the form \( r(\lambda) = \lambda^\beta \) for some \( \beta \).
- So, \( \psi'_a(\lambda \cdot y_a) = r(\lambda) \cdot \psi'_a(y_a) = \lambda^\beta \cdot \psi'_a(y_a) \).
- For \( \lambda = z \) and \( y_a = 1 \), we get \( \psi'_a(z) = \psi'_a(1) \cdot z^\beta \), i.e., that \( \psi'_a(y_a) = c_a \cdot y_a^\beta \) for some \( c_a \).
- Integrating, for \( \beta \neq -1 \), for \( y_a > 0 \), we get \( \psi_a(y_a) = k_a \cdot y_a^p + C_a \) for \( p = \beta + 1 \), \( k_a \overset{\text{def}}{=} \frac{c_a}{\beta + 1} \).
- Since \( \psi_i(y_i) \) is even, we get \( \psi_i(y_i) = k_i \cdot |y_i|^p + C_i \).
- So, the condition \( \sum_{i=1}^{n} \psi(y_i) \leq c \) takes the super-ellipsoid form \( \sum_{i=1}^{n} k_i \cdot |y_i|^p \leq c_0 \overset{\text{def}}{=} c - \sum_{i=1}^{n} C_i \).
- For this super-ellipsoid to be bounded, we need to have \( p > 0 \).
15. **Proof (final)**

- To complete the proof, it is sufficient to consider the case when $\beta = -1$.

- For $\beta = -1$, integration leads to
  \[ \psi_i(y_i) = k_i \cdot \ln(|y_i|) + C_i. \]

- So the condition $\sum_{i=1}^{n} \psi_i(y_i) \leq c$ takes the form
  \[ \sum_{i=1}^{n} k_i \cdot \ln(|y_i|) \leq c_0 \overset{\text{def}}{=} c - \sum_{i=1}^{n} C_i. \]

- Exponentiating both sides, we get $\prod_{i=1}^{n} |y_i|^{k_i} \leq \exp(C)$, for which the corresponding set $S_t$ is unbounded.

- So, in the bounded cases, we always have a super-ellipsoid. The result is proven.
16. Acknowledgments

This work was supported in part:

- by the National Science Foundation grants:
  - HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and
  - DUE-0926721, and
- by an award from Prudential Foundation.