Understanding Gaussian Filtering and KDE-Based Parameter Selection

Jeff Tuche :)

1 Understanding the Graph Components

Each density plot consists of the following elements:

- **Blue Curve (”All Robust”)**: KDE for all parameter sets that passed the Random Forest (RF) filtering.

- **Orange Curve (”Stable”)**: KDE for the subset of parameters that passed the Gaussian KDE filtering.

- **Red Dashed Line (Median)**: Median of the KDE-filtered stable parameters.

- **Red Shaded Area (Interquartile Range, Q25–Q75)**: The range where 50% of the KDE-filtered stable parameters fall.

2 Understanding the Median and the Red Shaded Area

2.1 How the Median is Computed

The median represents the central value in the KDE-filtered stable parameters. This means:

- It does not necessarily fall in the center of the red shaded area (Q25–Q75).

- The median is the value where half of the KDE-filtered stable parameters are below, and half are above.

- If the distribution is **skewed**, the median can shift towards the longer tail.

2.2 Understanding the Red Shaded Area (Interquartile Range)

- The red shaded area represents the **Interquartile Range (IQR)**, which contains the middle 50% of the stable parameters.

- Q_{25} (1st quartile): 25% of the KDE-filtered stable parameters are below this value.

- Q_{75} (3rd quartile): 75% of the KDE-filtered stable parameters are below this value.

- The width of the red area indicates the spread of the stable parameter values.

3 Why Might the Median Not Be in the Middle of the Red Area?

If the KDE stable distribution is not symmetrical, the median can be off-center due to:

1. **Skewness**: If the KDE stable distribution is skewed left or right, the median shifts toward the denser side.

2. **Multi-Modality**: If there are two peaks (bimodal distribution), the median may fall closer to one peak rather than the center of the full distribution.

3. **Data Imbalance**: If there are fewer points in one tail of the stable KDE distribution, the median will lean toward the denser region.

4 Examples to Illustrate

4.1 Scenario 1: Symmetric Distribution

- KDE Stable Curve: Looks like a perfect bell shape (normal distribution).

- Median: Right in the middle of Q_{25}–Q_{75}.

- Interpretation: The distribution is balanced, and the most stable values are well-centered.

4.2 Scenario 2: Skewed Distribution (Right-Skewed)

- KDE Stable Curve: Peaks on the left and has a long tail to the right.

- Median: Closer to the peak on the left, rather than the center of Q_{25}–Q_{75}.

- Interpretation: The most stable values are concentrated in a smaller range, but there are still some outliers pulling the Q_{75} boundary.

4.3 Scenario 3: Multi-Peak Distribution

- KDE Stable Curve: Has two peaks (bimodal).

- Median: Falls between the two peaks, not necessarily in the center of Q_{25}–Q_{75}.

- Interpretation: There are two possible stable regions, but they are separated.

5 How to Use This Information?

- If the **median is far from the red area**, check why the distribution is skewed.

- If the **red area is wide**, the stable parameters vary a lot, meaning you might need stricter filtering.

- If the **median is on one side of the red area**, focus on the more concentrated region of stability.

6 Final Check: Is the Median Always Inside the Red Shaded Area?

To confirm:

- Is your median always inside the red shaded area, or do you have cases where it is completely outside?

- If it's outside, investigate how KDE-filtered parameters are distributed and adjust filtering thresholds if needed.