DEALING WITH UNCERTAINTY

From the GEOPRIV motivational series

Drafts

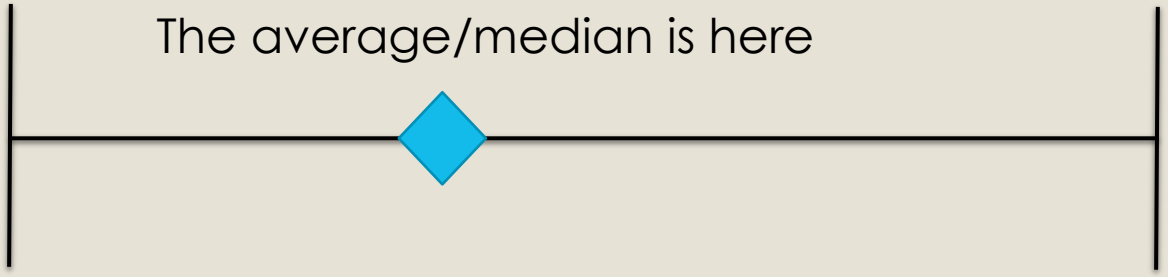
- draft-thomson-geopriv-uncertainty-08
  - What uncertainty (and confidence) mean and are good for
  - A bunch of shortcuts for dealing with uncertainty
  - Intended status: Informational
- draft-thomson-geopriv-confidence-04
  - A small addition to PIDF-LO
  - Intended status: Proposed Standard

UNCERTAINTY

draft-thomson-geopriv-uncertainty-08

Statistics Refresher: Confidence Intervals

- It's common to describe measurements of stochastic processes (i.e., random $#!^) as confidence intervals
- Graphs with the following are very common in scientific literature:



95% of the time, it's between here…

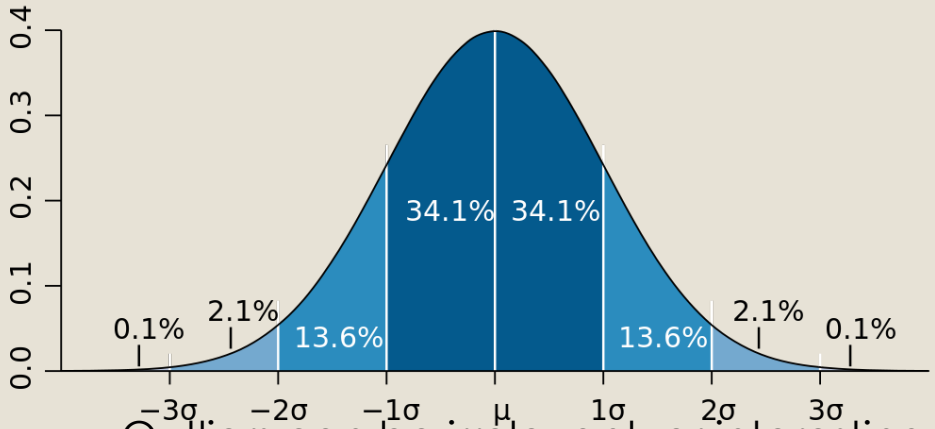
and here.

Terminology is surprisingly important

- Accuracy =
  - fuzzy, feel good term
  - qualitative, no numbers, use it when talking in the abstract
- Uncertainty =
  - quantitative, concrete, supported with numbers
  - useless without confidence
- Confidence =
  - probabilistic measure for uncertainty
  - quantitative, concrete, has numbers [0, 1) or [0, 100%)
- Combine uncertainty and confidence:
  - 95% of the time (confidence), the value is between X and Y (uncertainty range)

Lies, damned lies, and…

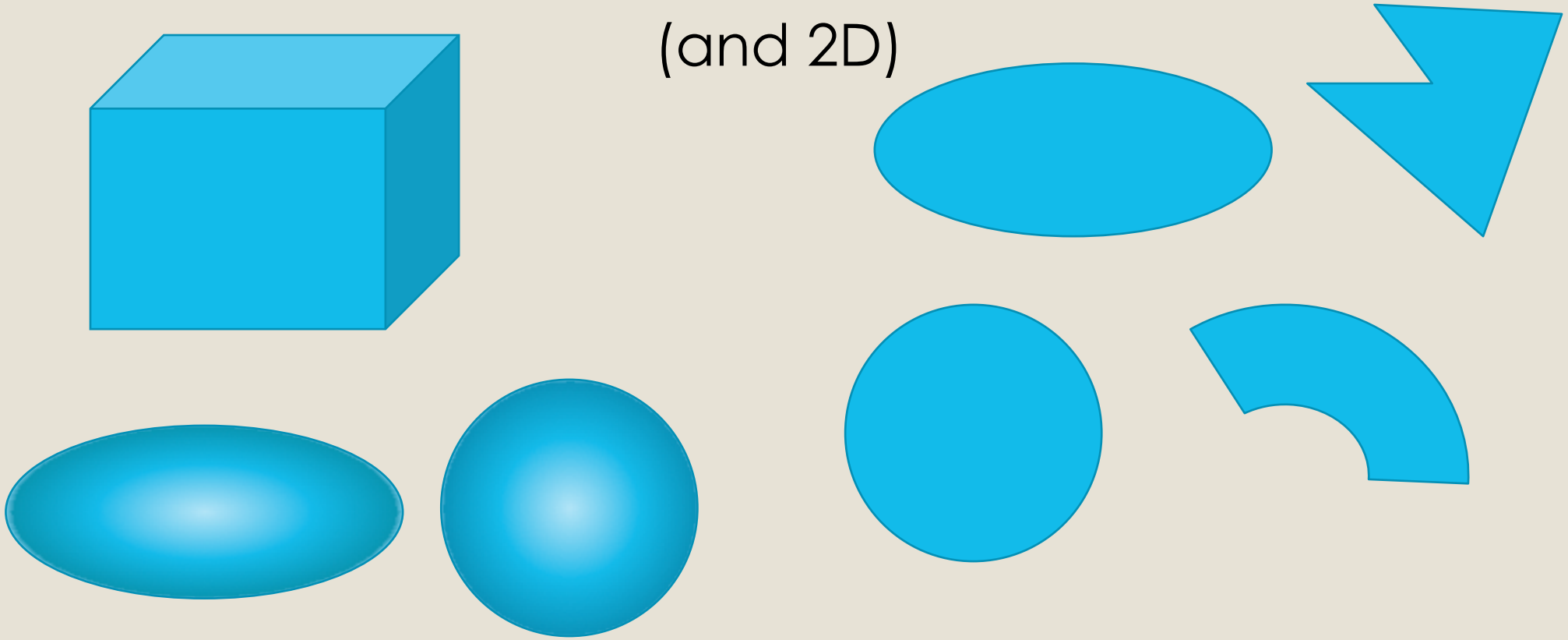
- The error bars hide a lot of details
  - The observed probability distribution is rarely perfectly normal



  - Outliers can be irrelevant, or interesting, but they disappear
  - Other interesting points like mean, median, variance, all go
- But that's OK, because it's hard to process more detailed information

RFC 5491 defines error bars in 3D

(and 2D)

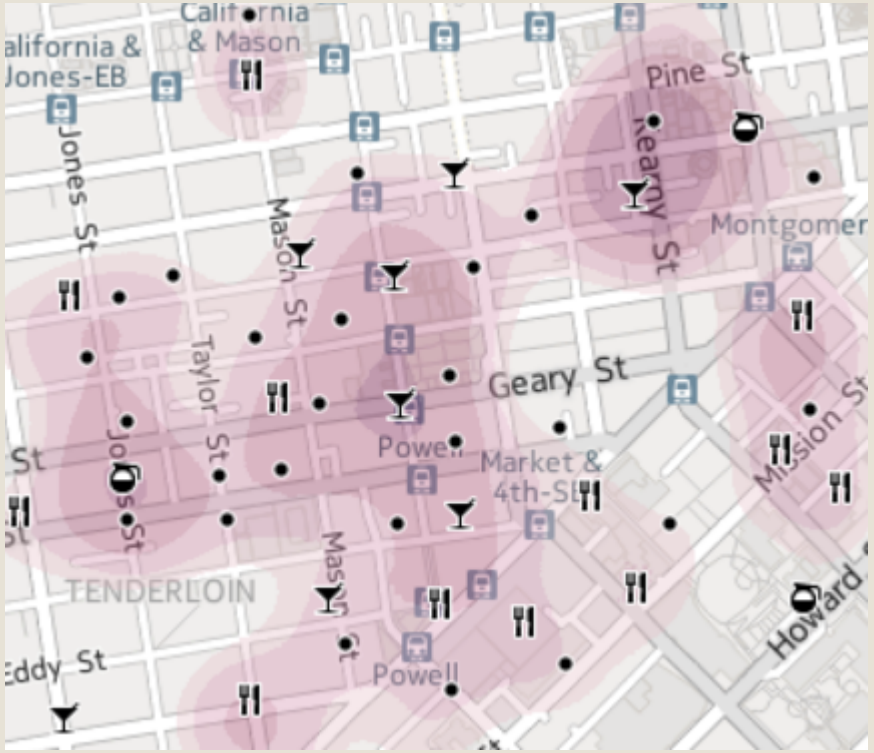


Still masking greater complexity

Ellipse/ellipsoid are pretty good for capturing the product of least squares or Kalman filters

Particle filters are much harder to capture

…draft-hoene-geopriv-bli



http://here.com/37.7873082,-122.4066945,16,0,0,gray.day

Mo' data, mo' troubles

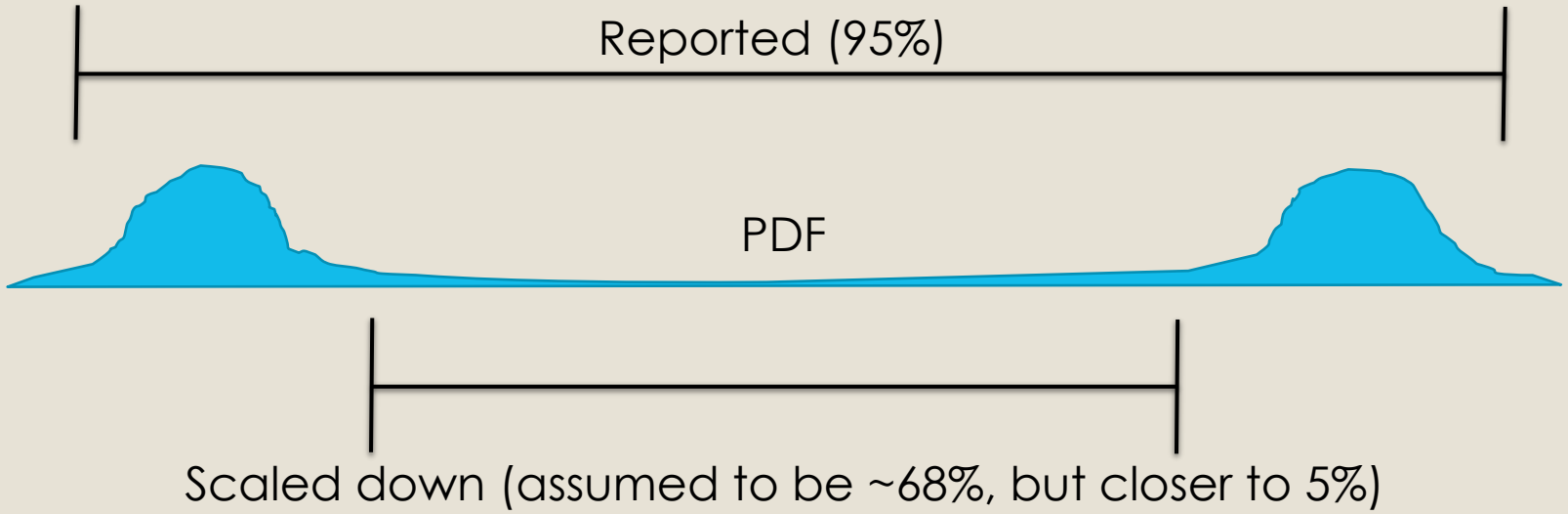
- Even the simplified information can be too much
- There are a bunch of things you can't do safely/easily
  - Most amount to the fact that you can't invent information you don't have
  - E.g., can't scale uncertainty without information loss
- Some applications require very little information
  - A point
  - Maybe a circle/sphere radius (so they can report "accuracy")
  - Is this location estimate "the same" as this other one
- So how do we get there?
  - draft-thomson-geopriv-uncertainty contains a bunch of cheats

Cheats

- Convert to point:
  - A simple method for calculating centroids of all the RFC 5491 shapes
  - Not so easy for polygons, but a robust approximation method provided
- Get single number for uncertainty:;
  - Two cheats: convert to circle
  - Use point calculation and find furthest point
  - Scale uncertainty based on probability distribution assumptions

Scaling

- Not always a good idea
  - Relies on assumptions
  - Big mistakes possible



- Scaling down is risky, scaling up is basically impossible
  - Unless you have some extra information.

CONFIDENCE

draft-thomson-geopriv-confidence-04

PIDF-LO assumes 95% confidence

- …and doesn't allow for divergence from this number
- That's a problem for implementations that are required to convert
  - Many existing systems produce estimates at other values
  - Conversion without sufficient knowledge requires assumptions
  - Assumptions cause data loss and errors

Impossibilities

- Sometimes 95% is unattainable
  - A >5% absolute error rate can happen
  - Maybe you are operating from a source that it just that bad
  - No alteration of the uncertainty value (other than to have it encompass the entire planet) can compensate for the errors
- e.g., Location determination based on a data set that is completely, irretrievably wrong 13.4% of the time
  - Confidence cannot be 86.6% or higher

Scaling hint

- Help with scaling by having an optional hint on PDF shape
  - Normal – scale up or down safely
  - Rectangular – scale down safely
  - Unknown – scale at own risk

Backwards compatibility

- None
- Intentionally – confidence changes everything
- …but it does more damage when you solve for backward compatibility
  - Scaling = bad
  - Alternative is no location information at all
- Only real solution is to admonish not to use confidence unless you are reasonably sure that the recipient will understand

THERE IS HOPE

Adopt