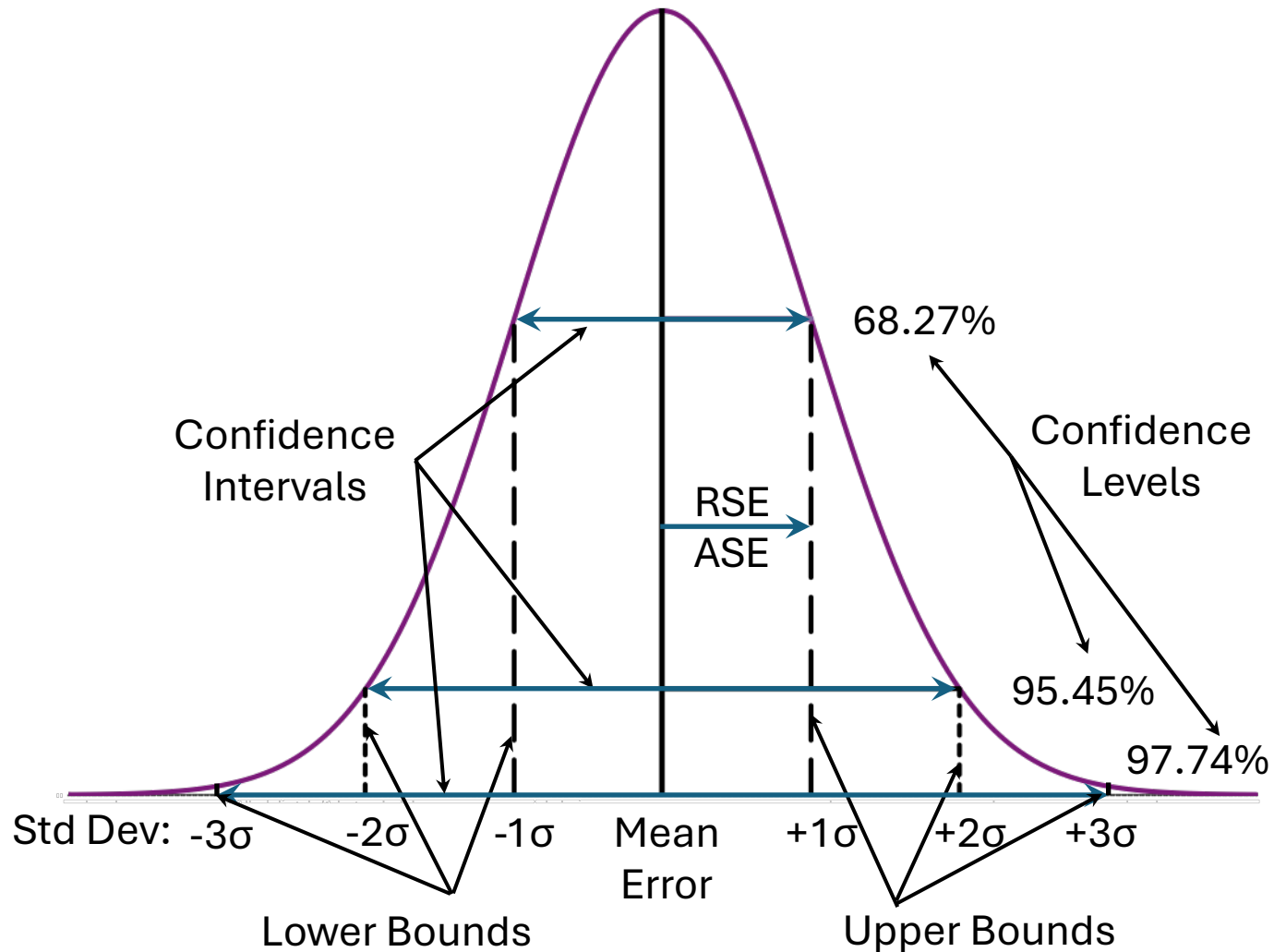# Outline

- Anatomy of the Gaussian Distribution of Error

- Multiplicative or Relative Standard Error (RSE)

  - HLL and Theta Sketches as examples

- Additive or Absolute Standard Error (ASE)

  - Quantile Sketches as examples

Lee Rhodes
Apache DataSketches PMC Chair, ASF Member
leerho@apache.org
13 Jan 2026

# Anatomy of the Gaussian Distribution of Error

For example, with the cardinality sketches (HLL, Theta, Tuple, CPC, FDT):

- The user selects *k* when building the sketch, which determines the size and basic accuracy (RSE) of the sketch.

- The user loads the sketch with a stream of data and then calls *getEstimate()*.

- The user can also query the sketch for the Upper Bound (UB) and the Lower Bound (LB) also choosing a *sigma* of 1, 2 or 3, which specifies a Confidence Level.

- The quantity (UB – LB) is the Confidence Interval. This reveals the quality of the estimate and is important to know!

Understanding Multiplicative or Relative Standard Error

The Sketch Error Specification:
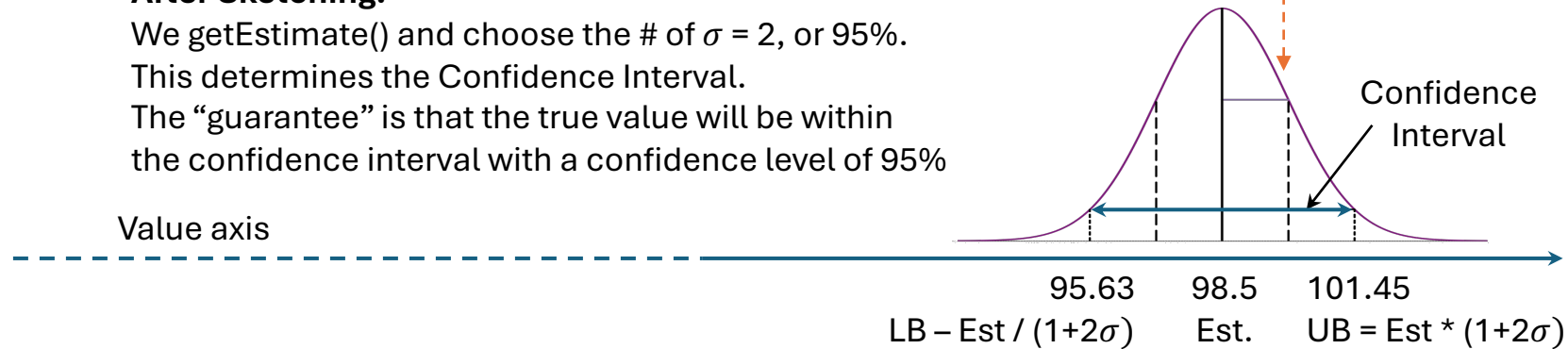With a specific *k*, the sketch will state that the error
(RSE) is, e.g., 1.5%.
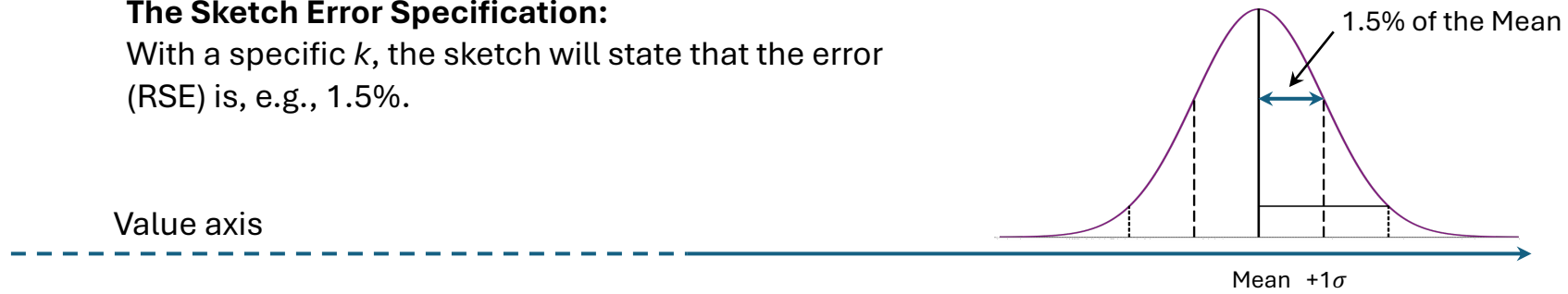
1.5% of the Mean

Value axis

Mean   +1σ

Understanding Multiplicative or Relative Sketch Error
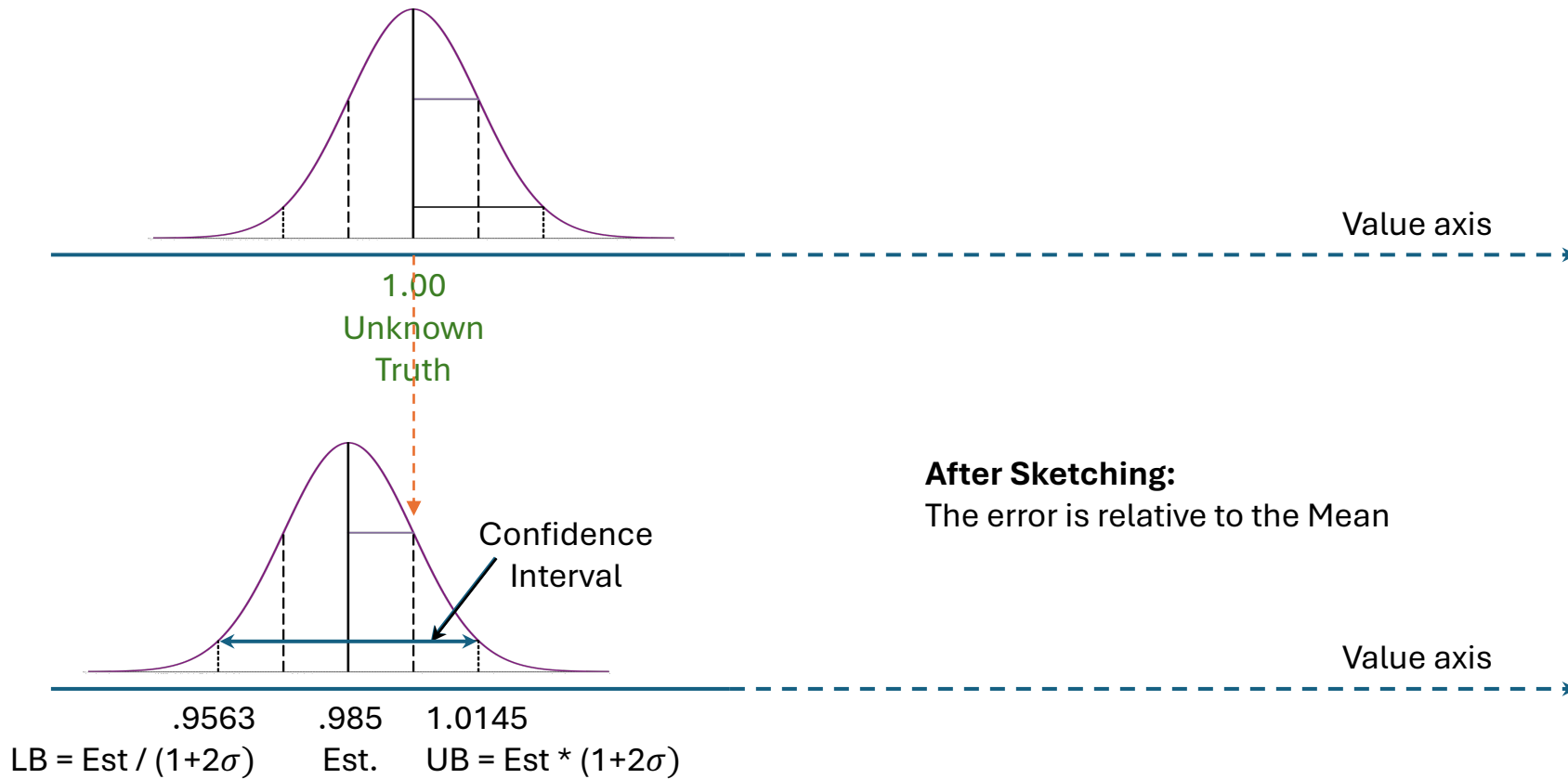# Large Values

Value axis
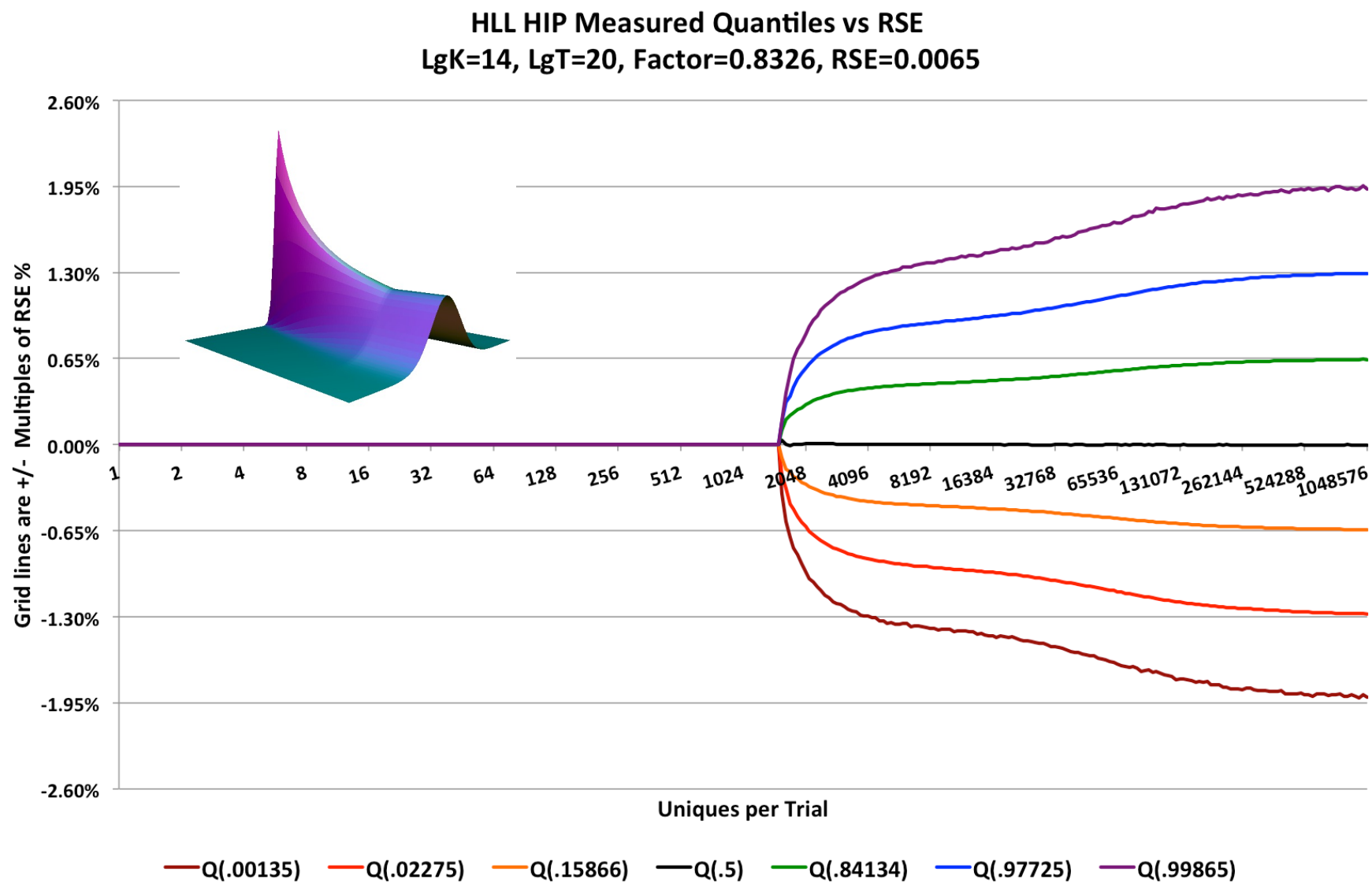
100
Unknown
Truth

After Sketching:
We getEstimate() and choose the # of σ = 2, or 95%.
This determines the Confidence Interval.
The "guarantee" is that the true value will be within
the confidence interval with a confidence level of 95%

Confidence
Interval

Value axis

95.63   98.5   101.45
LB – Est / (1+2σ)   Est.   UB = Est * (1+2σ)

Understanding Multiplicative or Relative Sketch Error
Small Values

Value axis

1.00
Unknown
Truth

After Sketching:
The error is relative to the Mean

Confidence
Interval

Value axis

.9563   .985   1.0145

LB = Est / (1+2σ)   Est.   UB = Est * (1+2σ)

Example (HLL) of Multiplicative or Relative Sketch Error

HLL HIP Measured Quantiles vs RSE
LgK=14, LgT=20, Factor=0.8326, RSE=0.0065

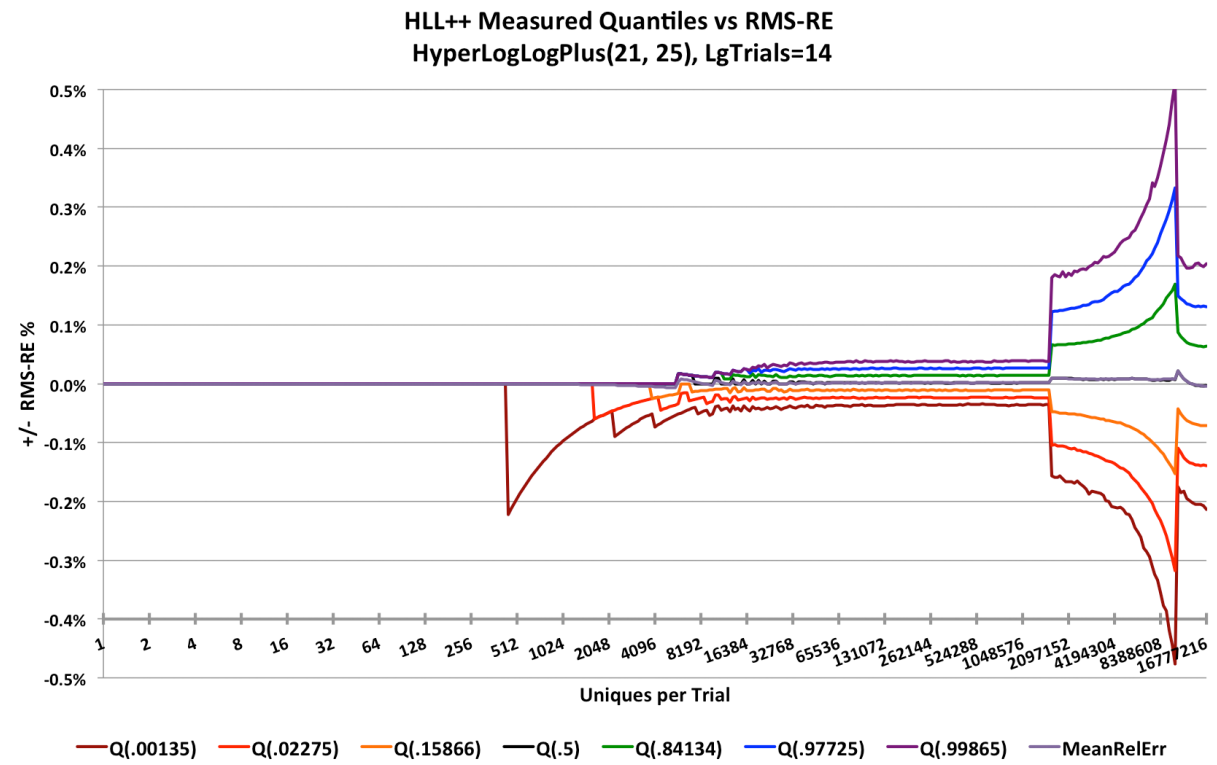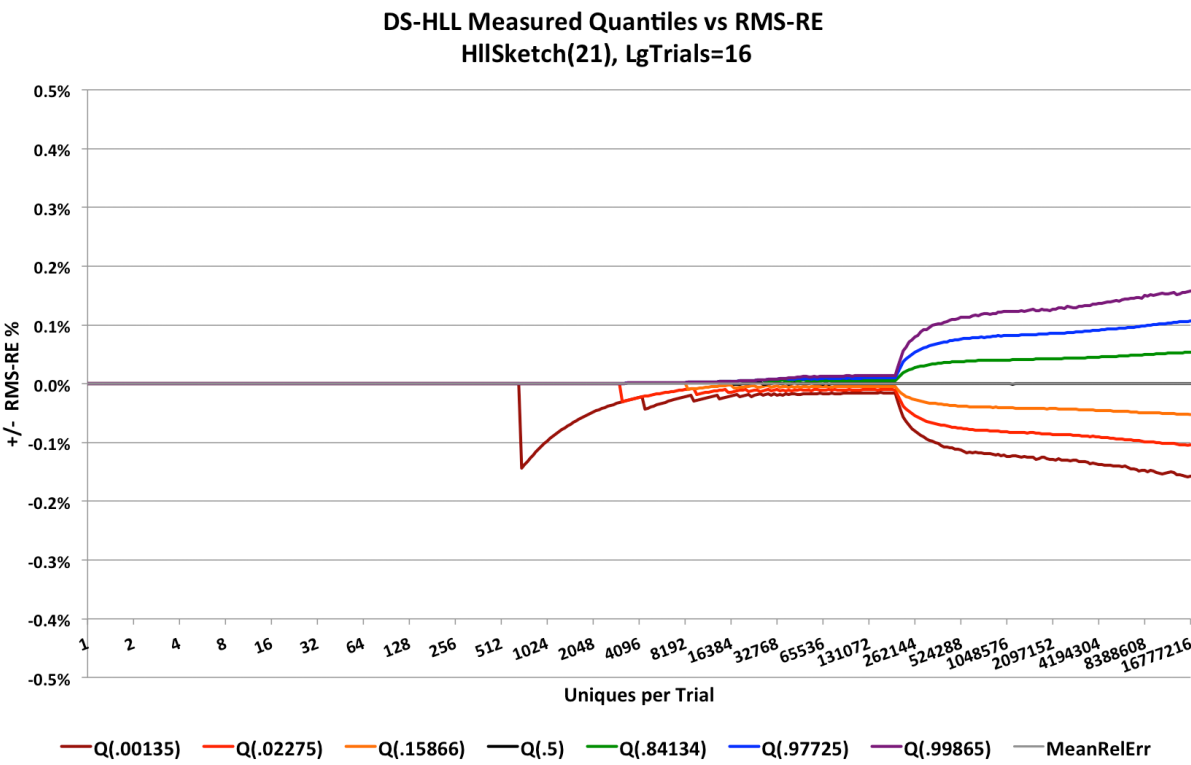Q(.00135) Q(.02275) Q(.15866) Q(.5) Q(.84134) Q(.97725) Q(.99865)

Example (Theta) of Multiplicative or Relative Sketch Error

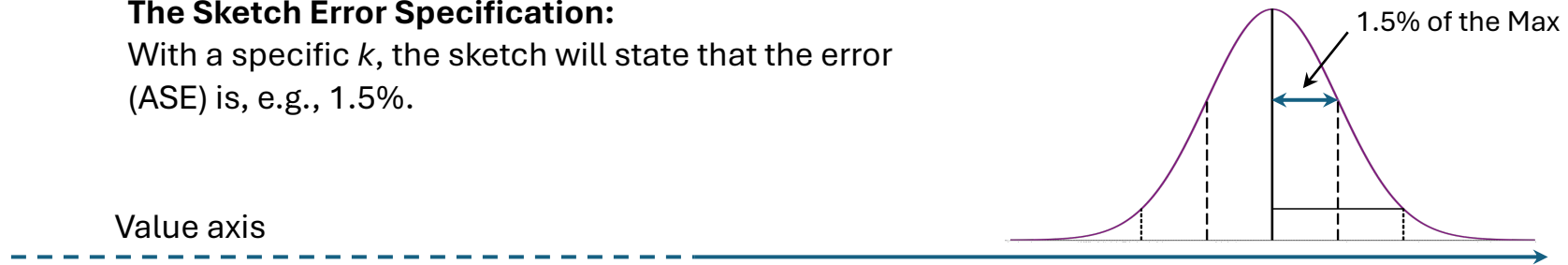Pitchfork Quantiles, k=4K, T=4K, p=1.0
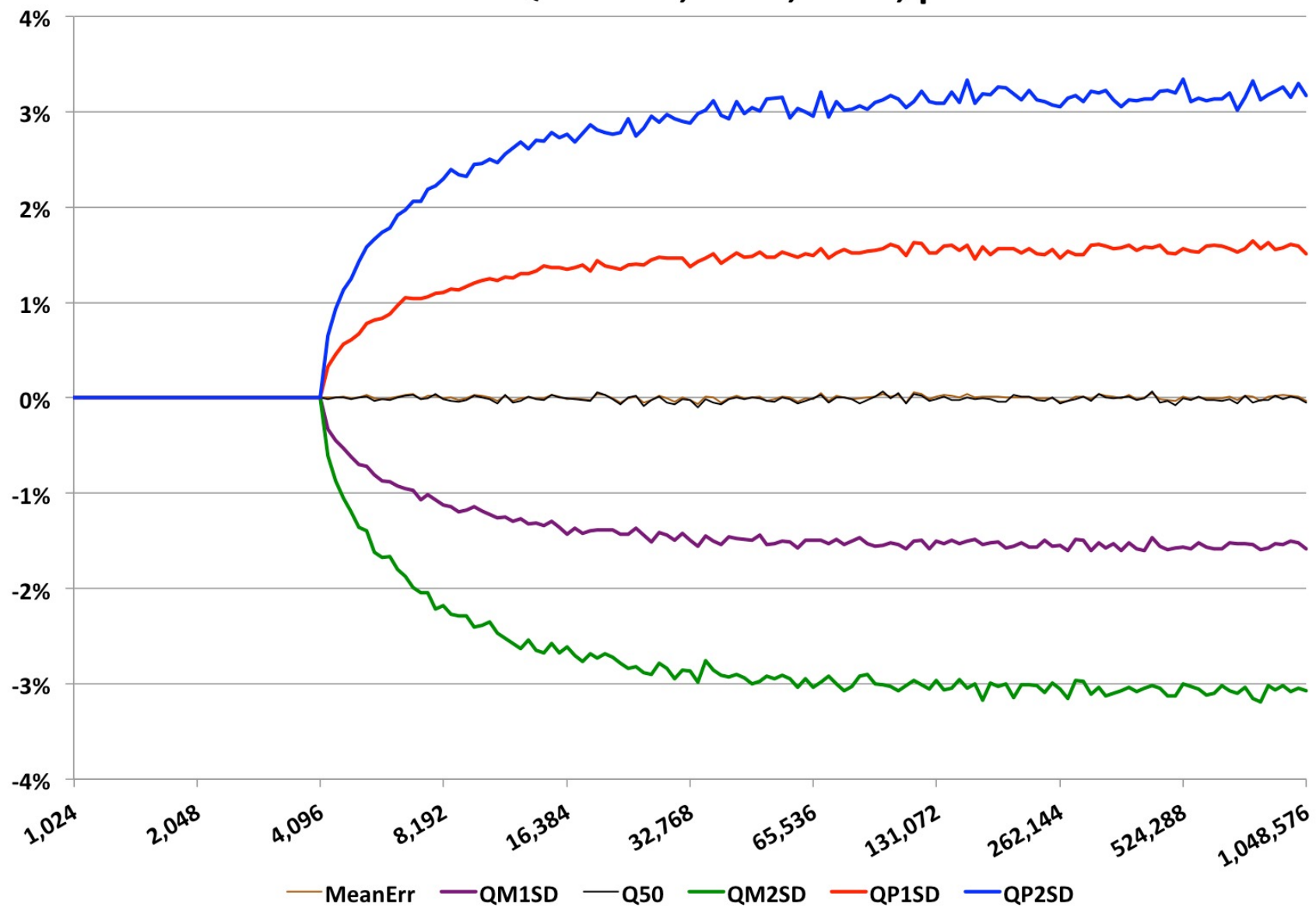
DataSketches HLL vs Clearspring Technologies HLL++
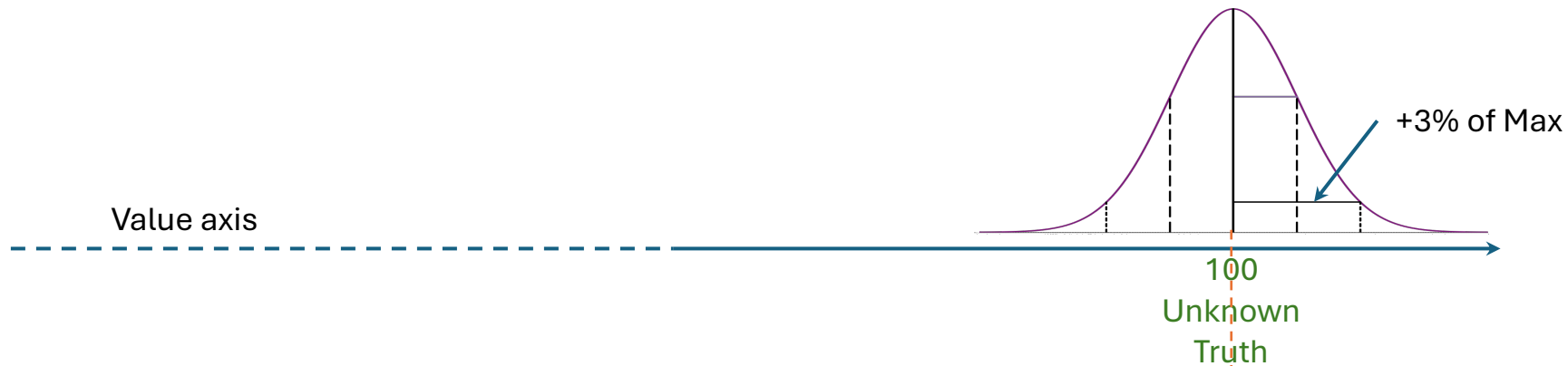
Understanding Additive or Absolute Standard Error

The Sketch Error Specification:
With a specific k, the sketch will state that the error
(ASE) is, e.g., 1.5%.

1.5% of the Max

Value axis

Understanding Additive or Absolute Sketch Error
Large Values

+3% of Max

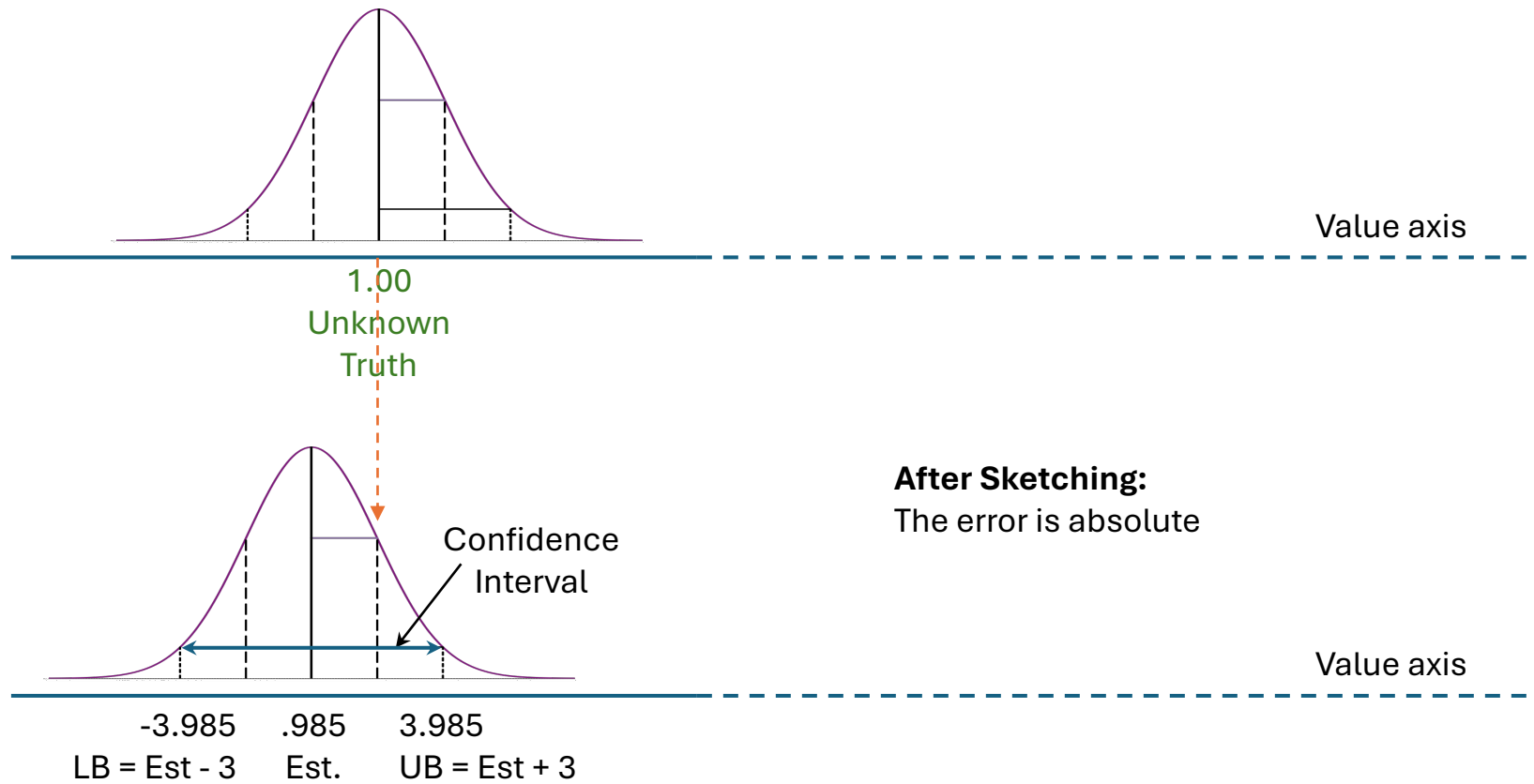100
Unknown
Truth

After Sketching:
We getEstimate() and choose the # of σ = 2, or 95%.
This determines the Confidence Interval.
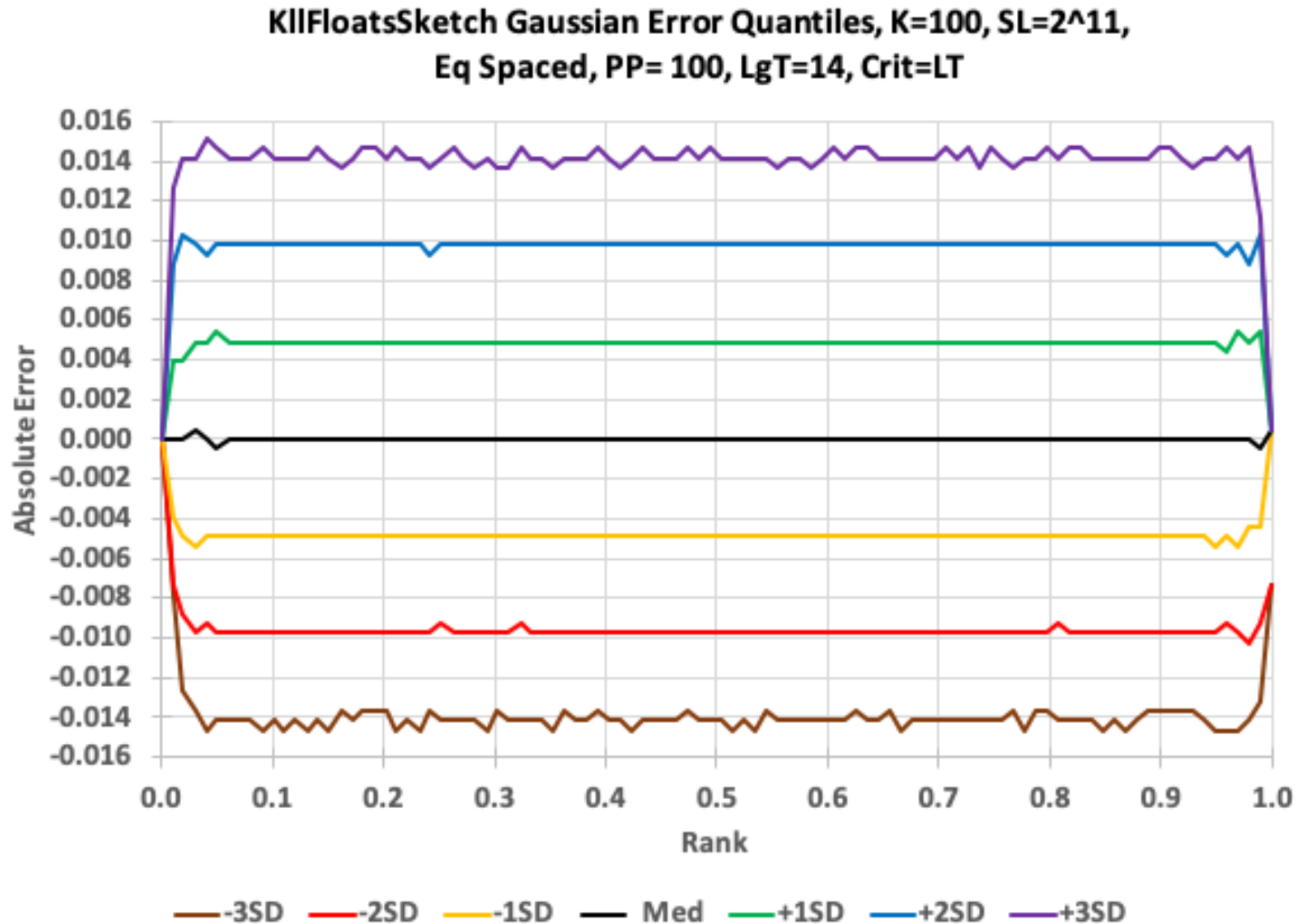The "guarantee" is that the true value will be within
the confidence interval with a confidence level of 95%

Value axis

Confidence
Interval

Value axis

95.5	98.5	101.5
LB = Est - 3	Est.	UB = Est + 3

Understanding Additive or Absolute Sketch Error
Small Values

Value axis

1.00
Unknown
Truth

After Sketching:
The error is absolute

Confidence
Interval

Value axis

-3.985   .985   3.985
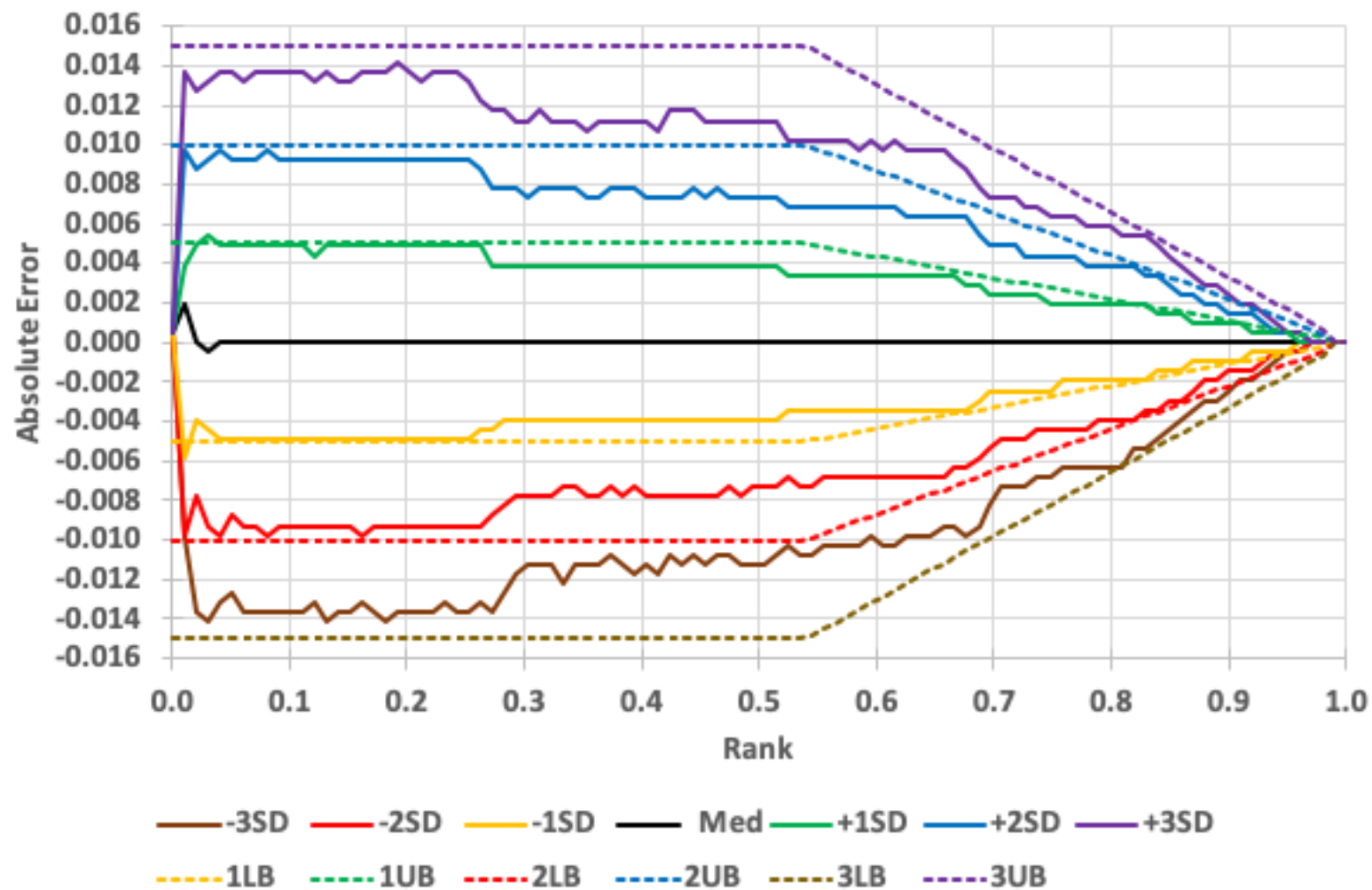LB = Est - 3   Est.   UB = Est + 3
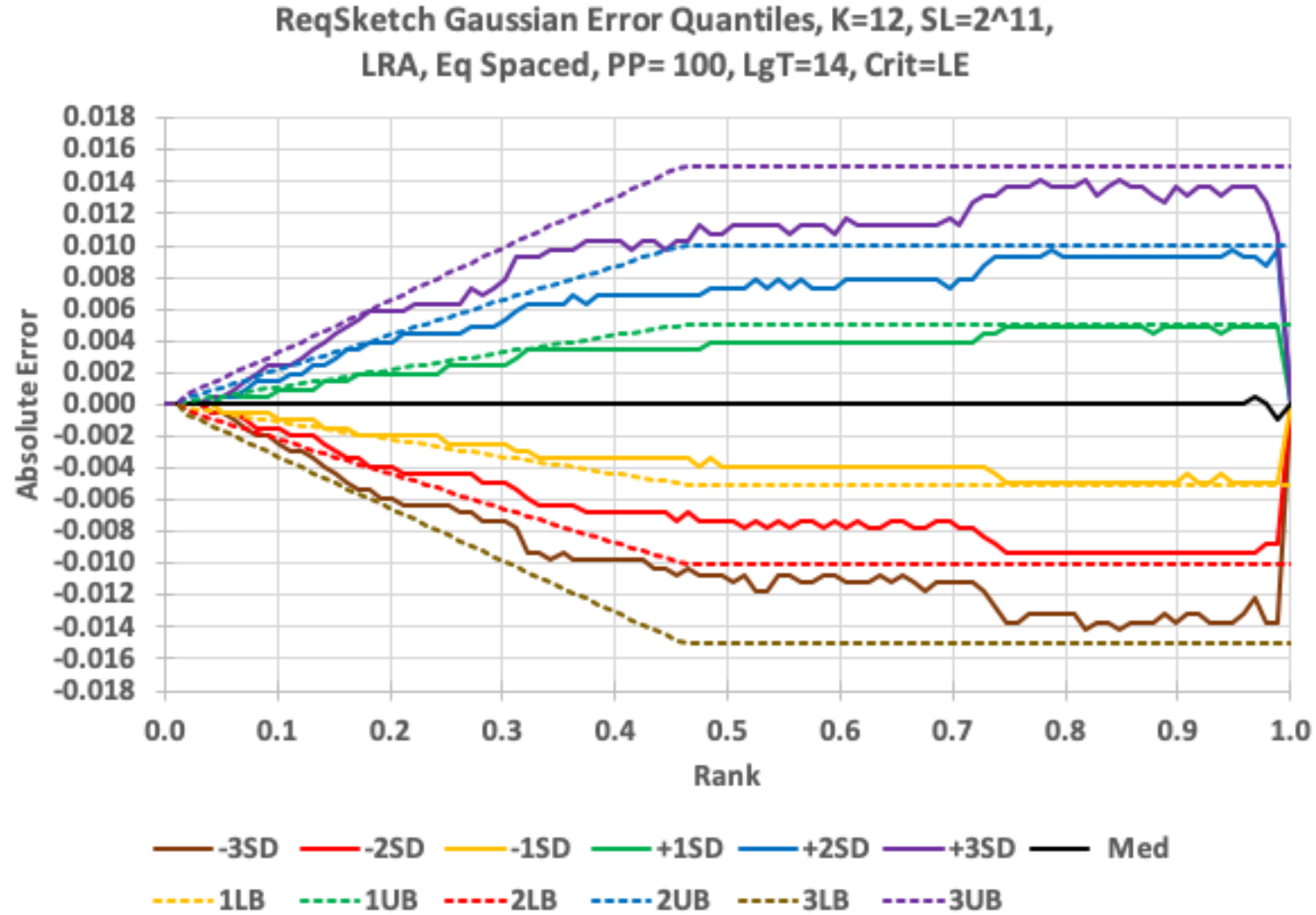
Example (KLL) of Additive or Absolute Sketch Error

Example (REQ-HRA) of Absolute + Relative Sketch Error

ReqSketch Gaussian Error Quantiles, K=12, SL=2^11,
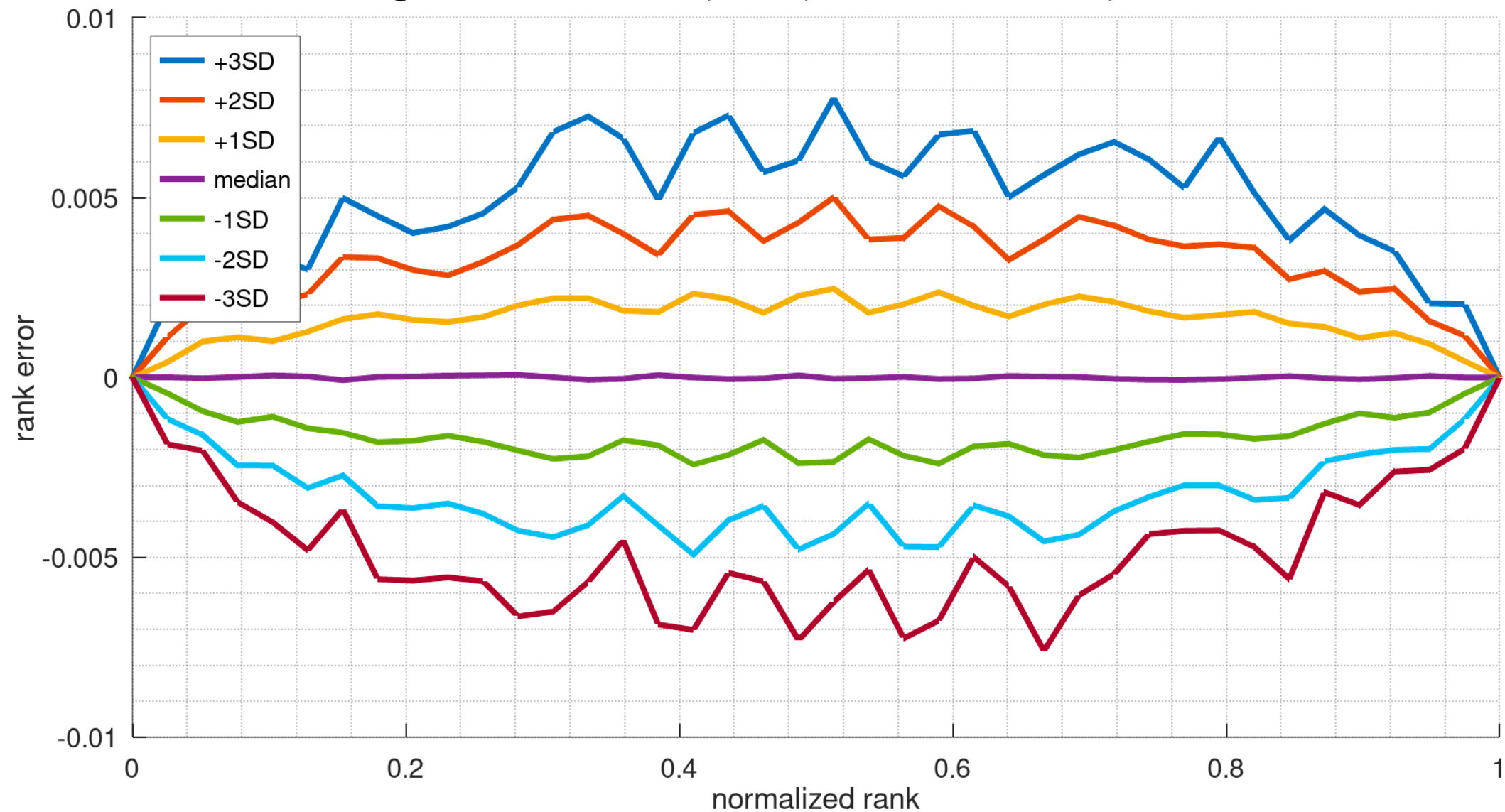LRA, Eq Spaced, PP= 100, LgT=14, Crit=LE

Example (REQ-LRA) of Absolute + Relative Sketch Error

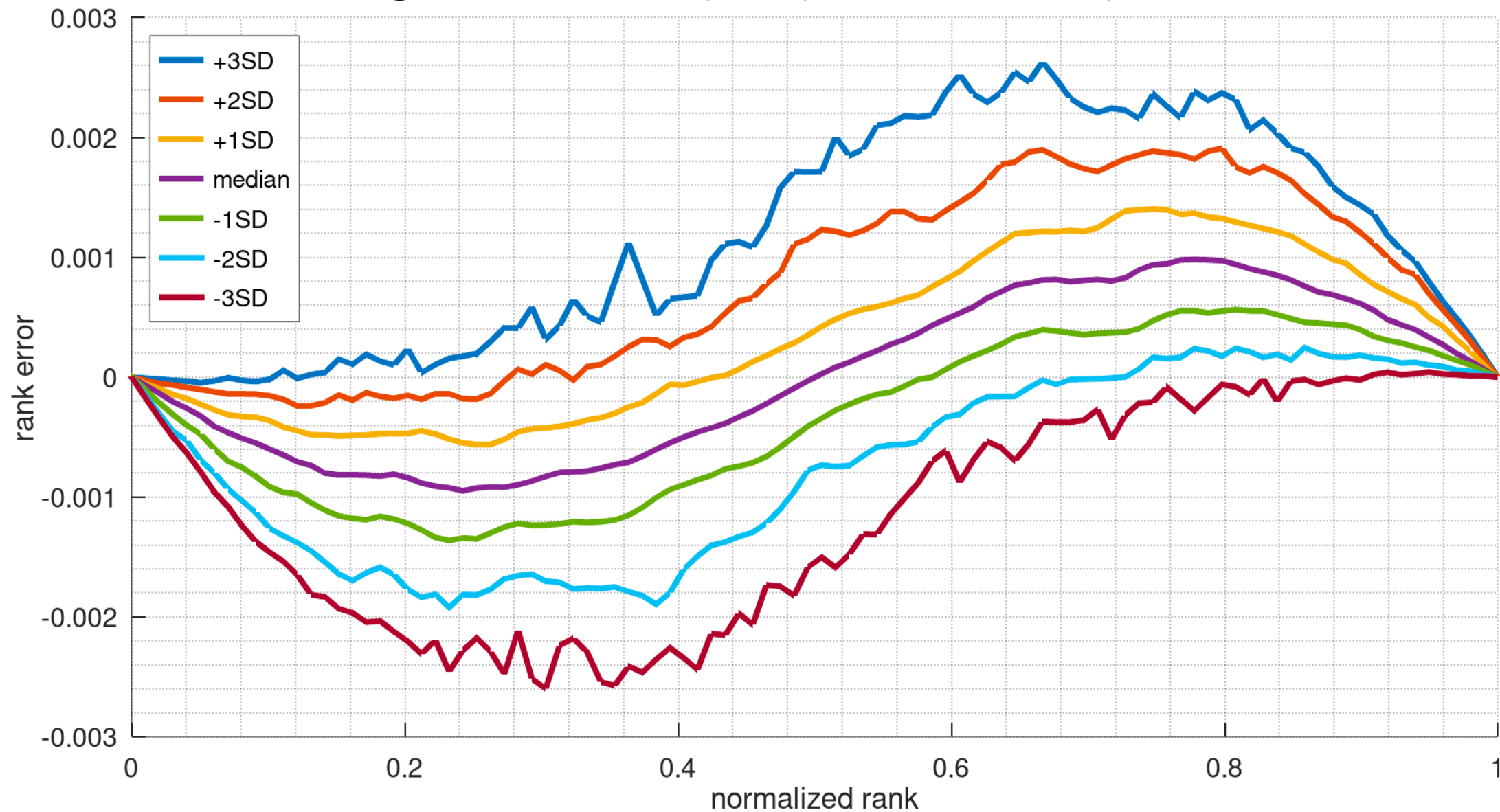ReqSketch Gaussian Error Quantiles, K=12, SL=2^11,
LRA, Eq Spaced, PP= 100, LgT=14, Crit=LE

Example (Tdigest), No Error Guarantees, Only Numerics, N = 1K

TDigest rank error K=100, N=2^{10}, uniform distribution, 10000 trials

Example (Tdigest), No Error Guarantees, Only Numerics, N = 32M

TDigest rank error K=100, N=2^{25}, uniform distribution, 10000 trials

Thank You!

Open Invitation for Collaboration

https://datasketches.apache.org