# Bootstrap: Limits and Criticalities

Giuseppe Frisella — Introduzione alla Complessità – A.A. 24/25



BASIC CONCEPTS

# The Bootstrap Method

It is a resampling technique with replacement that allows the study of an estimator's distribution. Starting from a sample, other “bootstrap” samples of the same size are generated, and the estimator is calculated for each of them, obtaining its empirical distribution.

### Resampling

Repeatedly drawing observations from the original dataset

### Empirical Estimate

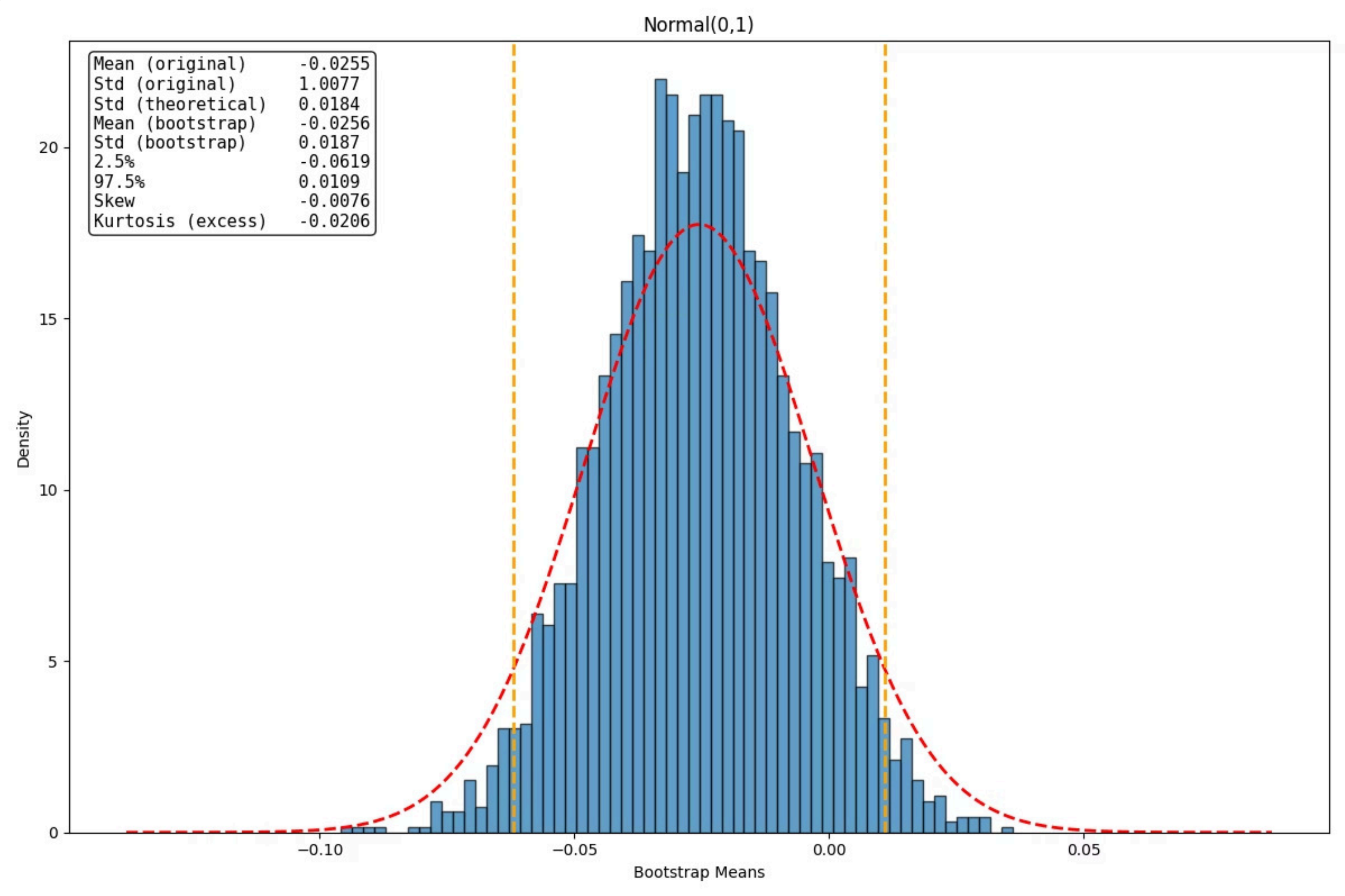
Building the sample distribution without relying on assumptions

### Confidence Intervals

Assessing the uncertainty and precision of estimates

# Bootstrap on Gaussian Distribution

The method works correctly and provides accurate results



CHALLENGES OF BOOTSTRAP

# Intrinsic limits of the Bootstrap method

Despite its versatility, the method has some limitations. Its effectiveness relies on two fundamental assumptions about the data: finite variance and independence of observations.

Ignoring these conditions can nullify the benefits of resampling.

### Infinite Variance

The Central Limit Theorem does not apply, preventing the sample mean from converging to a Gaussian distribution.

### Memory

The lack of independence leads to distorted estimates of variability and uncertainty.

EVALUATING THE ACCURACY

# Core Metrics

To assess the limitations and reliability of the bootstrap, it is necessary to analyze the distribution of the means using quantitative metrics.

### Percentiles

Divide a distribution into 100 equal probability parts. Comparing empirical and normal percentiles reveals departures from normality.

### Skewness

Defined as $\mu_3/\sigma^3$. It is a measure of asymmetry: positive values indicate right skewness, negative values left skewness.

### Kurtosis

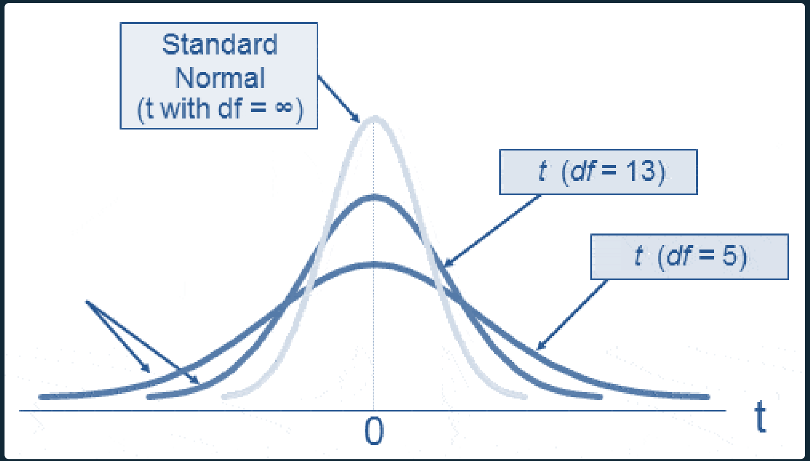
Defined as $\mu_4/\sigma^4$. It measures tail thickness: values above 3 indicate heavier tails than normal, while values below 3 indicate lighter tails.

CRITICAL CASES: FAT-TAILED DISTRIBUTIONS

# The Student's t-Distribution

The **Student's t-distribution** is fundamental for describing phenomena with **heavy tails**, such as when working with **small samples** and having **little information** about the original population. The parameter of the **degrees of freedom** ν determines how thick the tails are: low values of ν lead to a higher frequency of extreme observations.

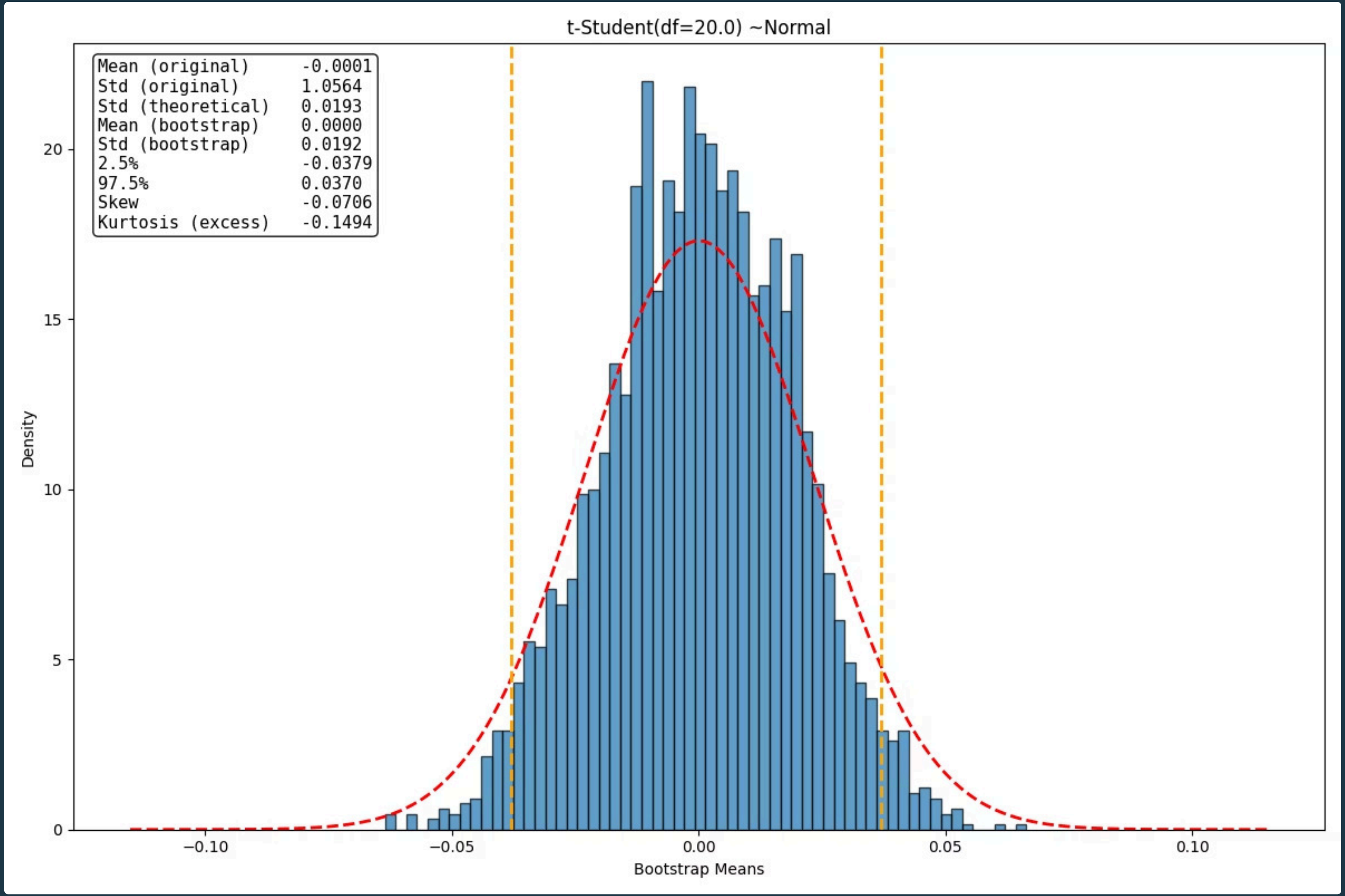
- The mean is finite only if $\nu > 1$
- The variance is finite only if $\nu > 2$



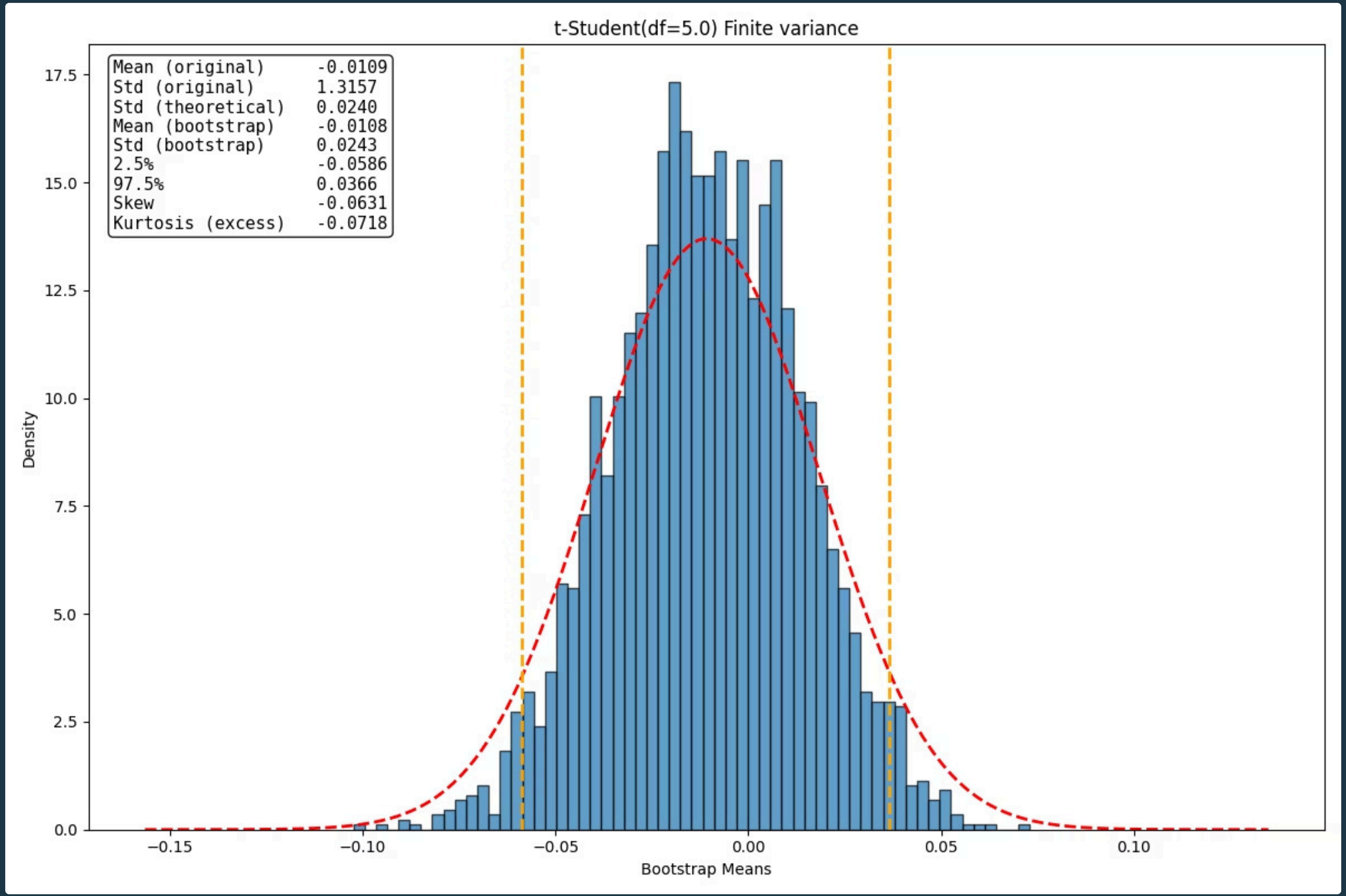

The implementation of the Student's t-distribution was carried out with the Python libraries used in the code.

$$f(t) = \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{\frac{-(\nu+1)}{2}}$$

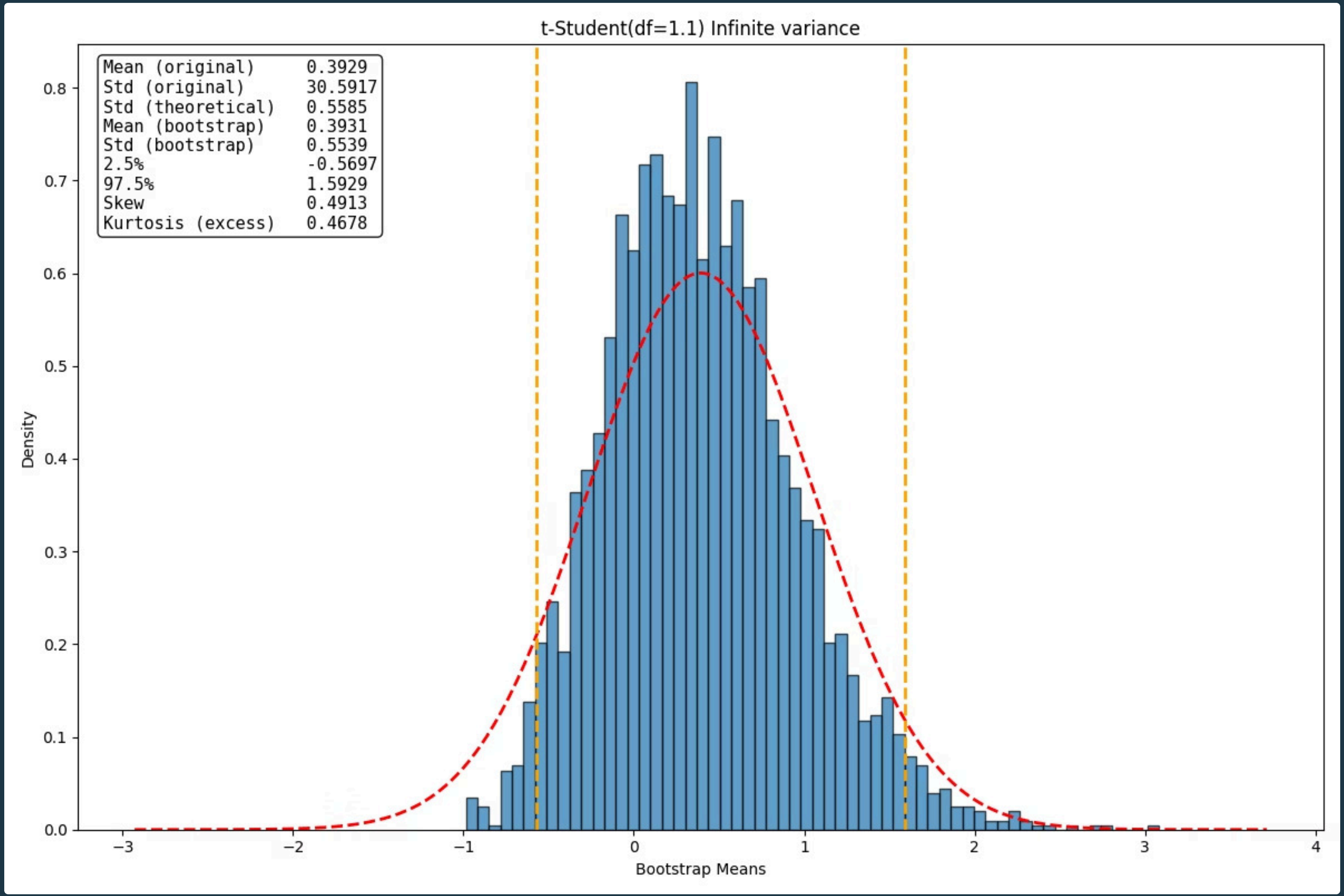
# Bootstrap and t-Student



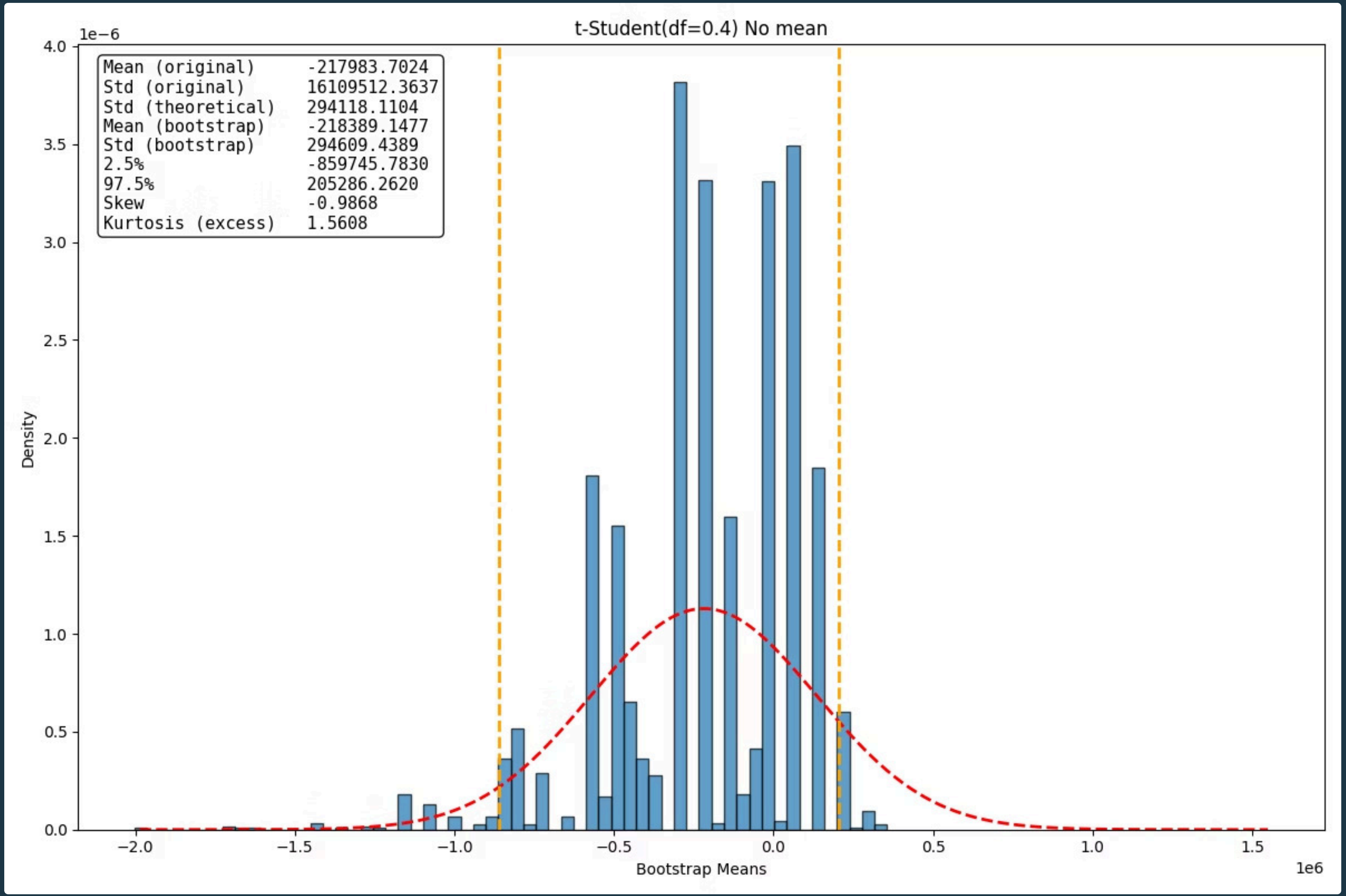
# Bootstrap and t-Student



# When the Bootstrap fails

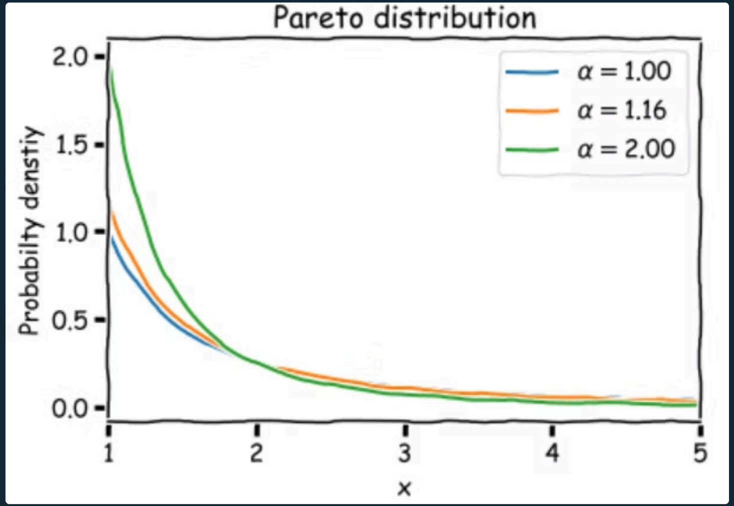


# When the Bootstrap fails



ANOTHER HEAVY-TAILED EXAMPLE

# The Pareto Distribution



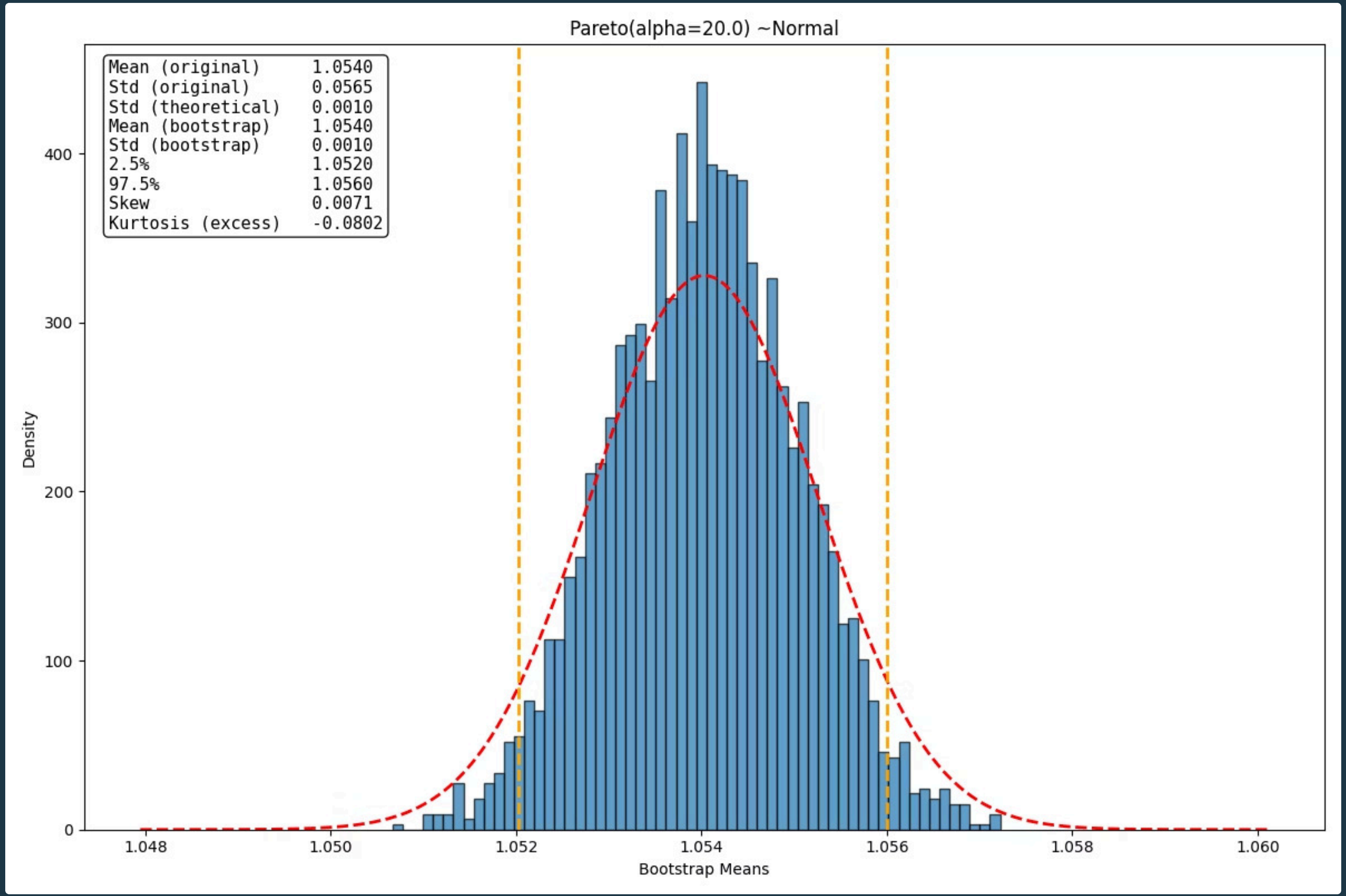

The Pareto distribution describes phenomena where a small percentage of factors are responsible for the majority of effects (e.g., wealth distribution, financial risks, natural disasters). It is a classic example of a heavy-tailed distribution, where rare events can have enormous impacts. The parameter $\alpha$ governs its behavior.

- The mean is finite if $\alpha > 1$
- The variance is finite only if $\alpha > 2$

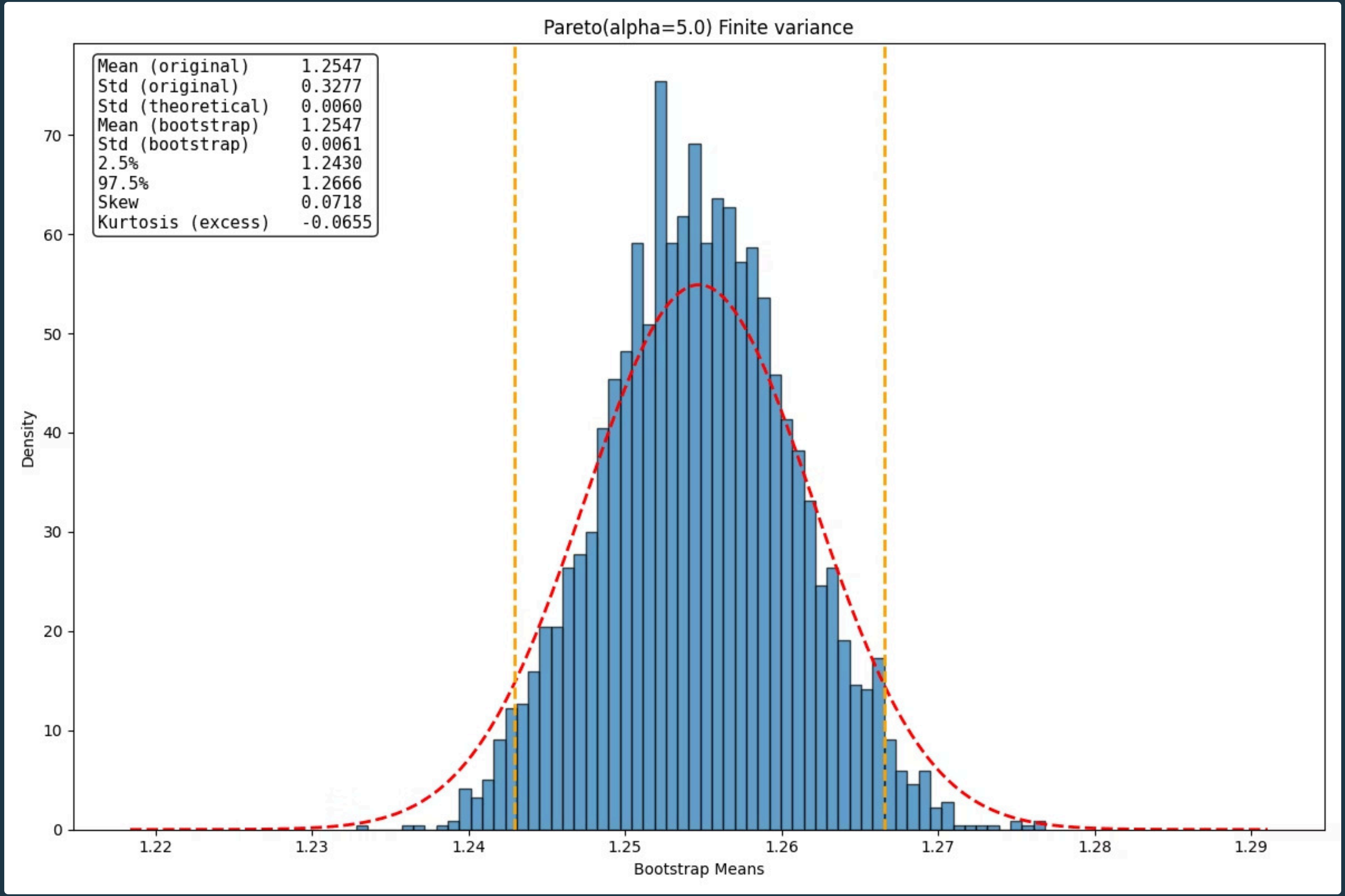
$$f_X(x) = \begin{cases} \dfrac{\alpha x_{\mathrm{m}}^{\alpha}}{x^{\alpha+1}} & x \geq x_{\mathrm{m}} \\ 0 & x < x_{\mathrm{m}} \end{cases}$$

The implementation of the Pareto distribution was carried out with the Python libraries used in the code.

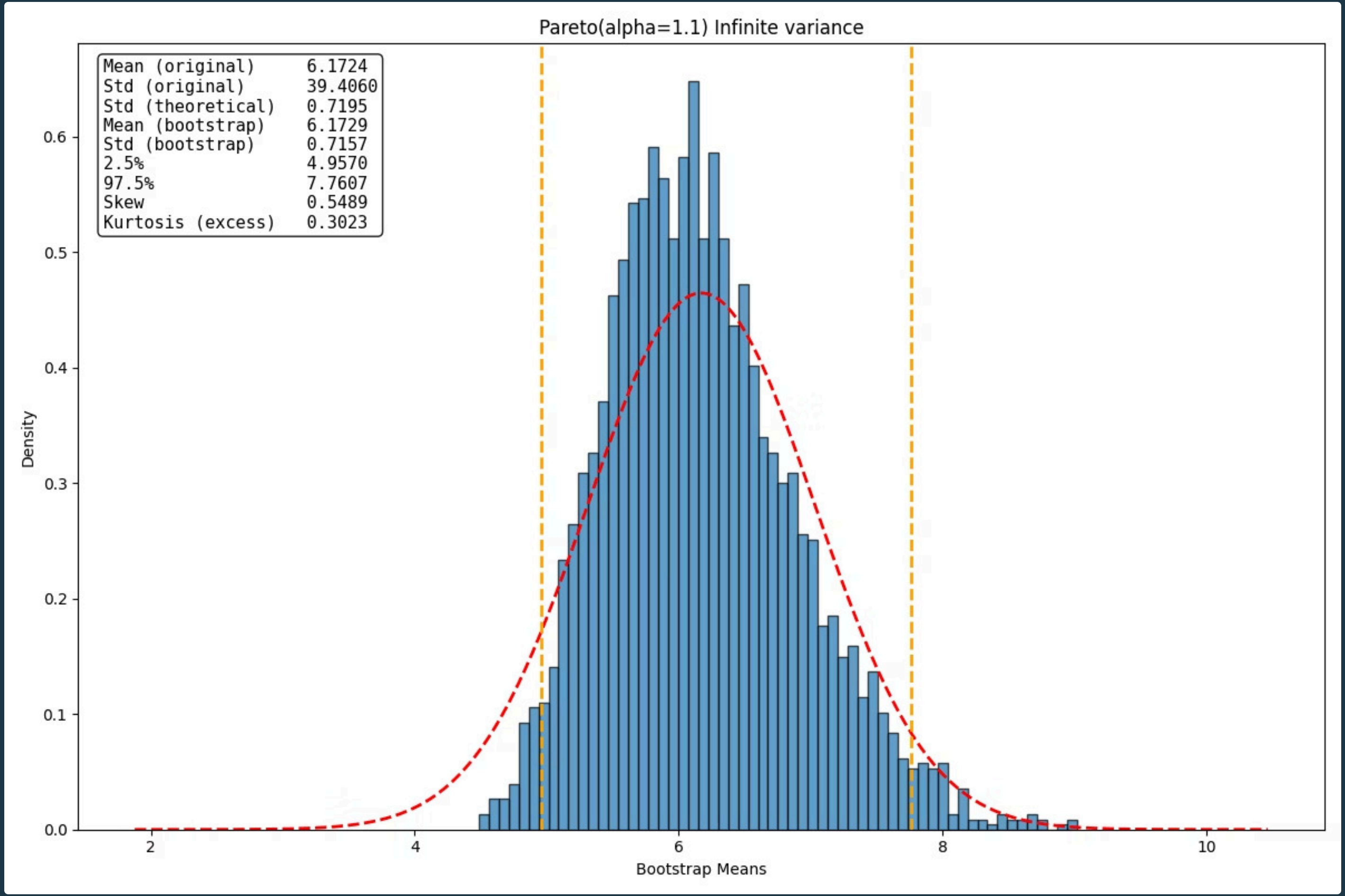
# Bootstrap and Pareto



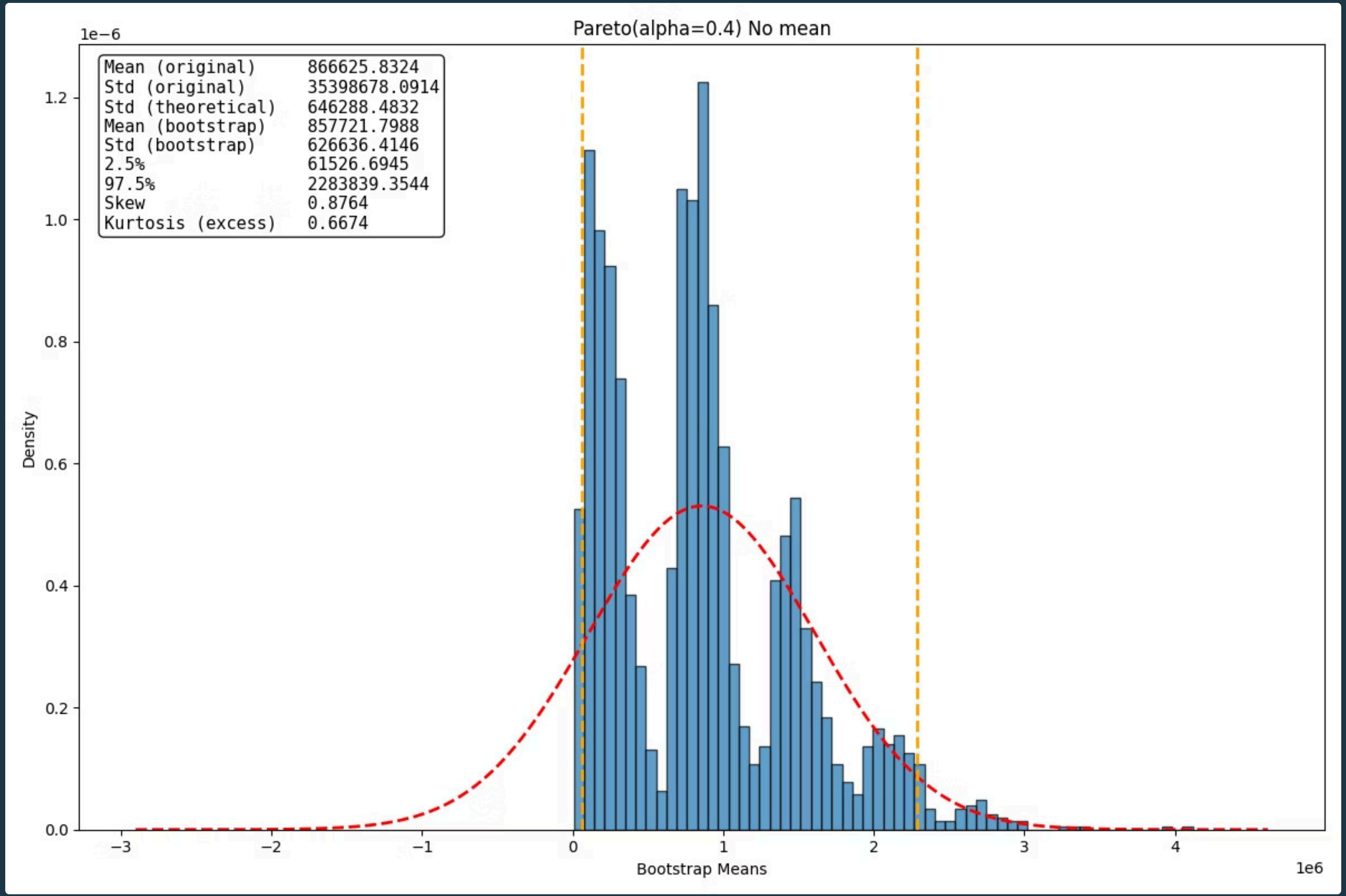
# Bootstrap and Pareto



# When the Bootstrap fails

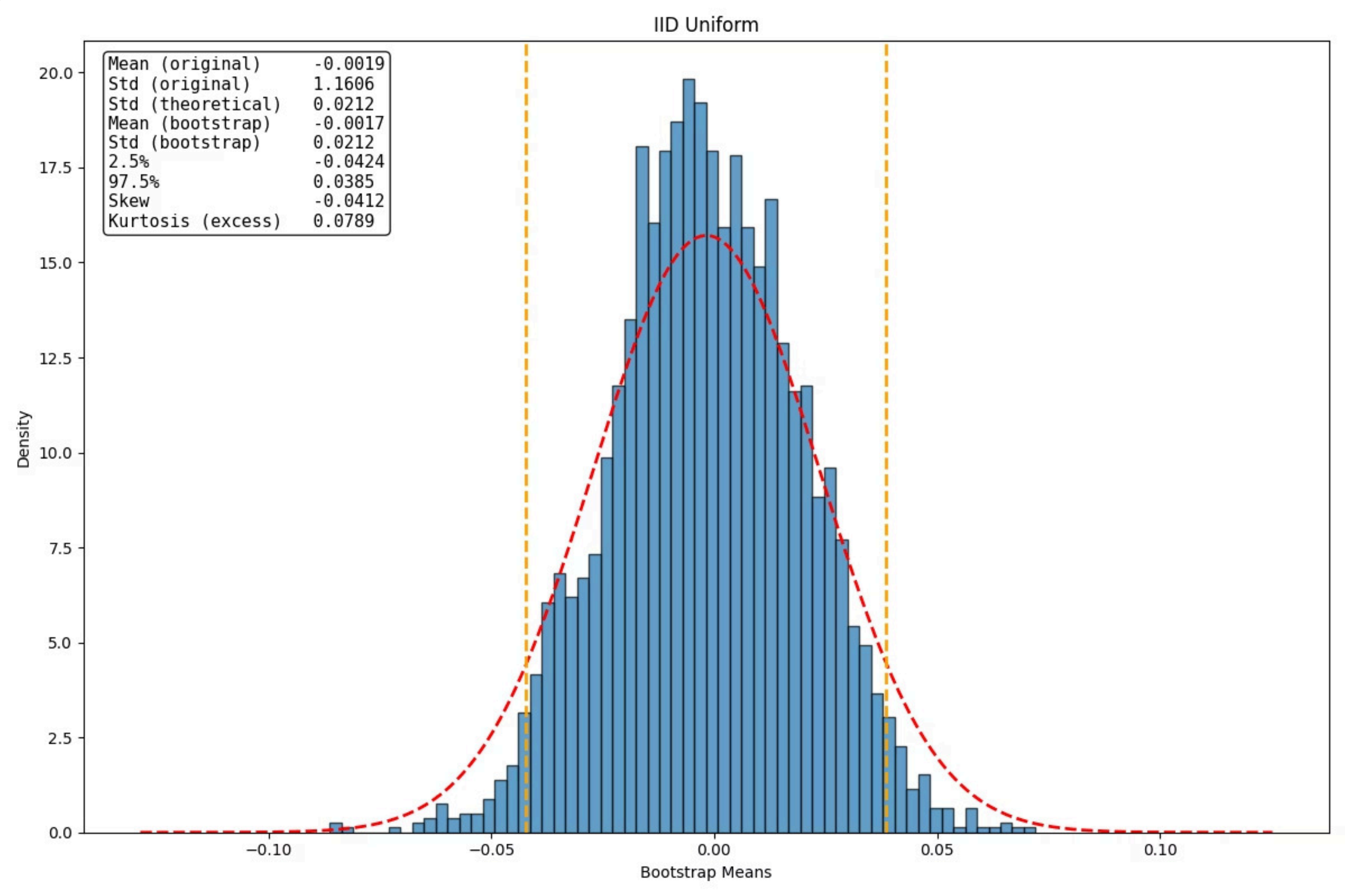


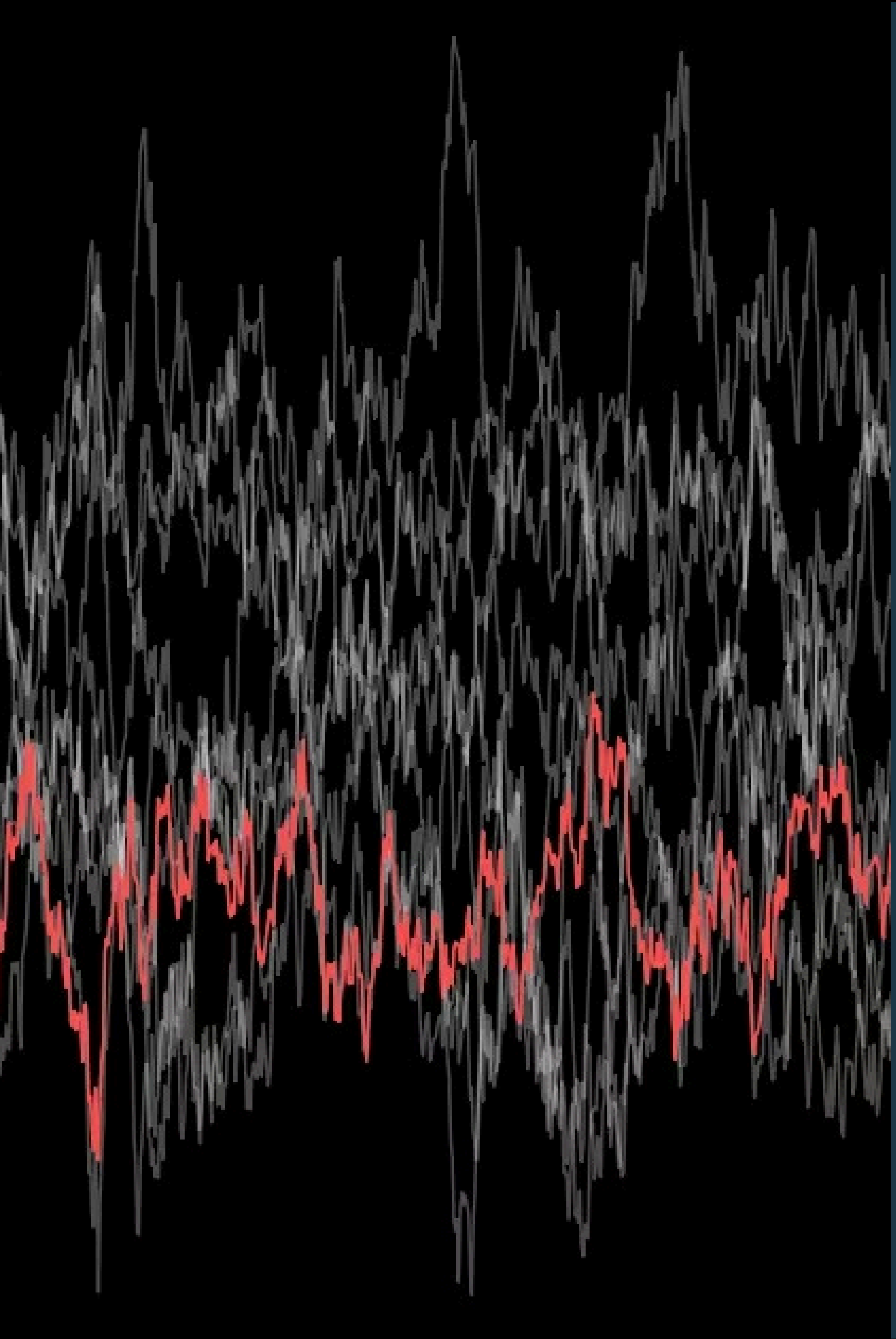
# When the Bootstrap fails



# Memory in Data and Implications

Bootstrap methods rely on the independence of observations. In the presence of autocorrelation, this assumption breaks down





# Ornstein-Uhlenbeck Process

The Ornstein–Uhlenbeck process describes mean-reverting stochastic dynamics. It is defined by the stochastic differential equation:

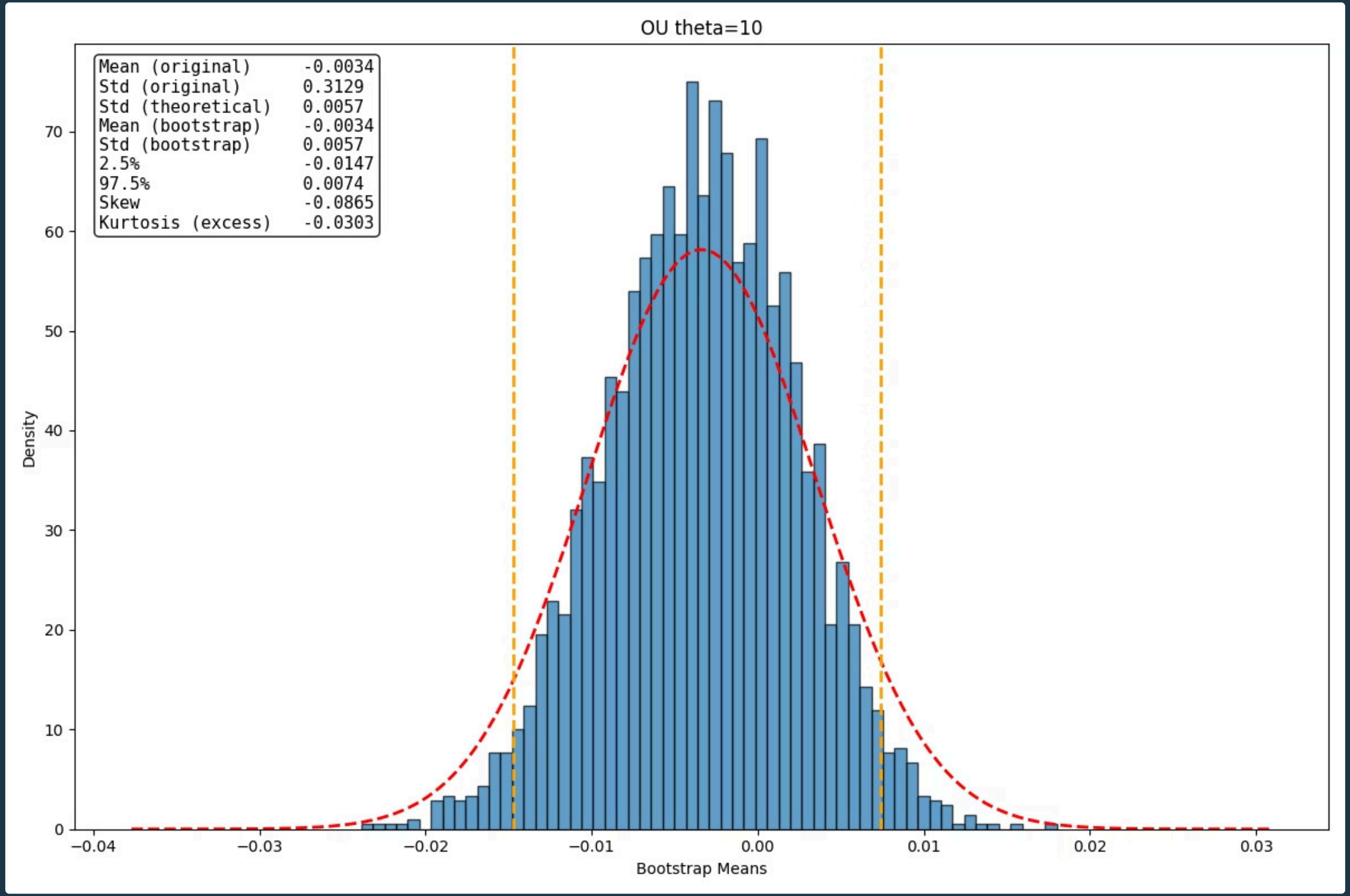
$$dX_t = -\theta(X_t - \mu)dt + \sigma dW_t$$

The autocorrelation decays exponentially with the lag $\tau$ according to: $\rho(\tau) = e^{-\theta\tau}$ which implies short memory.

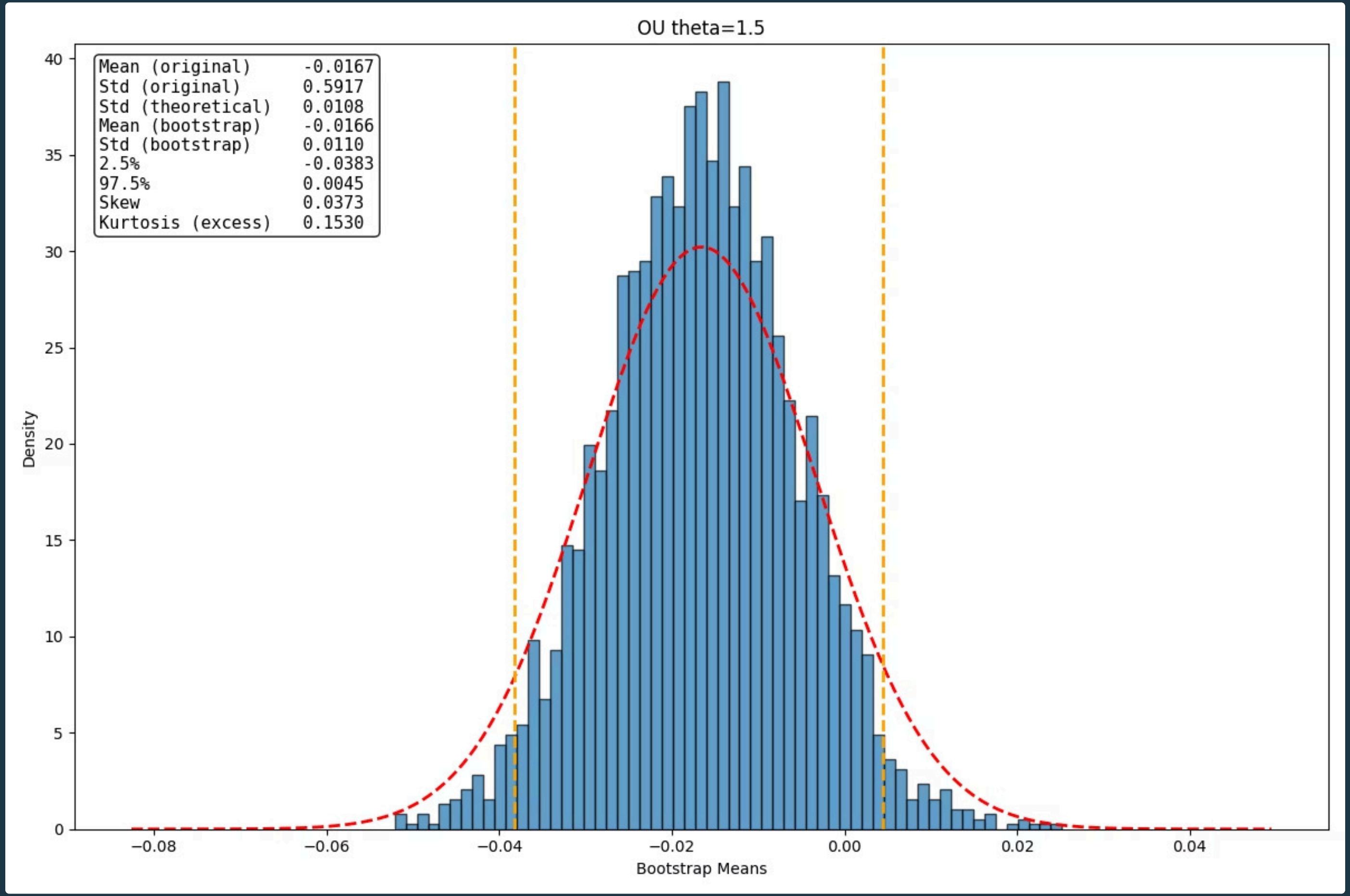
Large values of $\theta$ lead to a rapid return to the mean and short memory, while smaller values produce more persistent dynamics.

The numerical simulation is obtained by discretizing the stochastic equation using the Euler–Maruyama method: $X_{t+\Delta t} = X_t + \theta(\mu - X_t)\Delta t + \sigma\sqrt{\Delta t}Z_t$ con$Z_t \sim \mathcal{N}(0,1)$

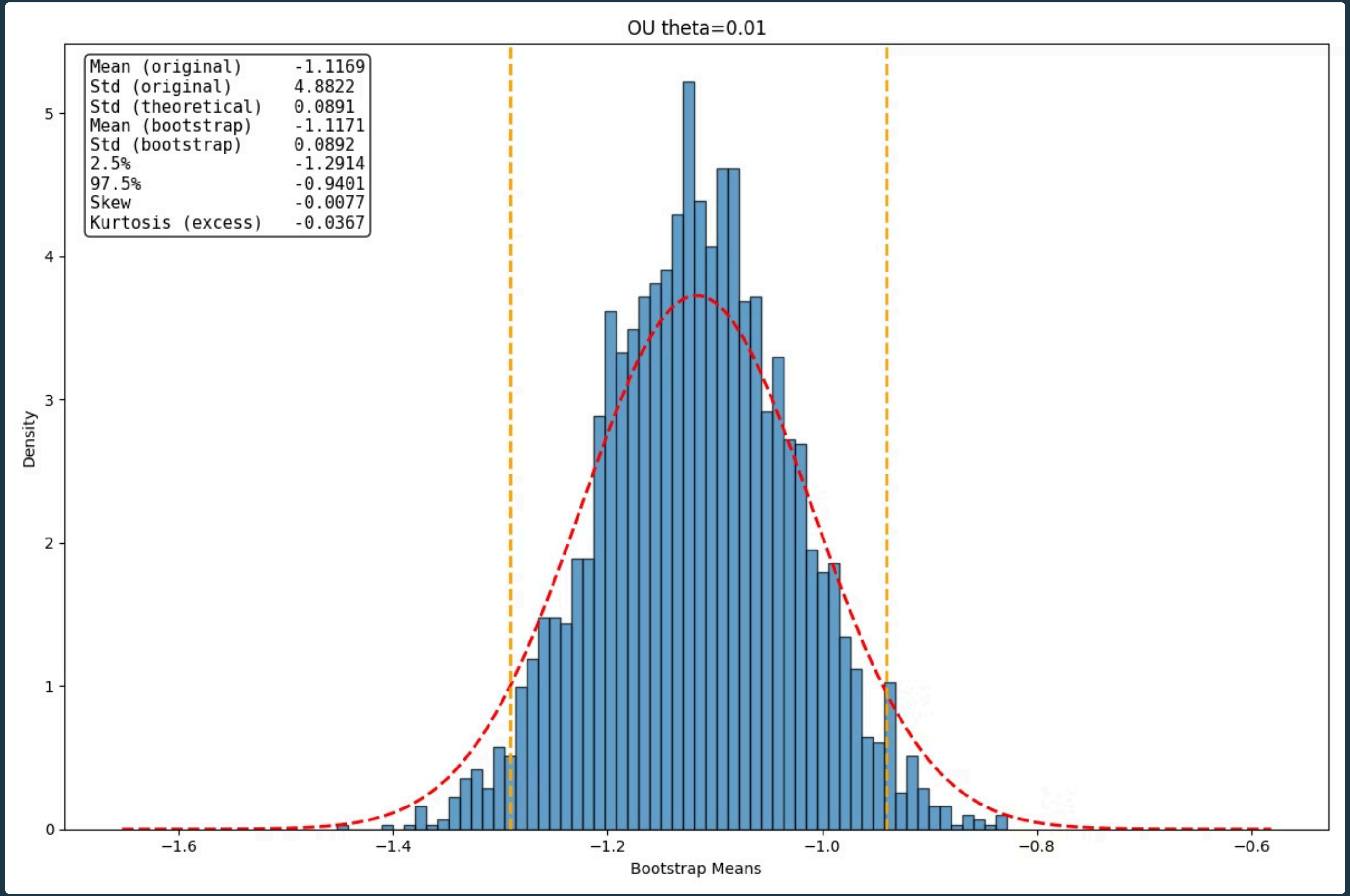
# Bootstrap and Ornstein-Uhlenbeck

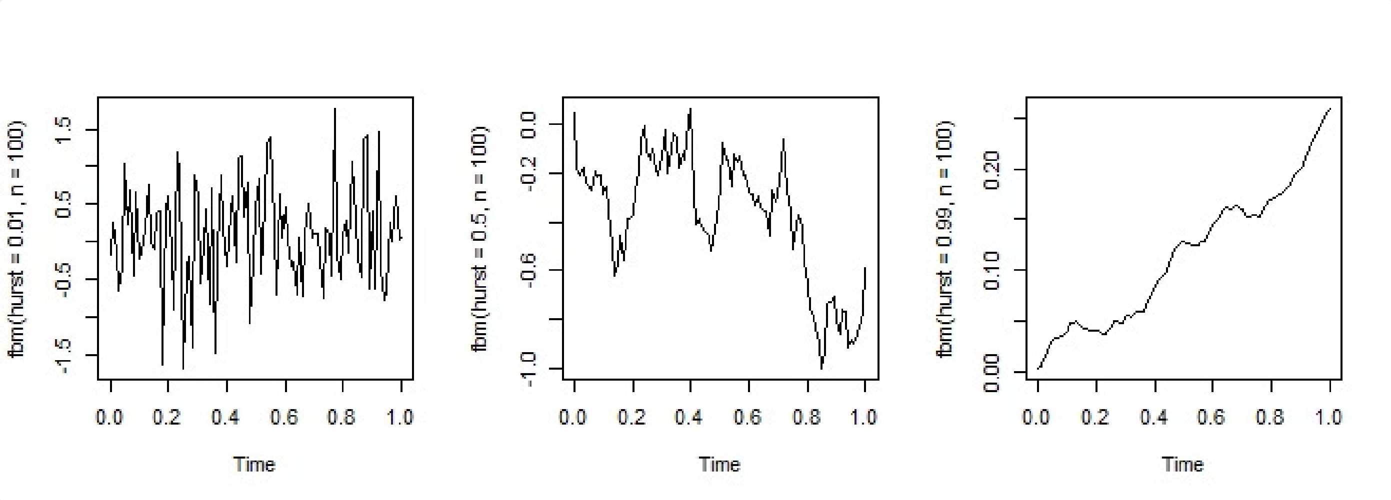


# Bootstrap and Ornstein-Uhlenbeck



# Bootstrap and Ornstein-Uhlenbeck





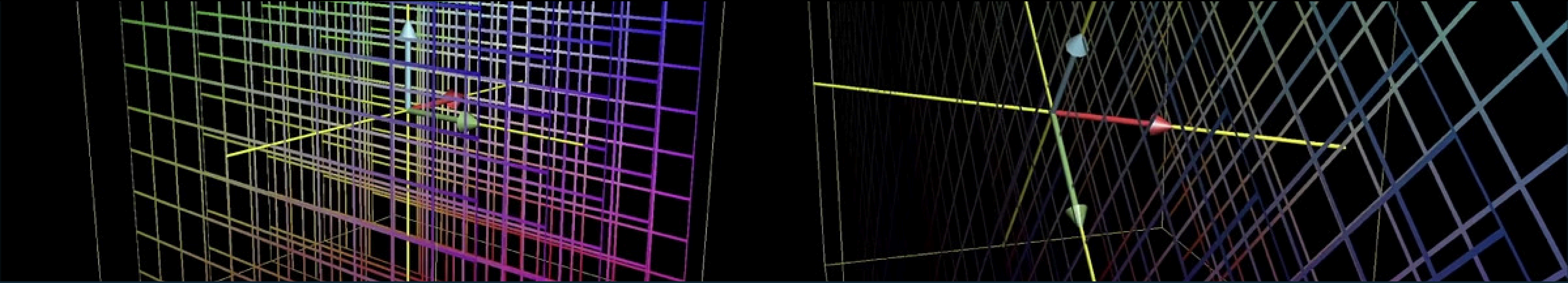

# Fractional Gaussian Noise

Fractional Gaussian noise (fGn) is a process with long-range correlations, used to model phenomena with memory such as hydrological flows, biological signals, network traffic, or financial time series. It is obtained as the discrete increment of a fractional Brownian motion (fBm) $B_H(t)$, a generalization of Brownian motion, with Hurst exponent $H$: $X_k = B_H(k+1) - B_H(k)$

The autocorrelation of fGn derives from that of fBm: $\mathrm{Cov}\big(B_H(t), B_H(s)\big) = \frac{1}{2}\Big(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\Big)$

The parameter $H$ controls the memory of the process: $H > 0.5$ indicates persistence, $H < 0.5$ anti-persistence, $H = 0.5$ white noise.

There is no simple closed-form formula to generate fGn. One needs a vector of Gaussian random variables whose covariance decays according to a power law as the lag increases: $\gamma(k) \sim H(2H-1)k^{2H-2}$



# fGn Implemetation

It computes the autocovariance at lags $k$, defining the temporal correlation of the process.

It constructs a Toeplitz matrix, where each descending diagonal has constant values, thus forming a symmetric covariance matrix.

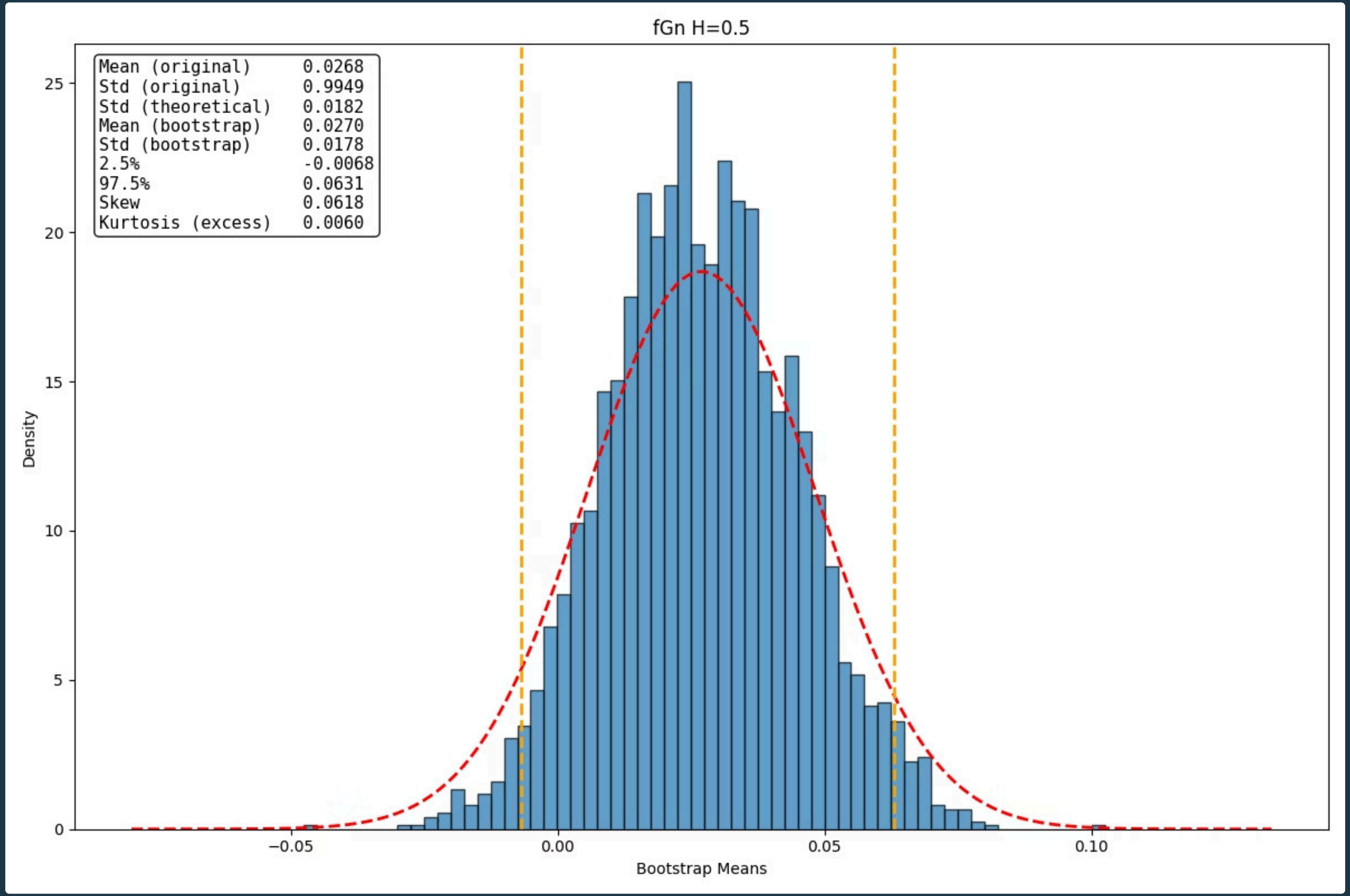
The covariance matrix $\Sigma$ is factorized using the Cholesky decomposition: $\Sigma = L\,L^\top$, with $L$ a lower-triangular matrix.

Fractional Gaussian noise is obtained by multiplying $L$ by a vector $Z$ of independent standard normal variables: $fGn = L\,Z, Z \sim N(0, I)$

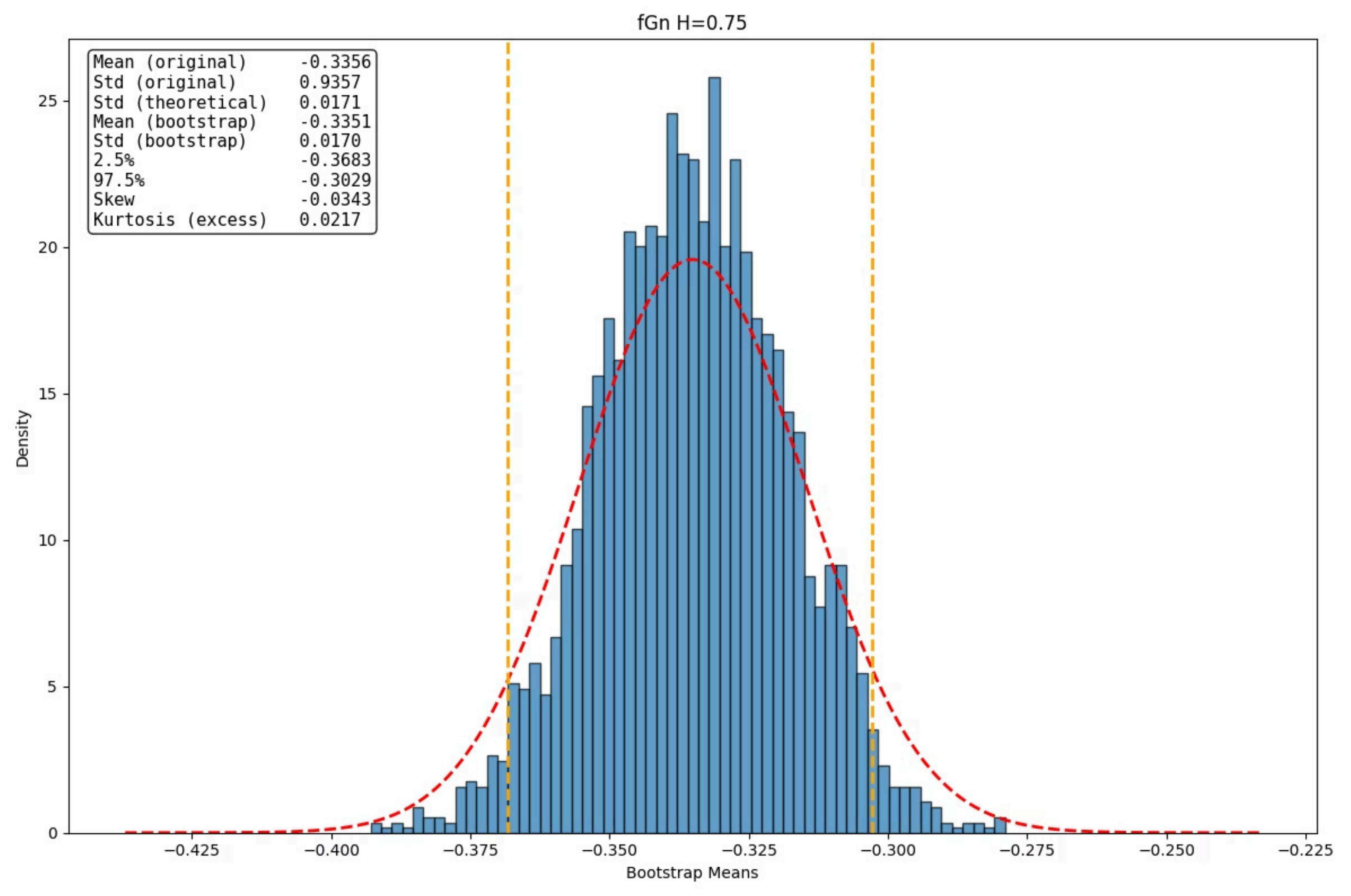
This guarantees that $\mathrm{Cov}(LZ) = L\,\mathrm{Cov}(Z)\,L^\top = LIL^\top = \Sigma$

In this way, uncorrelated white noise is transformed into a correlated process that exactly reproduces the covariance structure imposed by the Hurst exponent.

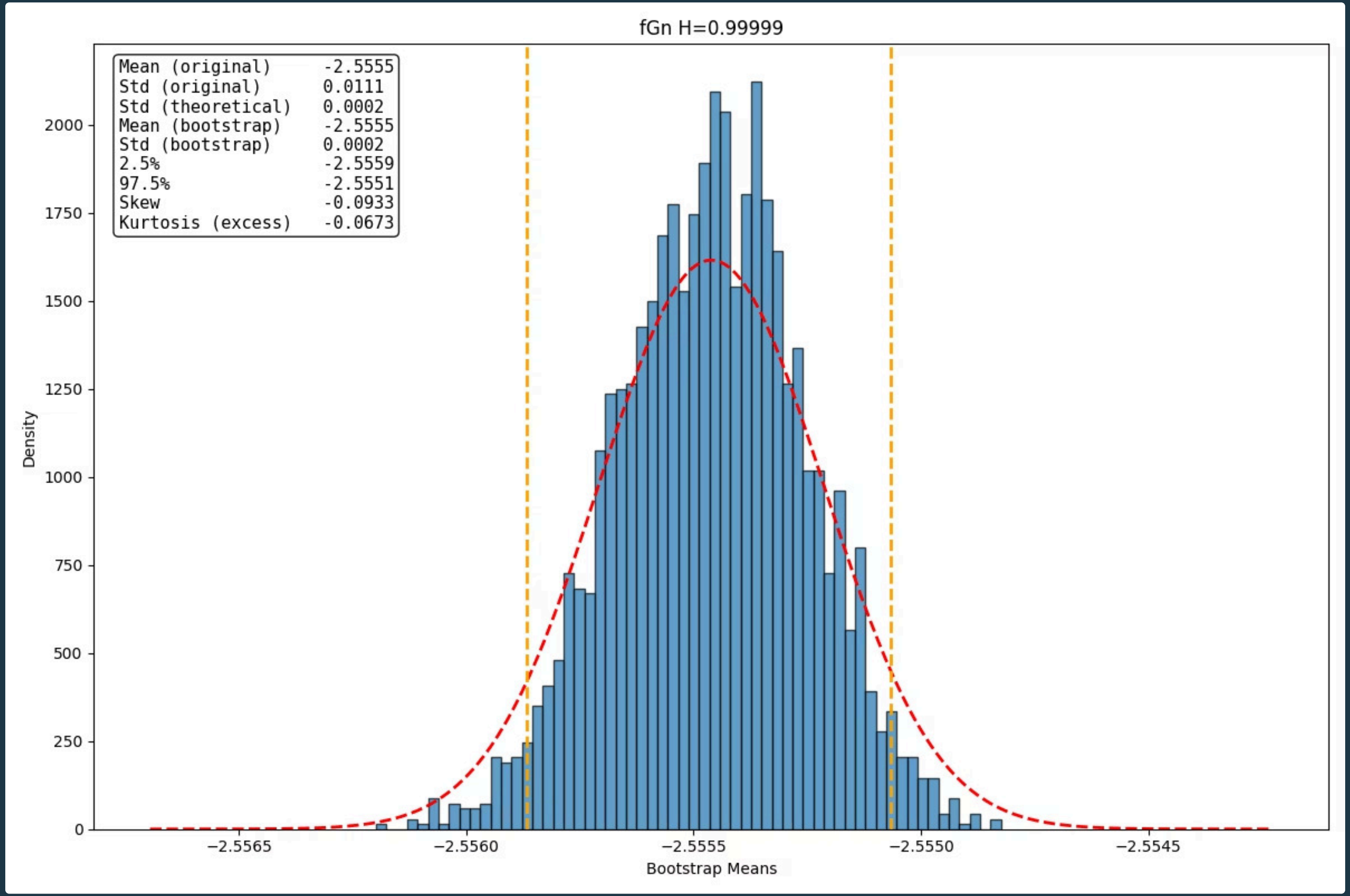
# Bootstrap and Fractional Gaussian Noise



# Bootstrap and Fractional Gaussian Noise



# Bootstrap and Fractional Gaussian Noise



# Empirical Coverage and Confidence Intervals

## Empirical Coverage

Measures the frequency with which a confidence interval contains the true parameter in repeated sampling.

$$\hat{c} = \frac{1}{R} \sum_{r=1}^{R} \mathbf{1}\{\theta \in CI_r\}$$

The method is well-calibrated if the empirical coverage $\hat{c}$ is close to the nominal level $\Pr(\theta \in CI_{pct}) \approx 1 - \alpha$.

## Percentile Bootstrap

From each dataset, $B$ bootstrap replicates $\{\hat{\theta}^{(b)}\}_{b=1}^{B}$ and the empirical bootstrap CDF $\hat{F}_{boot}(t)$ are obtained.

$$\hat{F}_{boot}(t) = \frac{1}{B} \sum_{b=1}^{B} \mathbf{1}\{\hat{\theta}^{(b)} \leq t\}$$

The percentile confidence interval at level $1 - \alpha$ is:

$$CI_{pct} = [\hat{F}_{boot}^{-1}(\alpha/2), \hat{F}_{boot}^{-1}(1 - \alpha/2)]$$

# Reliability of Bootstrap Percentiles

Confidence interval calibration is assessed by evaluating the bootstrap CDF at the true parameter for each replication:

$$u_r = F_{\text{boot},r}(\theta) = \frac{1}{B}\sum_{b=1}^{B} 1\{\hat{\theta}^{(b)} \leq \theta\}$$

If the bootstrap correctly approximates the sampling distribution, the values $u_1, \ldots, u_R$ should be approximately $\text{Uniform}(0, 1)$. Deviations from uniformity indicate that the confidence intervals may not have the desired nominal coverage.

# Cramér–von Mises Statistic

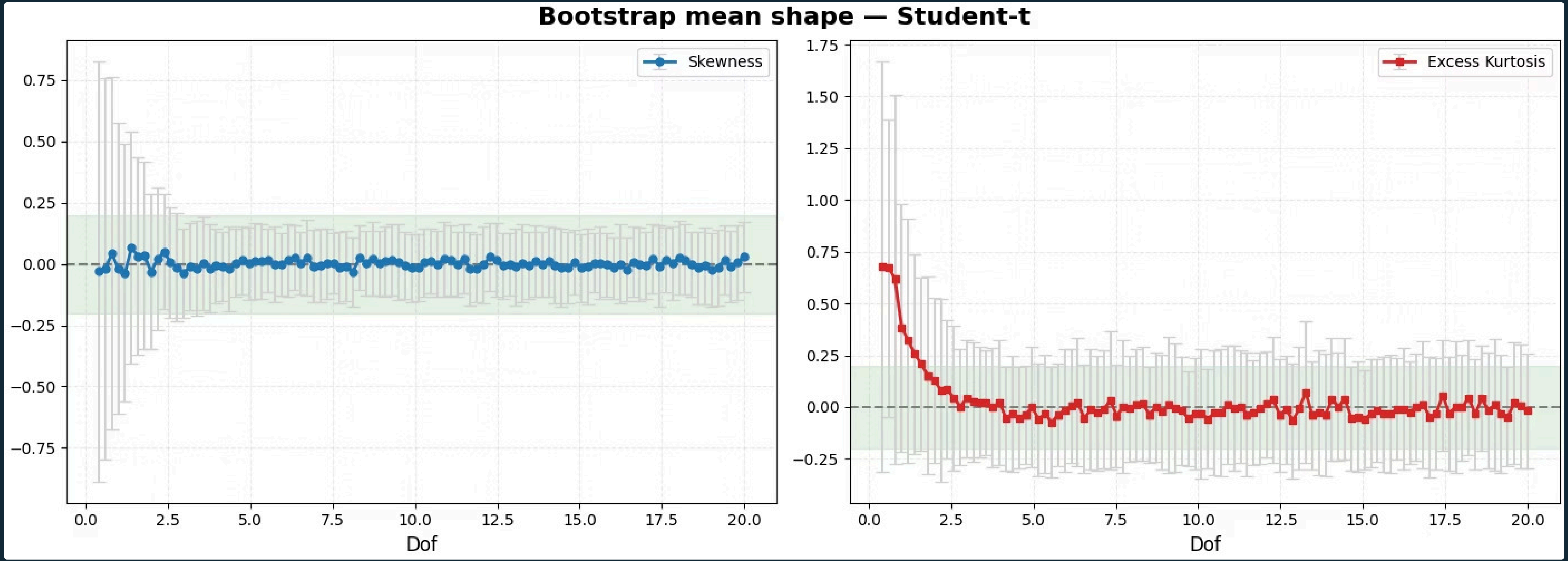
Uniformity is quantified using the Cramér–von Mises statistic:

$$W^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(u_{(i)} - \frac{2i-1}{2n}\right)^2$$

Where $u_{(i)}$ are the ordered values of $u_r$ and $n$ is the number of replications. Large values of $W^2$ signal significant deviations from uniformity.

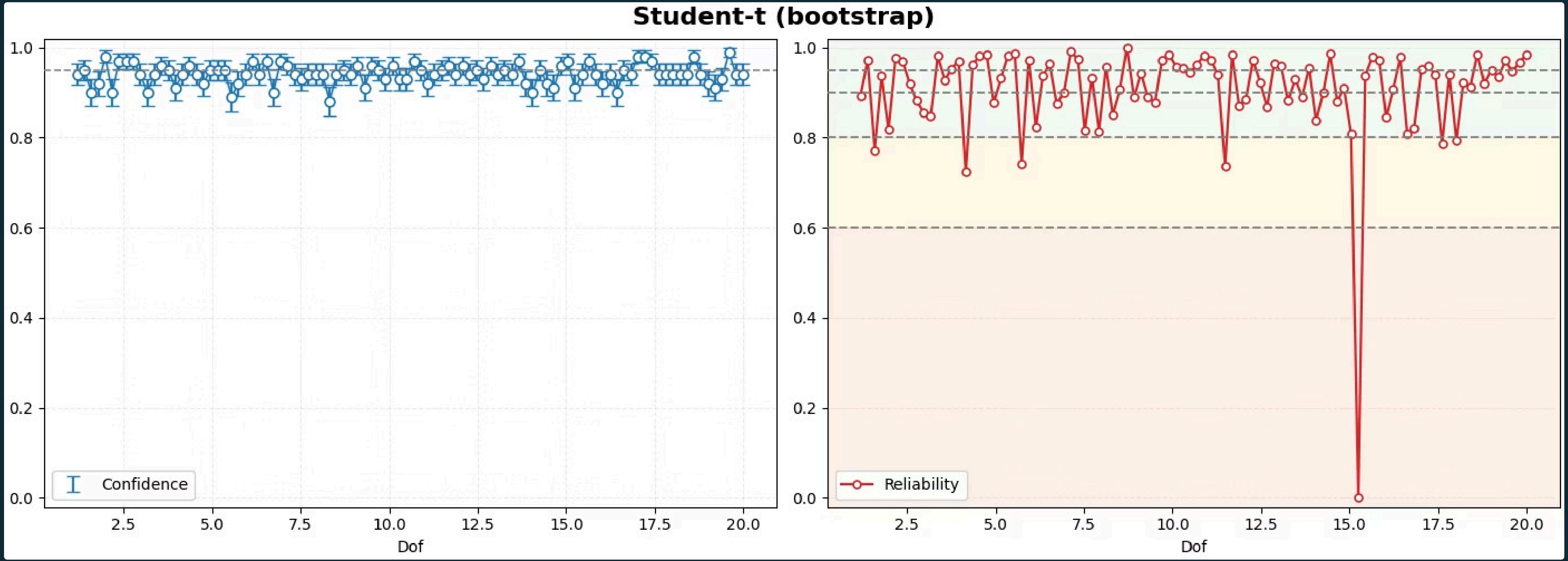
# Results

# Student's t: Skewness and Kurtosis



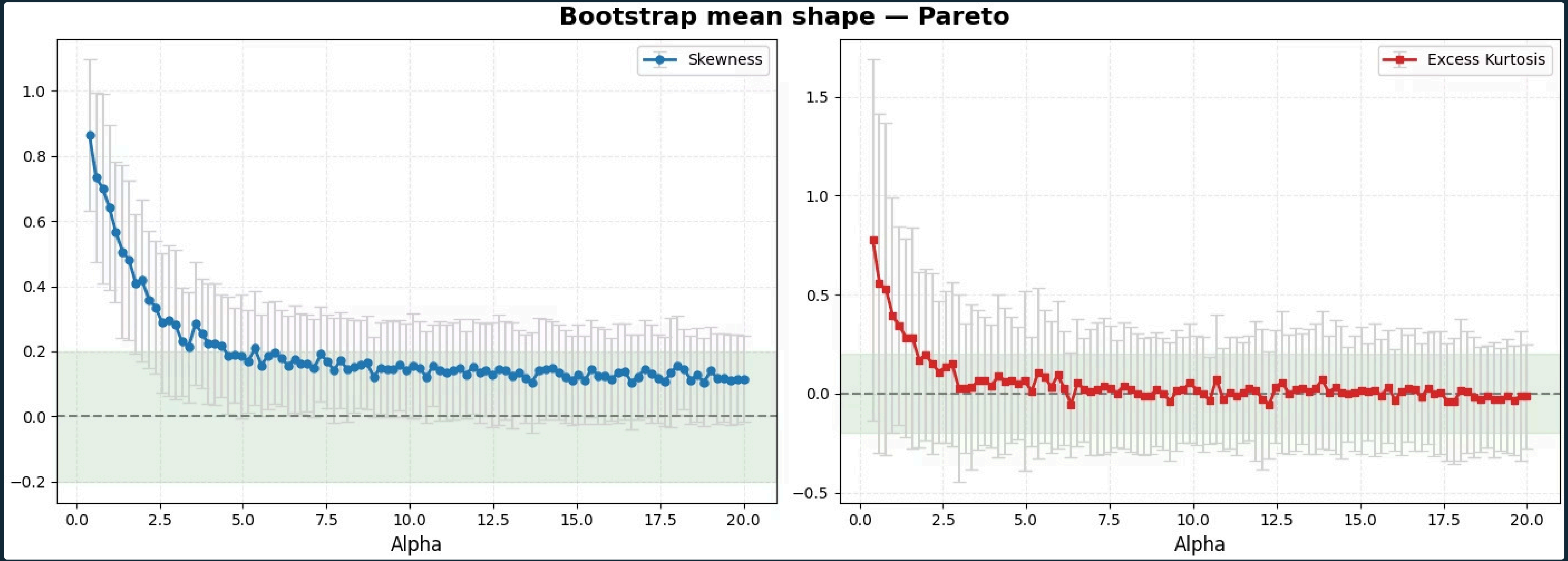

For low degrees of freedom, skewness fluctuates around zero due to the symmetry of the original data, while heavy tails keep kurtosis consistently above it; both eventually stabilize.

# Student's t: Empirical Coverage



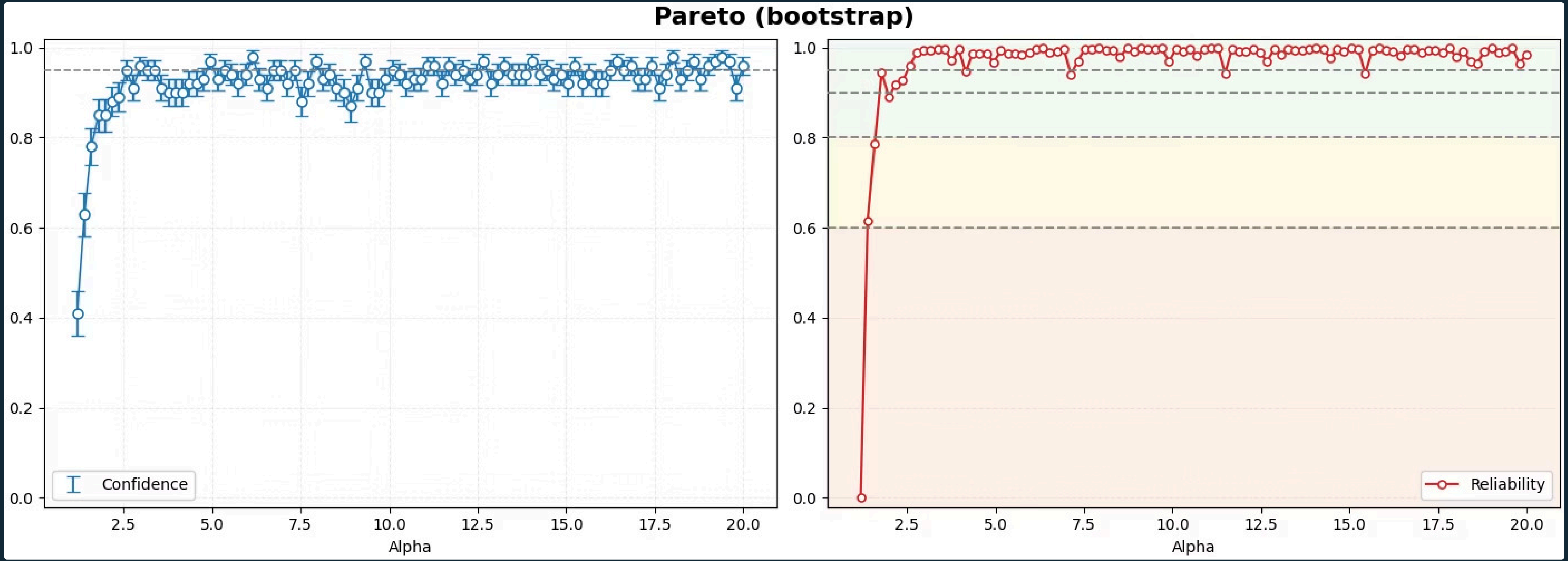

Coverage appears close to nominal for all degrees of freedom, but reliability fluctuates wildly even for large df, likely due to the distribution's symmetry and heavy tails.

# Pareto: Skewness and Kurtosis



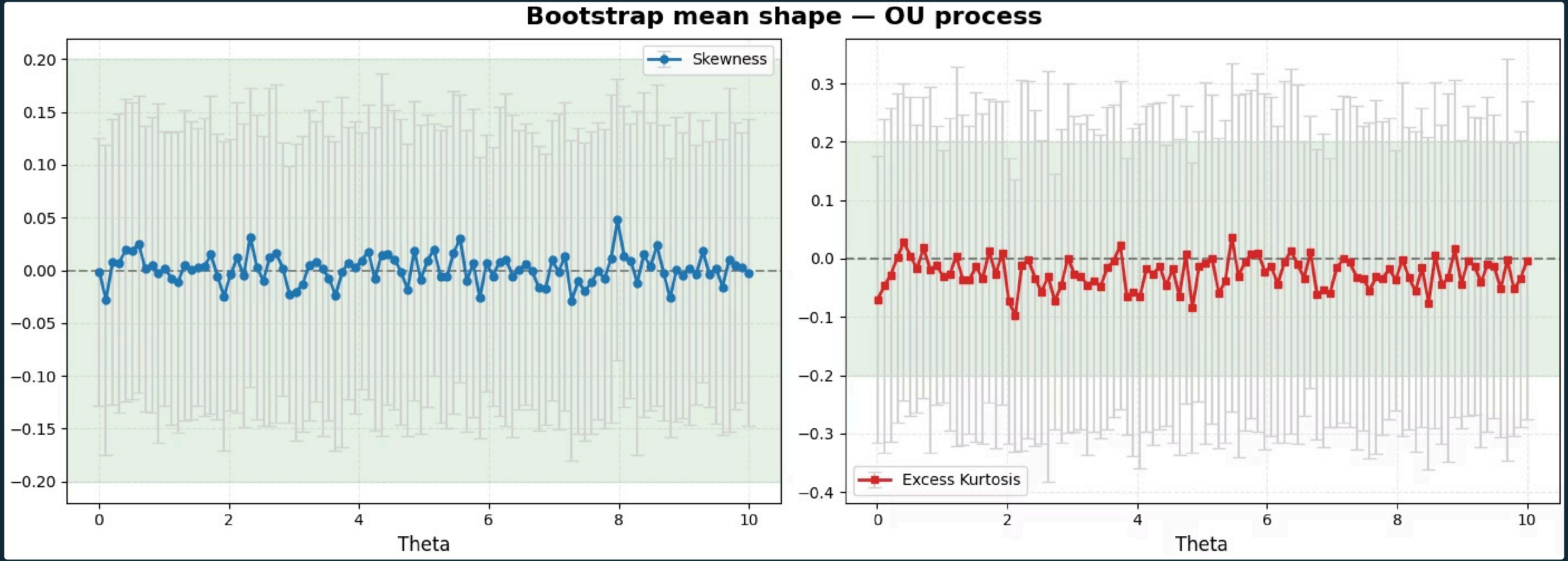

Kurtosis starts high and decreases toward zero as extremes are averaged out for larger α, while skewness remains positive due to the distribution's inherent right skew.

# Pareto: Empirical Coverage



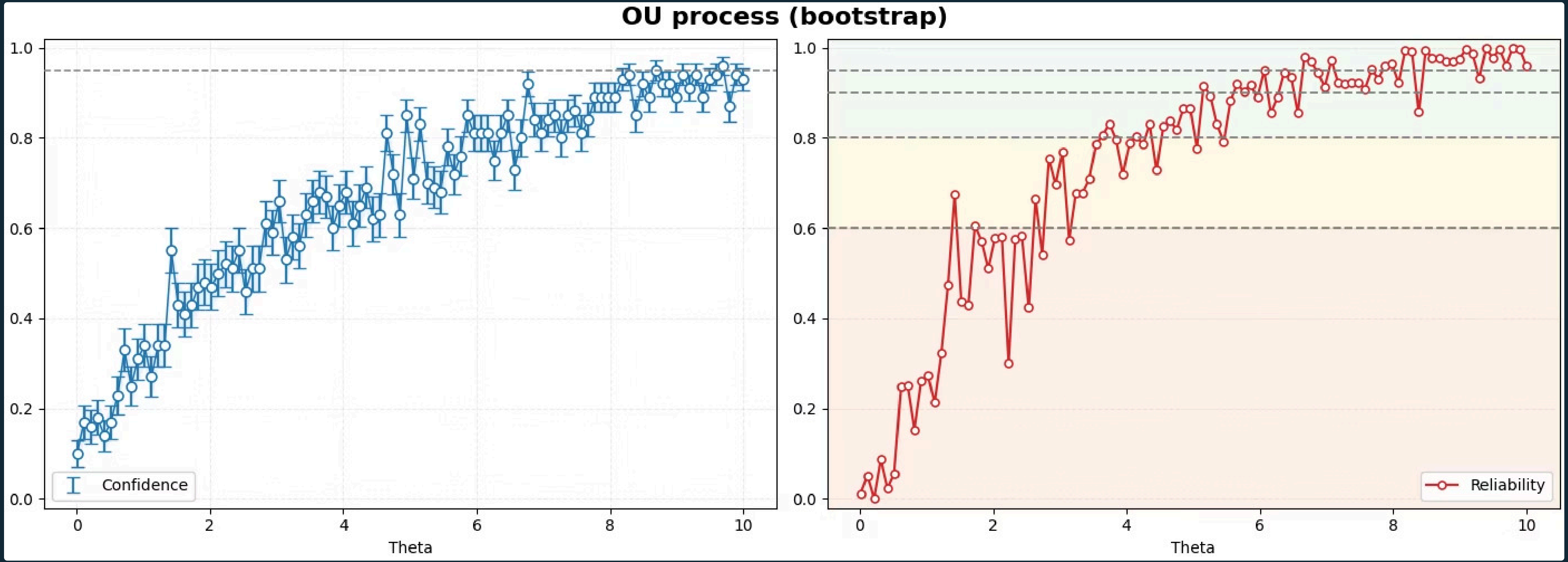

Coverage and reliability are poor for small α due to heavy-tailed extremes destabilizing bootstrap estimates, but improve as α increases and tails lighten.

# Ornstein-Uhlenbeck: Skewness and Kurtosis



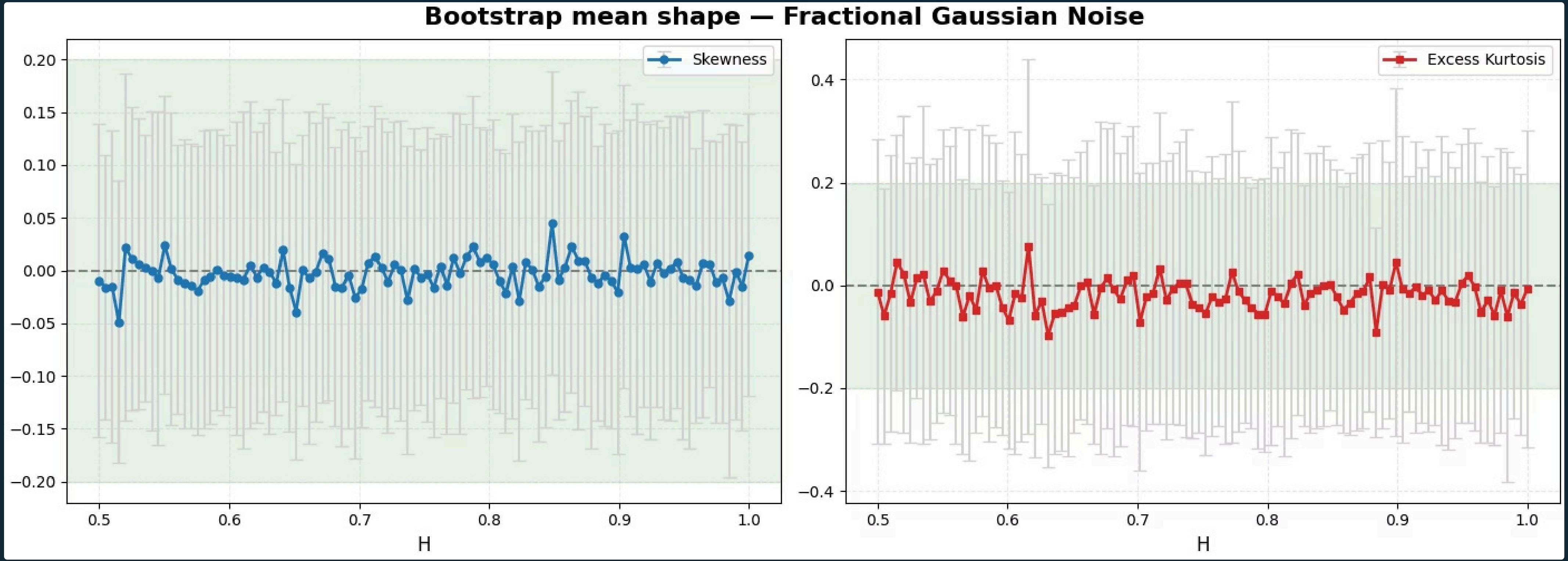

Skewness and kurtosis fluctuate around zero because the OU process is Gaussian; finite samples induce small bootstrap deviations, not genuine non-Gaussianity.

# Ornstein-Uhlenbeck: Empirical Coverage



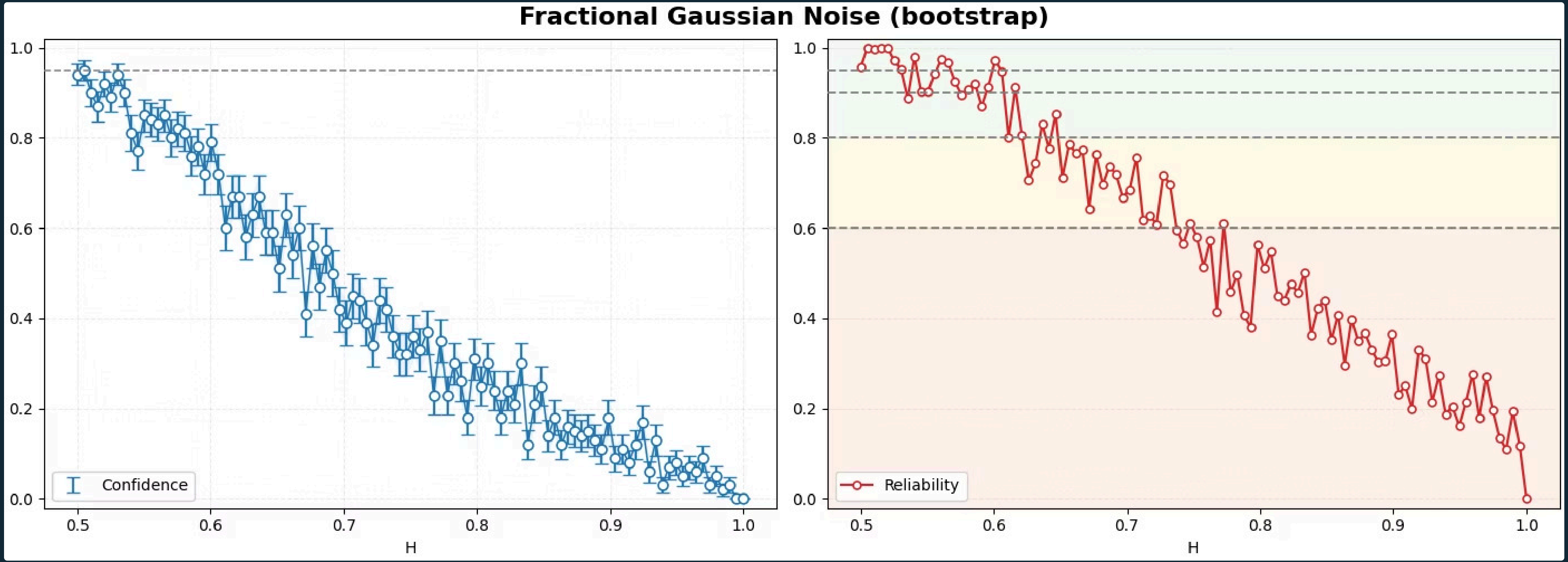

Reliability and coverage increase as memory and persistence decrease; weaker correlations make bootstrap estimates more stable and consistent.

# fGN: Skewness and Kurtosis




For fractional Gaussian noise, as for the Ornstein–Uhlenbeck process, skewness and kurtosis remain close to zero, showing only small fluctuations due to finite-sample effects.

# fGN: Empirical Coverage




Coverage and its reliability decrease as the Hurst coefficient increases. Stronger long-range dependences make bootstrap estimates less reliable.

# Conclusion

## Discussion

Bootstrap samples may retain features of the original distribution, including heavy tails, and estimates' reliability may depend on the original distribution's symmetry.

On autocorrelated data, bootstrap resamples tend to be overly centered and underestimate dispersion.

## Future Improvements

To handle heavy tails and undefined means or variances, alternative estimators robust to outliers, such as the median, can be used.

The use of block bootstrap, on the other hand, can help to preserve the temporal dependence structure in data with memory.

# Thank you