STA 542: Introduction to Time Series Analysis

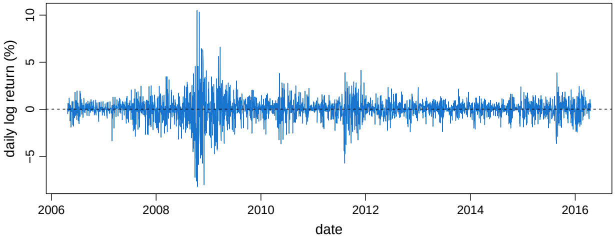
Lecture 7: From an estimated variance to Gaussian inference

Lasse Vuursteen

Duke University

Wed Sep 16, 2026

What would justify a shorter interval for this mean?



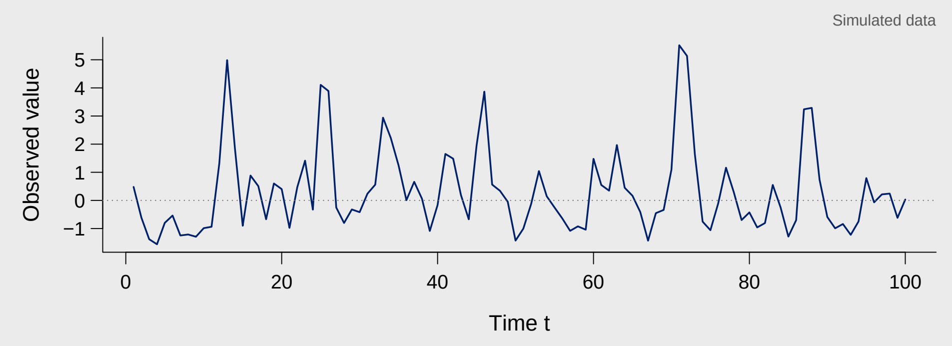
A ratio-consistent variance estimate supports a conservative interval.

- At 95%, its multiplier is about 4.47.
- What additional argument would permit 1.96?

Real data: `astsa::djia`, April 2006–April 2016; 2,517 daily log returns in percent.

Can averages of a skewed process look Gaussian?

Let $E_t \overset{\text{iid}}{\sim} \text{Exponential}(1)$, $Z_t := E_t - 1$, and $X_t := Z_t + 0.7Z_{t-1}$.

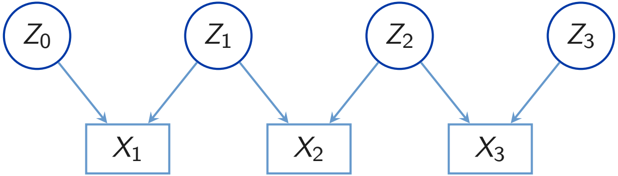


Each observation is non-Gaussian. Adjacent observations share an innovation.

Simulated data: one stationary MA(1) path, $T = 100$; centered exponential innovations.

Which observations share an innovation?

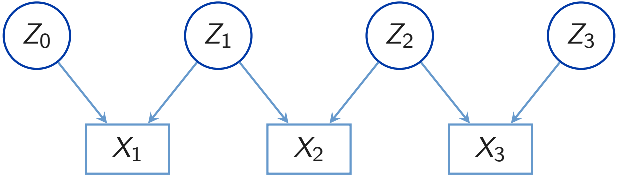
The rule is $X_t = Z_t + 0.7Z_{t-1}$, with independent innovations.



Which pairs of observations are independent?

Which observations share an innovation?

The rule is $X_t = Z_t + 0.7Z_{t-1}$, with independent innovations.



Which pairs of observations are independent?

More generally: observations through time s use Z_j with $j \le s$; observations from $s+2$ onward use Z_j with $j \ge s+1$.

A fixed gap can separate independent groups

For a nonnegative integer m, a process is *m-dependent* if, for every integer s,

$$(\ldots, X_{s-1}, X_s) \quad \text{and} \quad (X_{s+m+1}, X_{s+m+2}, \ldots) \quad \text{are independent.}$$

- The moving average just shown is 1-dependent.
- Under weak stationarity, m-dependence gives $\gamma(h) = 0$ for $|h| > m$.

A fixed gap can separate independent groups

For a nonnegative integer m, a process is *m-dependent* if, for every integer s,

$$(\ldots, X_{s-1}, X_s) \quad \text{and} \quad (X_{s+m+1}, X_{s+m+2}, \ldots) \quad \text{are independent.}$$

- The moving average just shown is 1-dependent.
- Under weak stationarity, m-dependence gives $\gamma(h) = 0$ for $|h| > m$.

Does a covariance cutoff give this independence property?

Finite dependence supplies a central limit theorem

Suppose (X_t) is strictly stationary and m-dependent, with mean μ, finite second moment, and $\tau^2 = \sum_{h=-m}^{m} \gamma(h) > 0$.

$$\frac{\sqrt{T}(\bar{X}_T - \mu)}{\tau} \xrightarrow{d} N(0, 1).$$

- Gaussian observations are not an assumption.
- For our MA(1), the covariance calculation gives $\mu = 0$ and $\tau = 1.7$.

Workbook: *A moving average with non-Gaussian innovations — covariance calculation.*

Which distribution does the CLT describe?

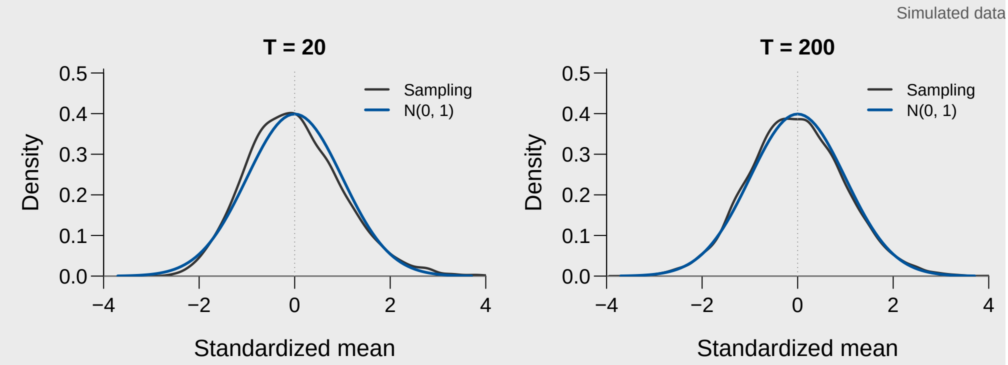
For the centered-exponential MA(1), record

$$U_T := \frac{\sqrt{T}\,\bar{X}_T}{1.7}.$$

1. Generate an entire path of length T.
2. Compute one value of U_T from that path.
3. Repeat with independent paths, keeping the process law fixed.

Would a histogram of the observations within one path answer the same question?

Longer averages are closer to the Gaussian reference



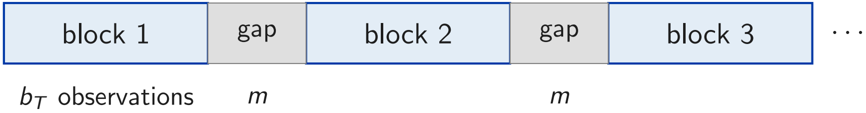
Which features change as T increases? Which distribution stays skewed?

Simulated data: 10,000 independent centered-exponential MA(1) paths at each length; coefficient 0.7, $\tau = 1.7$.

Workbook: *The Moving-Average CLT in Simulation — code and sampling summaries.*

Why leave gaps between long blocks?

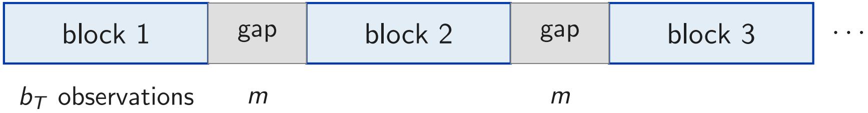
For a strictly stationary m-dependent process, choose $b_T = \lfloor T^{1/3} \rfloor$ and form blocks of b_T observations.



What do the gaps buy us? What must be shown about the observations we discard?

Why leave gaps between long blocks?

For a strictly stationary m-dependent process, choose $b_T = \lfloor T^{1/3} \rfloor$ and form blocks of b_T observations.



What do the gaps buy us? What must be shown about the observations we discard?

The retained block sums are independent. The gaps and final remainder must be negligible on the $\sqrt{T}$ scale.

Discarding the gaps has a vanishing cost

Let D_T contain the gaps and final remainder, and write $R_T := \sum_{t \in D_T} (X_t - \mu)$.

$$\mathrm{Var}\left(\frac{R_T}{\sqrt{T}}\right) \leq (2m+1)\gamma(0)\frac{|D_T|}{T} \longrightarrow 0.$$

- Each discarded observation has at most $2m+1$ nonzero covariance terms.
- The count is $|D_T| = O(T/b_T + b_T)$, so $|D_T|/T \to 0$.

Discarding the gaps has a vanishing cost

Let D_T contain the gaps and final remainder, and write $R_T := \sum_{t \in D_T} (X_t - \mu)$.

$$\mathrm{Var}\left(\frac{R_T}{\sqrt{T}}\right) \le (2m+1)\gamma(0)\frac{|D_T|}{T} \longrightarrow 0.$$

- Each discarded observation has at most $2m+1$ nonzero covariance terms.
- The count is $|D_T| = O(T/b_T + b_T)$, so $|D_T|/T \to 0$.

Chebyshev now gives $R_T/\sqrt{T} \xrightarrow{P} 0$.

No bounded block can dominate the total

For this proof sketch, first suppose $|X_t - \mu| \le M$. Let $B_{j,T}$ be the centered sum in retained block j.

$$\max_j \frac{|B_{j,T}|}{\sqrt{T}} \le \frac{Mb_T}{\sqrt{T}} \longrightarrow 0.$$

The block sums are independent, and their total variance divided by T tends to $\tau^2 > 0$. An independent-array CLT applies.

No bounded block can dominate the total

For this proof sketch, first suppose $|X_t - \mu| \le M$. Let $B_{j,T}$ be the centered sum in retained block j.

$$\max_j \frac{|B_{j,T}|}{\sqrt{T}} \le \frac{Mb_T}{\sqrt{T}} \longrightarrow 0.$$

The block sums are independent, and their total variance divided by T tends to $\tau^2 > 0$. An independent-array CLT applies.

Adding back $R_T/\sqrt{T} \to 0$ preserves the limit. Extending the proof to unbounded observations uses truncation, taken on faith.

Workbook: *The Blocking Argument — optional details.*

What if dependence continues at every lag?

A stationary causal AR(1) with $0 < |\phi| < 1$ and innovation variance $\sigma^2 > 0$ has

$$\gamma(h) = \frac{\sigma^2}{1 - \phi^2} \phi^{|h|}.$$

- No finite covariance cutoff exists.
- Consequently it is not m-dependent for any finite m.

How could a CLT still hold when old innovations keep contributing?

Summable coefficients control distant innovations

Consider a causal linear process

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \qquad \sum_{j=0}^{\infty} |\psi_j| < \infty.$$

Assume (Z_t) is i.i.d., with mean zero and variance $\sigma^2 < \infty$.

- ψ_j is the weight on the innovation j time steps ago.
- A finite truncation is a finite moving average.
- Summability controls the approximation as the truncation grows.

The linear-process limit uses the sum of coefficients

Under the i.i.d., finite-variance, and absolute-summability conditions just stated,

$$\sqrt{T}(\bar{X}_T - \mu) \xrightarrow{d} N(0, \tau^2), \qquad \tau^2 = \sigma^2 \left(\sum_{j=0}^{\infty} \psi_j \right)^2 .$$

- This τ^2 equals the long-run covariance sum.
- If the coefficient sum is zero, the limit is a point mass at zero.
- Dividing by τ requires $\tau^2 > 0$.

Workbook: *A Central Limit Theorem for Causal Linear Processes — result taken on faith.*

The AR(1) is a quick application

For the stationary causal AR(1), $\psi_j = \phi^j$ with $|\phi| < 1$. With centered i.i.d. innovations of variance $\sigma^2 > 0$,

$$\tau^2 = \sigma^2 \left(\sum_{j=0}^{\infty} \phi^j \right)^2 = \frac{\sigma^2}{(1-\phi)^2}.$$

The AR(1) is a quick application

For the stationary causal AR(1), $\psi_j = \phi^j$ with $|\phi| < 1$. With centered i.i.d. innovations of variance $\sigma^2 > 0$,

$$\tau^2 = \sigma^2 \left(\sum_{j=0}^{\infty} \phi^j \right)^2 = \frac{\sigma^2}{(1-\phi)^2}.$$

Where did we require Gaussian innovations? What would be missing if the innovations were only uncorrelated?

A known limiting scale gives an oracle Gaussian interval

Suppose $\sqrt{T}(\bar{X}_T - \mu)/\tau \xrightarrow{d} N(0,1)$, with $\tau > 0$ known. Let z_p be the standard normal p-quantile.

$$C_T^{\mathrm{oracle}} := \left[\bar{X}_T - z_{1-\alpha/2}\frac{\tau}{\sqrt{T}}, \quad \bar{X}_T + z_{1-\alpha/2}\frac{\tau}{\sqrt{T}}\right], \qquad 0 < \alpha < 1.$$

At 95%, $z_{0.975} \approx 1.96$.

What event in the CLT statistic corresponds to coverage?

Coverage becomes a standard normal probability

The interval covers exactly when

$$\mu \in C_T^{\text{oracle}} \quad \Longleftrightarrow \quad \left| \frac{\sqrt{T}(\bar{X}_T - \mu)}{\tau} \right| \leq z_{1-\alpha/2}.$$

The CLT and continuity of the normal CDF give $\mathbb{P}\{\mu \in C_T^{\text{oracle}}\} \to 1 - \alpha$.

Coverage becomes a standard normal probability

The interval covers exactly when

$$\mu \in C_T^{\text{oracle}} \quad \Longleftrightarrow \quad \left| \frac{\sqrt{T}(\bar{X}_T - \mu)}{\tau} \right| \leq z_{1-\alpha/2}.$$

The CLT and continuity of the normal CDF give $\mathbb{P}\{\mu \in C_T^{\text{oracle}}\} \to 1 - \alpha$.

Even the oracle interval is generally asymptotic. Knowing the variance does not make a non-Gaussian mean Gaussian.

Can the same path supply the denominator?

The uncertainty-estimation workbook supplied routes to a nonnegative estimate satisfying

$$\widehat{\tau}_T^2 \xrightarrow{P} \tau^2 > 0.$$

Known-range estimation or Bartlett consistency can provide this conclusion.

- What happens to the CLT when τ is replaced by $\sqrt{\widehat{\tau}_T^2}$?
- Must the estimated scale be independent of the sample mean?
- How fast must the scale estimate converge?

A factor converging to one preserves the distributional limit

Slutsky's theorem gives

$$U_T \xrightarrow{d} U, \qquad V_T \xrightarrow{P} 1 \quad \Longrightarrow \quad U_T V_T \xrightarrow{d} U.$$

- U_T and V_T may be dependent.
- No particular rate is required for $V_T \to 1$.

Which factor should approach a normal limit, and which should approach one?

Studentization uses the estimated standard error

Let $\widehat{\tau}_T := \sqrt{\widehat{\tau}_T^2}$, where $\widehat{\tau}_T^2 \geq 0$ and $\widehat{\tau}_T^2 \xrightarrow{P} \tau^2 > 0$. Define

$$S_T := \begin{cases} \sqrt{T}(\bar{X}_T - \mu)/\widehat{\tau}_T, & \widehat{\tau}_T > 0, \\ 0, & \widehat{\tau}_T = 0. \end{cases}$$

Consistency at a positive limit gives $\mathbb{P}(\widehat{\tau}_T = 0) \to 0$. The assigned value on that vanishing event is a convention for the statistic.

The CLT and scale consistency combine in one product

Suppose $\sqrt{T}(\bar{X}_T - \mu)/\tau \xrightarrow{d} N(0,1)$. On $\{\widehat{\tau}_T > 0\}$,

$$S_T = \underbrace{\frac{\sqrt{T}(\bar{X}_T - \mu)}{\tau}}_{\xrightarrow{d} N(0,1)} \underbrace{\frac{\tau}{\widehat{\tau}_T}}_{\xrightarrow{P} 1} \xrightarrow{d} N(0,1).$$

- Define the second factor to be one on the zero event.
- The displayed equality can fail only on an event whose probability vanishes.
- Slutsky therefore gives the limit for the defined S_T.

The Gaussian interval is now computable from one path

Under the CLT with $\tau^2 > 0$ and a nonnegative consistent estimate $\widehat{\tau}_T^2$, use

$$C_T^{\text{feasible}} := \left[\bar{X}_T - z_{1-\alpha/2} \frac{\widehat{\tau}_T}{\sqrt{T}}, \quad \bar{X}_T + z_{1-\alpha/2} \frac{\widehat{\tau}_T}{\sqrt{T}} \right].$$

- The standard error is $\widehat{\tau}_T / \sqrt{T}$.
- A zero estimate gives the interval $\{\bar{X}_T\}$.
- Its coverage must be checked from the interval itself.

The zero-estimate event does not change limiting coverage

When $\widehat{\tau}_T > 0$, coverage is equivalent to $|S_T| \le z_{1-\alpha/2}$. Hence

$$\left| \mathbb{P}\{\mu \in C_T^{\text{feasible}}\} - \mathbb{P}\{|S_T| \le z_{1-\alpha/2}\} \right| \le \mathbb{P}(\widehat{\tau}_T = 0) \longrightarrow 0.$$

The studentized limit gives $\mathbb{P}\{\mu \in C_T^{\text{feasible}}\} \to 1 - \alpha$.

The zero-estimate event does not change limiting coverage

When $\widehat{\tau}_T > 0$, coverage is equivalent to $|S_T| \le z_{1-\alpha/2}$. Hence

$$\left| \mathbb{P}\{\mu \in C_T^{\text{feasible}}\} - \mathbb{P}\{|S_T| \le z_{1-\alpha/2}\} \right| \le \mathbb{P}(\widehat{\tau}_T = 0) \longrightarrow 0.$$

The studentized limit gives $\mathbb{P}\{\mu \in C_T^{\text{feasible}}\} \to 1 - \alpha$.

The estimated scale and Gaussian shape are both asymptotic approximations.

The same standard error supports two different multipliers

At 95%, both feasible intervals use the same estimated standard error. When that standard error is positive,

$$\frac{\text{Gaussian width}}{\text{Chebyshev width}} = \frac{z_{0.975}}{1/\sqrt{0.05}} \approx 0.438.$$

- Ratio consistency gives the conservative limiting lower bound.
- Adding a CLT gives Gaussian coverage tending to 95%.

The shorter interval uses an additional distributional conclusion.

Return to the variance estimator from the MA(2) exercise

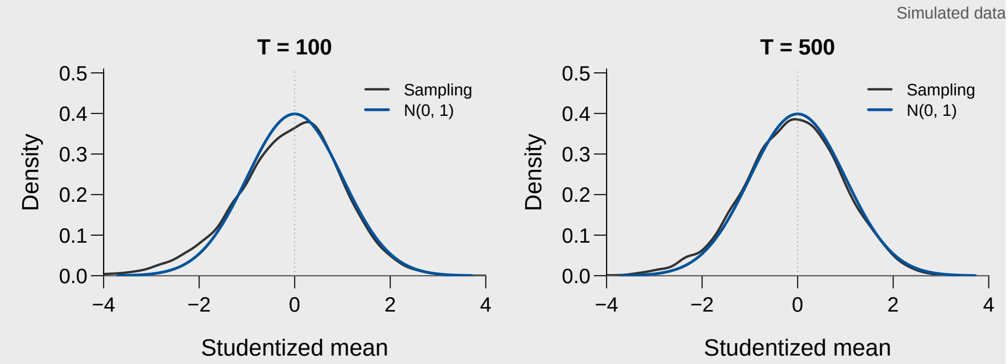
Use $X_t = \mu + Z_t + 0.5Z_{t-1} + 0.25Z_{t-2}$, with $Z_t = E_t - 1$ and i.i.d. $E_t \sim \text{Exponential}(1)$. Its known-range estimate is

$$\widehat{q}_T := \max\left\{\widehat{\gamma}_T(0) + 2\left(1 - \frac{1}{T}\right)\widehat{\gamma}_T(1) + 2\left(1 - \frac{2}{T}\right)\widehat{\gamma}_T(2),\ 0\right\}.$$

- The estimated standard error is $\sqrt{\widehat{q}_T / T}$.
- Strict stationarity and 2-dependence supply the CLT.
- Covariance learning gives $\widehat{q}_T \xrightarrow{P} \tau^2 = 49/16$.

Workbook: *Two multipliers for one estimated standard error — derivations and code.*

Estimating the scale affects the finite-sample shape



What differences from the normal reference remain at each record length?

Simulated data: the preceding MA(2) with $\mu = 0$; 5,000 independent paths per length.

Workbook: *Two multipliers for one estimated standard error — sampling summaries.*

Known and estimated scales give different finite-sample coverage

Coverage (%), with Monte Carlo standard errors in percentage points in parentheses:

Method	$T = 100$	$T = 500$
Oracle Gaussian	94.82 (0.31)	94.32 (0.33)
Feasible Gaussian	91.90 (0.39)	94.24 (0.33)
Feasible Chebyshev	99.84 (0.06)	100.00 (0.00)

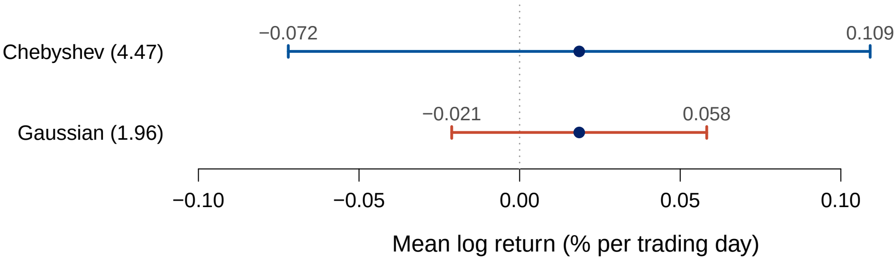
Which comparison isolates the remaining Gaussian-shape approximation?

The oracle uses the exact variance q_T/T, with $q_T = 49/16 - 9/(4T)$. Both feasible rules use the same estimated variance.

Simulated data: the same 5,000 MA(2) paths at each length. Zero-width intervals are retained; an observed 100% is not a population guarantee.

The two multipliers give different intervals for the same mean

Use the same Bartlett estimate with bandwidth $H = 20$ for both intervals.



The Gaussian calibration requires both a CLT and a consistent positive scale. These calculations do not verify those assumptions for the return record.

Real data: `astsa::djia`, April 2006–April 2016; 2,517 daily log returns in percent.

Which series could have one common mean and covariance structure?

Our inference arguments started with a stationary model for the observed record.

Trends, seasonal patterns, and changing spread can challenge that premise.

- Which adjustment addresses which feature of a record?
- What assumptions would make the adjusted series stationary?

Workbook 8 studies this filtering question.