STA 542: Introduction to Time Series Analysis

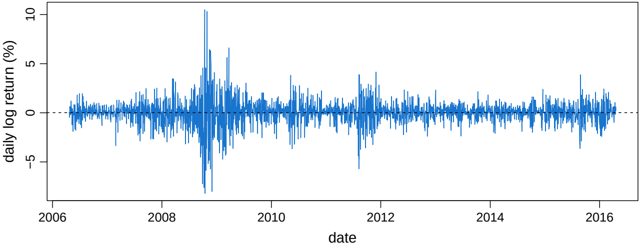
Lecture 5: Learning covariances from one path

Lasse Vuursteen

Duke UNIVERSITY

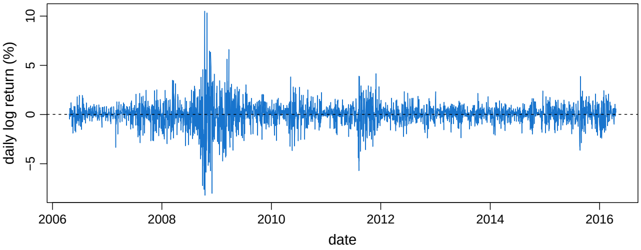
Wed Sep 9, 2026

How precisely does one record determine a mean?



We have 2,517 daily returns. Suppose we model them by a weakly stationary process with mean μ.

How precisely does one record determine a mean?



We have 2,517 daily returns. Suppose we model them by a weakly stationary process with mean μ.

- What determines the error of their sample mean?
- Is the marginal variance enough?

Real data: `astsa::djia`, April 2006–April 2016; daily log returns in percent. Figure reused from Lecture 2.

Consistency controls the probability of a fixed error

Let θ be a fixed feature of a process law, and let $\widehat{\theta}_T$ be a statistic computed from $X_1, \ldots, X_T$.

Consistency

For every $\varepsilon > 0$,

$$\mathbb{P}\big(|\widehat{\theta}_T - \theta| > \varepsilon\big) \longrightarrow 0.$$

We write $\widehat{\theta}_T \xrightarrow{P} \theta$.

Consistency controls the probability of a fixed error

Let θ be a fixed feature of a process law, and let $\widehat{\theta}_T$ be a statistic computed from $X_1, \ldots, X_T$.

Consistency

For every $\varepsilon > 0$,

$$\mathbb{P}\big(|\widehat{\theta}_T - \theta| > \varepsilon\big) \longrightarrow 0.$$

We write $\widehat{\theta}_T \xrightarrow{P} \theta$.

How large is the error at the sample size we actually have?

A second moment gives a probability bound

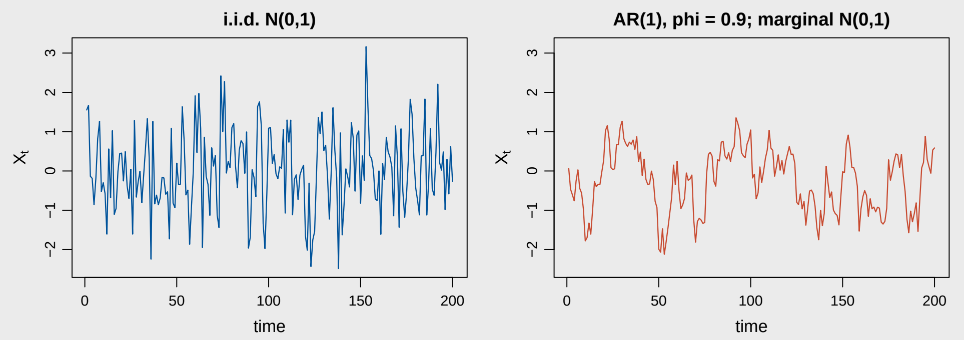
The stronger mean-square convergence criterion used earlier was $\mathbb{E}[(\widehat{\theta}_T - \theta)^2] \to 0$.

$$\mathbb{P}\big(|\widehat{\theta}_T - \theta| > \varepsilon\big) \le \frac{\mathbb{E}[(\widehat{\theta}_T - \theta)^2]}{\varepsilon^2}.$$

- A vanishing numerator implies consistency.
- A *known* numerator gives a bound at each finite T.

For the sample mean, we therefore start by computing its variance.

The marginal distributions agree. Do the mean errors?



Both processes have $X_t \sim N(\mu, 1)$; each path has length $T = 200$.

- Left: i.i.d. observations. Right: stationary Gaussian AR(1), $\phi = 0.9$.
- Which process allows for an easier estimation of the mean? Why?

The variance of a mean uses every covariance pair

Suppose (X_t) is weakly stationary with mean μ and ACVF γ. Write $\bar{X}_T := T^{-1} \sum_{t=1}^{T} X_t$.

$$\begin{aligned}
\operatorname{Var}(\bar{X}_T) &= \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t=1}^{T} \operatorname{Cov}(X_s, X_t) \\
&= \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t=1}^{T} \gamma(t-s).
\end{aligned}$$

Weak stationarity makes each entry depend only on its lag.

How often does each lag occur?

For $T = 4$, the entries below are the lags $t - s$:

$s \backslash t$	1	2	3	4
1	0	1	2	3
2	−1	0	1	2
3	−2	−1	0	1
4	−3	−2	−1	0

How often does each lag occur?

For $T = 4$, the entries below are the lags $t - s$:

$s \backslash t$	1	2	3	4
1	0	1	2	3
2	-1	0	1	2
3	-2	-1	0	1
4	-3	-2	-1	0

- Lag h occurs $T - |h|$ times.
- Positive and negative lags contribute equally because $\gamma(-h) = \gamma(h)$.

The ACVF determines the exact variance of the mean

Exact variance under weak stationarity

$$v_T(\gamma) := \mathrm{Var}(\bar{X}_T) = \frac{1}{T}\left\{\gamma(0) + 2\sum_{h=1}^{T-1}\left(1 - \frac{h}{T}\right)\gamma(h)\right\}.$$

- The factor $1 - h/T$ records how many pairs are available.
- For uncorrelated observations, $v_T(\gamma) = \gamma(0)/T$.
- No Gaussian assumption or limit theorem was used.

Can dependence make a mean more precise?

For the stationary causal AR(1) with $|\phi| < 1$ and marginal variance one, $\gamma(h) = \phi^{|h|}$.

At $T = 50$, order $\phi = -0.6$, 0, and 0.6 by the variance of $\bar{X}_T$.

Can dependence make a mean more precise?

For the stationary causal AR(1) with $|\phi| < 1$ and marginal variance one, $\gamma(h) = \phi^{|h|}$.

At $T = 50$, order $\phi = -0.6$, 0, and 0.6 by the variance of $\bar{X}_T$.

$$v_{50}(\phi = -0.6) < v_{50}(\phi = 0) < v_{50}(\phi = 0.6).$$

Positive correlations add to the variance; alternating correlations can cancel part of it.

The sign and size of the covariance terms determine the effect.

If the ACVF were supplied, we could give an interval

Suppose γ is known, so $v_T(\gamma)$ is known. Fix $0 < \alpha < 1$.

Known-ACVF Chebyshev interval

$$C_T^{\mathrm{known}} := \left[\bar{X}_T - \sqrt{\frac{v_T(\gamma)}{\alpha}},\ \bar{X}_T + \sqrt{\frac{v_T(\gamma)}{\alpha}} \right].$$

Chebyshev gives $\mathbb{P}\{\mu \in C_T^{\mathrm{known}}\} \geq 1 - \alpha$.

If the ACVF were supplied, we could give an interval

Suppose γ is known, so $v_T(\gamma)$ is known. Fix $0 < \alpha < 1$.

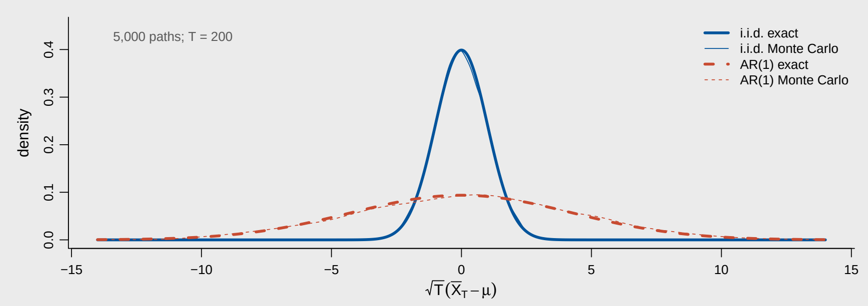
Known-ACVF Chebyshev interval

$$C_T^{\mathrm{known}} := \left[\bar{X}_T - \sqrt{\frac{v_T(\gamma)}{\alpha}},\ \bar{X}_T + \sqrt{\frac{v_T(\gamma)}{\alpha}}\right].$$

Chebyshev gives $\mathbb{P}\{\mu \in C_T^{\mathrm{known}}\} \geq 1 - \alpha$.

At 95%, the half-width is $\sqrt{20}$ times the standard deviation of $\bar{X}_T$.

Equal marginal variances can hide different mean errors



At $T = 200$, $T\,\mathrm{Var}(\bar{X}_T)$ is 1 for i.i.d. data and 18.1 for the stationary AR(1) with $\phi = 0.9$.

The 95% Chebyshev half-widths are 0.316 and 1.345.
Simulated data; 5,000 paths/model, seed 542. Thick: exact Gaussian. Thin: estimated densities.

What if the estimated mean enters a forecast?

Return to the random walk, now with a fixed unknown drift δ:

$$X_t = X_{t-1} + \delta + Z_t, \qquad t \geq 1, \qquad (Z_t) \sim \mathrm{WN}(0, \sigma^2).$$

- $X_0 = x_0$ is fixed; $0 < \sigma^2 < \infty$ is known.
- White noise supplies zero means and zero cross-time covariances.
- Independence and Gaussianity are not assumed.

Write $\mathbb{P}_\delta$ for any fixed law satisfying these conditions.

The average increment estimates the drift

Observe $X_0, \dots, X_T$. The increments $D_t := X_t - X_{t-1}$ satisfy

$$\widehat{\delta}_T := \frac{1}{T}\sum_{t=1}^{T} D_t = \delta + \frac{1}{T}\sum_{t=1}^{T} Z_t.$$

- The levels are nonstationary; the increments are weakly stationary.
- $\mathbb{E}_\delta[\widehat{\delta}_T] = \delta$ and $\mathrm{Var}_\delta(\widehat{\delta}_T) = \sigma^2/T$.

Estimation error remains in the fitted drift even when the variance is known.

Derive the error before choosing an interval

For a chosen integer horizon $h \geq 1$, use the fitted forecast

$$\widehat{X}_{T+h|T} := X_T + h\widehat{\delta}_T.$$

1. Write $X_{T+h} - \widehat{X}_{T+h|T}$ using only the Z_t's.
2. Identify the contribution from future noise.
3. Identify the contribution from estimating δ.

Under white noise alone, this is a forecast rule; we have not identified it as the conditional mean.

There are two sources of forecast error

Use $X_{T+h} = X_T + h\delta + \sum_{j=1}^{h} Z_{T+j}$ and subtract the forecast:

$$\begin{aligned} e_{T,h} &:= X_{T+h} - \widehat{X}_{T+h|T} \\ &= \underbrace{\sum_{j=1}^{h} Z_{T+j}}_{\text{future noise}} - \underbrace{\frac{h}{T}\sum_{t=1}^{T} Z_t}_{\text{drift estimation error}} . \end{aligned}$$

There are two sources of forecast error

Use $X_{T+h} = X_T + h\delta + \sum_{j=1}^{h} Z_{T+j}$ and subtract the forecast:

$$
\begin{aligned}
e_{T,h} &:= X_{T+h} - \widehat{X}_{T+h|T} \\
&= \underbrace{\sum_{j=1}^{h} Z_{T+j}}_{\text{future noise}} - \underbrace{\frac{h}{T}\sum_{t=1}^{T} Z_t}_{\text{drift estimation error}} .
\end{aligned}
$$

The fitted drift is reused h times. Its error is multiplied by h.

Zero cross-time covariances let us add the variances

The two sums in $e_{T,h}$ use disjoint times, so their covariance is zero.

$$\mathrm{Var}_\delta(e_{T,h}) = \underbrace{h\sigma^2}_{\text{future noise}} + \underbrace{\frac{h^2\sigma^2}{T}}_{\text{drift estimation}} .$$

- Also $\mathbb{E}_\delta[e_{T,h}] = 0$, so this is the mean squared forecast error.
- No independence between the sums is needed for this calculation.

Zero cross-time covariances let us add the variances

The two sums in $e_{T,h}$ use disjoint times, so their covariance is zero.

$$\mathrm{Var}_\delta(e_{T,h}) = \underbrace{h\sigma^2}_{\text{future noise}} + \underbrace{\frac{h^2\sigma^2}{T}}_{\text{drift estimation}} .$$

- Also $\mathbb{E}_\delta[e_{T,h}] = 0$, so this is the mean squared forecast error.
- No independence between the sums is needed for this calculation.

If the drift were known, which term would disappear?

A prediction interval can include estimation uncertainty

Fix $0 < \alpha < 1$. Chebyshev applied to the full forecast error gives

$$I_{T,h} := \left[\widehat{X}_{T+h|T} - \sqrt{\frac{\sigma^2(h + h^2/T)}{\alpha}},\ \widehat{X}_{T+h|T} + \sqrt{\frac{\sigma^2(h + h^2/T)}{\alpha}} \right].$$

For every fixed δ and every law satisfying the moment assumptions,
$\mathbb{P}_\delta\{X_{T+h} \in I_{T,h}\} \geq 1 - \alpha$.

A prediction interval can include estimation uncertainty

Fix $0 < \alpha < 1$. Chebyshev applied to the full forecast error gives

$$I_{T,h} := \left[\widehat{X}_{T+h|T} - \sqrt{\frac{\sigma^2(h + h^2/T)}{\alpha}},\ \widehat{X}_{T+h|T} + \sqrt{\frac{\sigma^2(h + h^2/T)}{\alpha}}\right].$$

For every fixed δ and every law satisfying the moment assumptions, $\mathbb{P}_\delta\{X_{T+h} \in I_{T,h}\} \geq 1 - \alpha$.

The target is now a future observation. The drift remains fixed.

Which experiment does this prediction guarantee describe?

Repeat the observed history *and* its future under one fixed law.

$$\mathbb{P}_\delta\{X_{T+h} \in I_{T,h}\} \geq 1 - \alpha \qquad \text{for each chosen } h.$$

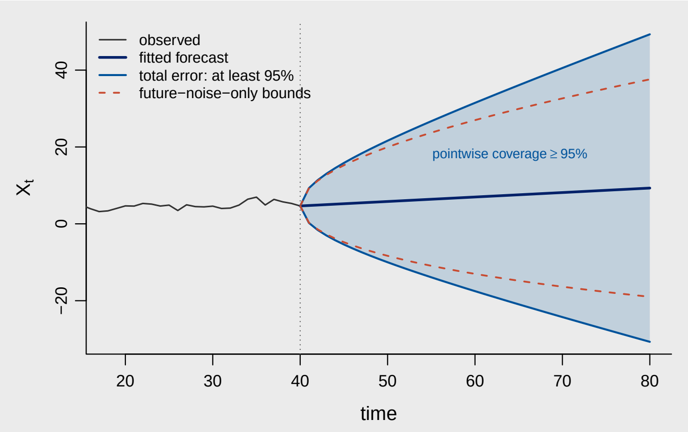
- Marginal coverage; no conditional-on-history guarantee is claimed.
- One horizon at a time; no simultaneous future-path guarantee is claimed.
- The bound uses known σ^2. Plugging in an estimate needs a new argument.

When does estimating the drift matter?

Estimation variance divided by future-noise variance is h/T.

At $h = T$, the two contributions are equal.

The half-width is then $\sqrt{2}$ times the future-noise-only half-width.



Simulated history: $T = 40$, $\delta = 0.2$, $\sigma^2 = 1$; i.i.d. Gaussian innovations, seed 542. Dashed bounds omit estimation error.

Usually the ACVF must be learned from the same path

We have treated γ as supplied. From $X_1, \ldots, X_T$, we instead compute

Sample ACVF

$$\widehat{\gamma}_T(h) := \frac{1}{T}\sum_{t=1}^{T-h}(X_t - \bar{X}_T)(X_{t+h} - \bar{X}_T), \qquad 0 \le h < T.$$

Use $\widehat{\gamma}_T(-h) = \widehat{\gamma}_T(h)$ and divisor T, not $T - h$.

Usually the ACVF must be learned from the same path

We have treated γ as supplied. From $X_1, \ldots, X_T$, we instead compute

Sample ACVF

$$\widehat{\gamma}_T(h) := \frac{1}{T} \sum_{t=1}^{T-h} (X_t - \bar{X}_T)(X_{t+h} - \bar{X}_T), \qquad 0 \leq h < T.$$

Use $\widehat{\gamma}_T(-h) = \widehat{\gamma}_T(h)$ and divisor T, not $T - h$.

If the mean is consistently estimated, must these covariances be consistent too?

A stationary process can retain one random scale forever

Draw $\Lambda \in \{1, 2\}$ with equal probabilities, once for the entire path. Independently, let (R_t) be i.i.d. with values -1 and 1 equally likely.

$$X_t := \Lambda R_t, \qquad t \in \mathbb{Z}.$$

- The scale is part of one fixed process law.
- A time shift preserves the joint distribution: the process is strictly stationary.

A stationary process can retain one random scale forever

Draw $\Lambda \in \{1, 2\}$ with equal probabilities, once for the entire path. Independently, let (R_t) be i.i.d. with values -1 and 1 equally likely.

$$X_t := \Lambda R_t, \qquad t \in \mathbb{Z}.$$

- The scale is part of one fixed process law.
- A time shift preserves the joint distribution: the process is strictly stationary.

What do the averages of X_t and X_t^2 learn?

The population variance averages over both scales

For $X_t = \Lambda R_t$, independence and the zero means of the R_t's give

$$\mathbb{E}[X_t] = 0, \qquad \gamma(0) = \mathbb{E}[\Lambda^2] = \frac{5}{2}, \qquad \gamma(h) = 0 \quad (h \neq 0).$$

The population variance averages over both scales

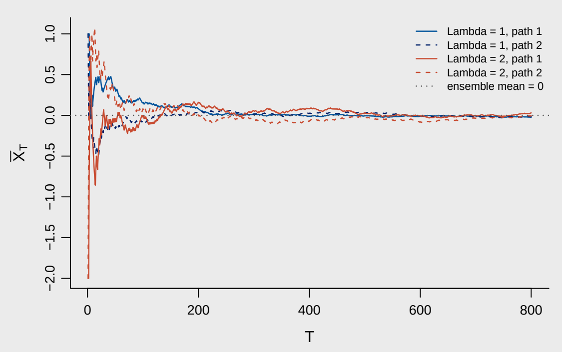
For $X_t = \Lambda R_t$, independence and the zero means of the R_t's give

$$\mathbb{E}[X_t] = 0, \qquad \gamma(0) = \mathbb{E}[\Lambda^2] = \frac{5}{2}, \qquad \gamma(h) = 0 \quad (h \neq 0).$$

Hence $\mathrm{Var}(\bar{X}_T) = 5/(2T) \to 0$.

The sample mean converges to zero in mean square. Does averaging the squares work as well?

The running means approach their population target



Both realized scales give the same limiting mean, zero.

Simulated data: two paths at each scale, $T = 800$, seed 542.

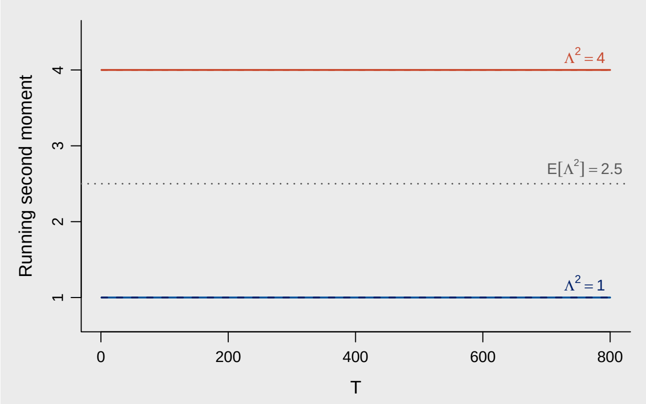
The squared observations retain the realized scale

Every observation satisfies $X_t^2 = \Lambda^2$.

Consequently,

$$\widehat{\gamma}_T(0) = \Lambda^2 - \bar{X}_T^2 \xrightarrow{P} \Lambda^2.$$

The limit is 1 or 4; the population variance is $5/2$.



Same simulated paths; the second moments coincide within each scale.

Which time averages does a covariance require?

Since $\gamma(h) = \mathbb{E}[X_0 X_h] - \mu^2$, we need more than the mean.

Quantity to learn	Function of a window	What we average
μ	$g(x) = x$	X_t
$\mathbb{E}[X_0^2]$	$g(x) = x^2$	X_t^2
$\mathbb{E}[X_0 X_h]$	$g_h(x_0, \ldots, x_h) = x_0 x_h$	$X_t X_{t+h}$

Which time averages does a covariance require?

Since $\gamma(h) = \mathbb{E}[X_0 X_h] - \mu^2$, we need more than the mean.

Quantity to learn	Function of a window	What we average
μ	$g(x) = x$	X_t
$\mathbb{E}[X_0^2]$	$g(x) = x^2$	X_t^2
$\mathbb{E}[X_0 X_h]$	$g_h(x_0, \ldots, x_h) = x_0 x_h$	$X_t X_{t+h}$

Each row needs a convergence statement for its own time average.

We ask for convergence for the function we use

Let (X_t) be strictly stationary. Fix a nonnegative integer r and $g : \mathbb{R}^{r+1} \to \mathbb{R}$ with $\mathbb{E}[|g(X_0, \ldots, X_r)|] < \infty$.

g-ergodicity

$$\frac{1}{T-r} \sum_{t=1}^{T-r} g(X_t, \ldots, X_{t+r}) \xrightarrow{P} \mathbb{E}[g(X_0, \ldots, X_r)].$$

- Strict stationarity gives every shifted window the same expectation.
- r and g stay fixed while T grows.
- The condition concerns this g; strict stationarity alone does not imply it.

For the ACVF, specify exactly the averages we need

Let (X_t) be weakly stationary; fix a nonnegative integer m.
We call it *ACVF-learnable through lag m* if

$$\bar{X}_T \xrightarrow{P} \mu, \qquad \frac{1}{T}\sum_{t=1}^{T-h} X_t X_{t+h} \xrightarrow{P} \mathbb{E}[X_0 X_h] \quad (h = 0, \ldots, m).$$

- The maximum lag m is fixed before T grows.
- Under strict stationarity, use $g(x) = x$ and $g_h(x_0, \ldots, x_h) = x_0 x_h$.
- The divisors $T - h$ and T are equivalent here because $(T - h)/T \to 1$.

Which familiar processes satisfy these conditions?

A fact we will use without proof

A well-defined process $X_t = F(Z_t, Z_{t-1}, \ldots)$, using the same function F at every time and an i.i.d. sequence (Z_t), is g-ergodic for every integrable fixed-window function.

- Finite moving averages give an immediate example.
- Selected causal AR, MA, ARMA and GARCH solutions also qualify when the required moments exist.

We must still check that the particular function we average is integrable.

Check the conditions for an MA(1)

Let

$$X_t = Z_t + \theta Z_{t-1}, \qquad Z_t \overset{\text{iid}}{\sim} N(0,1).$$

1. Which functions must be averaged to estimate $\gamma(0)$ and $\gamma(1)$?
2. What makes their expectations finite?
3. What are the two population targets?

Check the conditions for an MA(1)

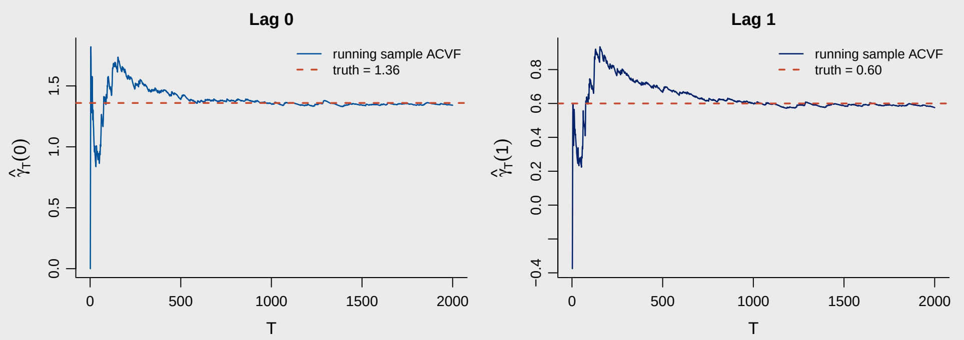
Let

$$X_t = Z_t + \theta Z_{t-1}, \qquad Z_t \overset{\text{iid}}{\sim} N(0,1).$$

1. Which functions must be averaged to estimate $\gamma(0)$ and $\gamma(1)$?
2. What makes their expectations finite?
3. What are the two population targets?

Use $g(x) = x$, $g_0(x) = x^2$, and $g_1(x_0, x_1) = x_0 x_1$. Finite second moments give integrability.

The two covariance estimates approach their targets



Each estimate uses the first T observations, its own sample mean, and divisor T.

The theorem concerns fixed lags. Here those lags remain 0 and 1.

Simulated Gaussian MA(1), coefficient $\theta = 0.6$; seed 542. Thin: estimates. Thick dashed: population ACVF.

These conditions give consistency at each fixed lag

Fixed-lag sample-ACVF consistency

If (X_t) is ACVF-learnable through lag m, then

$$\widehat{\gamma}_T(h) \xrightarrow{P} \gamma(h), \qquad h = 0, 1, \ldots, m.$$

The raw product averages are already covered by the assumptions. We still need to account for centering by $\bar{X}_T$.

Expand the centered products

For a fixed lag h, expand the definition:

$$\widehat{\gamma}_T(h) = \frac{1}{T}\sum_{t=1}^{T-h} X_t X_{t+h} - \bar{X}_T \frac{1}{T}\sum_{t=1}^{T-h} X_t - \bar{X}_T \frac{1}{T}\sum_{t=1}^{T-h} X_{t+h} + \left(1 - \frac{h}{T}\right)\bar{X}_T^2.$$

Expand the centered products

For a fixed lag h, expand the definition:

$$\widehat{\gamma}_T(h) = \frac{1}{T}\sum_{t=1}^{T-h} X_t X_{t+h} - \bar{X}_T \frac{1}{T}\sum_{t=1}^{T-h} X_t - \bar{X}_T \frac{1}{T}\sum_{t=1}^{T-h} X_{t+h} + \left(1 - \frac{h}{T}\right)\bar{X}_T^2.$$

Each partial mean omits only h terms from the full mean. Why does that omission vanish?

A fixed number of endpoint terms has vanishing weight

Weak stationarity and Cauchy–Schwarz give

$$\mathbb{E}\left[\left|\frac{1}{T}\sum_{t=T-h+1}^{T} X_t\right|\right] \leq \frac{h}{T}\sqrt{\mu^2+\gamma(0)} \longrightarrow 0.$$

The same bound holds at the left endpoint. Markov's inequality makes both omissions vanish in probability.

A fixed number of endpoint terms has vanishing weight

Weak stationarity and Cauchy–Schwarz give

$$\mathbb{E}\left[\left|\frac{1}{T}\sum_{t=T-h+1}^{T} X_t\right|\right] \le \frac{h}{T}\sqrt{\mu^2 + \gamma(0)} \longrightarrow 0.$$

The same bound holds at the left endpoint. Markov's inequality makes both omissions vanish in probability.

The four terms therefore converge to $\mathbb{E}[X_0 X_h] - \mu^2 - \mu^2 + \mu^2 = \gamma(h)$.

Fixed-lag consistency leaves a growing-sum problem

The exact variance formula still uses

$$T\,\mathrm{Var}(\bar{X}_T) = \gamma(0) + 2\sum_{h=1}^{T-1}\left(1 - \frac{h}{T}\right)\gamma(h).$$

We have consistency at each fixed lag. Does that control this growing sum?

Fixed-lag consistency leaves a growing-sum problem

The exact variance formula still uses

$$T\,\mathrm{Var}(\bar{X}_T) = \gamma(0) + 2\sum_{h=1}^{T-1}\left(1 - \frac{h}{T}\right)\gamma(h).$$

We have consistency at each fixed lag. Does that control this growing sum?

No bound or rate has been supplied for the combined estimation error over the increasing collection of lags.

What remains before we can calibrate an estimated mean?

- When does $T\,\mathrm{Var}(\bar{X}_T)$ settle to a finite limit?
- Can that limiting scale be estimated from one path?
- Does the standardized mean error have an approximately Gaussian shape?

Workbook 6 identifies and estimates the scale for conservative intervals. Workbook 7 studies the CLT and Gaussian calibration.