STATS 207: Time Series Analysis Autumn 2020

Lecture 2: Sample ACF and Basic Theoretical Constructs

Dr. Alon Kipnis
Slides credit: David Donoho, Dominik Rothenhäusler

September 16th 2020

Outline

White Noise

Autocovariance and Autocorrelation

Stationarity

Estimating Auto- and Crosscovariance

Genera Info

- Home assignment 1 will be posted on Monday 9/21/2020. It is Due two weeks later (Monday 10/5/2020).
- We will announce on Canvas and provide the link as-soon-as the assignment is posted.
- Lecture 1 slides and recording are available on Canvas.
- Syllabus with estimated week numbers for each topic is on Canvas.

Models for Time Series Data

Name	Example
White noise	$w_t \sim \mathcal{N}(0, \sigma^2)$
Moving Average	$x_t = (w_{t-1} + w_t + w_{t+1})/3$
Autoregression	$x_t = x_{t-1} - 0.9x_{t-2} + w_t$
Random Walk	$x_t = x_{t-1} + w_t$
Sinusoid in noise	$x_t = 2\cos(2\pi t/50 + 0.6\pi) + w_t$

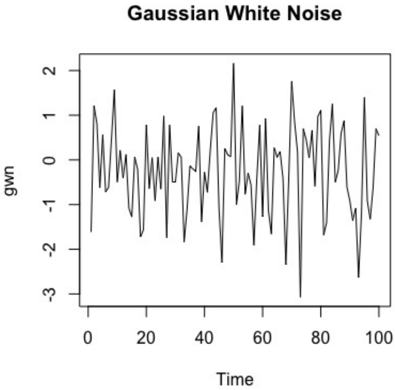
Recall lowercase notation: For each $t = 0, 1, \ldots$, x_t, is a random variables.

White Noise

3 Flavors of White Noise

A **Change** to the definitions in lecture 1:

- Definition: (w_t) is *white noise* if
  1. $\mathbb{E}[w_t] = 0$ for all $t = 1, 2, \ldots$ (zero mean)
  2. $\mathbb{E}\left[w_t^2\right] = \sigma_w^2$ for all $t = 1, 2, \ldots$ (finite and identical variance)
  3. $\mathbb{E}[w_s w_t] = 0$, for all $t \neq s$ (pairwise uncorrelated).
- Definition: (w_t) is *iid noise* if it is white noise and if $w_1, w_2, \ldots$ are independent and identically distributed.
- Definition: (w_t) is *white Gaussian noise* if it is iid noise and $w_t \sim N(0, \sigma_w^2)$.



- White noise has **no varying structure over time**.
- For most real time series data white noise is not a good model!
- Standard working pipeline:
  1. Transform data
  2. Fit model to data
  3. Check whether residuals are white noise

How to check whether white noise is a good model for data?

Autocorrelation Function

Data: $x_1, x_2, \ldots, x_n$.

- Definition: *Sample ACF*:

$$\hat{\rho}(h) \equiv \frac{\sum_{t=1}^{n-h}(x_t - \bar{x})(x_{t+h} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}, \quad h = 1, 2, \ldots, n-1.$$

- **For white noise** (from the CLT):

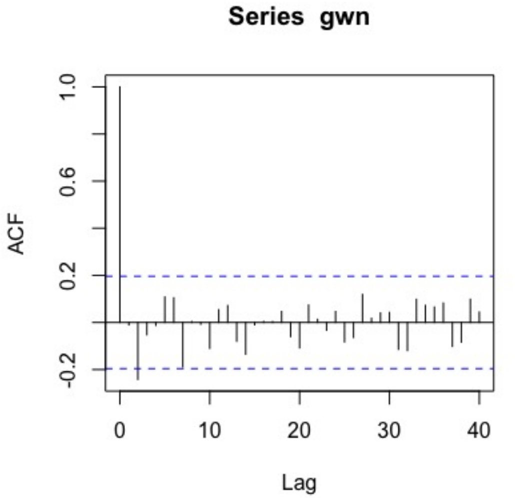
$$\hat{\rho}(h) \xrightarrow{D} \mathcal{N}(0, 1),$$

hence,

$$\Pr\left(|\hat{\rho}(h)| > 1.96/\sqrt{n}\right) \approx \Pr\left(|\mathcal{N}(0, 1)| > 1.96\right) = 0.05.$$

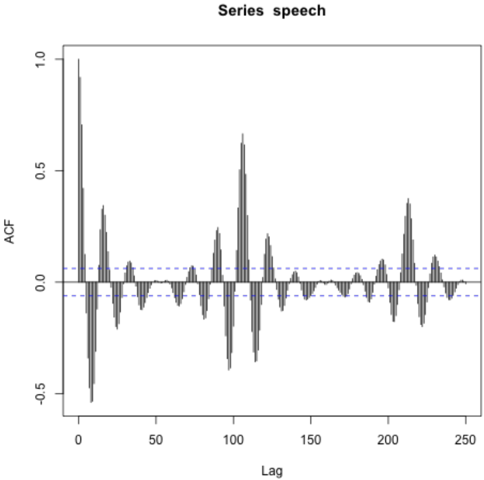
Correlogram: Plot the sample ACF for $h = 1, \ldots, k$ to check if white noise is a good model for data (R function `acf`).

Example of Sample ACF of White Noise



ACF of Speech Data

Example 1.27 in [Shumway & Stoffer]



Not a white noise!

Autocovariance and Autocorrelation

Joint random variables

- Definition: Joint CDF

$$F_{t_1,\ldots,t_n}(c_1,\ldots,c_n) = \Pr\left(x_{t_1} \leq c_1,\ldots,x_{t_n} \leq x_n\right)$$

- Example: white Gaussian noise $x_i \overset{iid}{\sim} \mathcal{N}(0,1)$,

$$F_{t_1,\ldots,t_n}(c_1,\ldots,c_n) = \prod_{i=1}^{n} \phi(c_i), \qquad \phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

- Definition: *Marginal* distribution function

$$F_t(x) = \Pr\left(x_t \leq x\right), \quad x \in \mathbb{R}$$

- Definition: Marginal *density* function

$$f_t(x) = \frac{d}{dx} F_t(x).$$

Mean Function

- Definition: The *mean* function

$$\mu_{xt} \equiv \mathbb{E}[x_t] = \int_{-\infty}^{\infty} x_t f_t(x) dx$$

- Example: (w_t) white noise,

$$\mu_{wt} = 0, \quad \forall t.$$

- Properties:
  - Mean is linear:

$$y_t = a \cdot x_t + b \cdot y_t \Rightarrow \mu_{yt} = a\mu_{xt} + b\mu_{zt}$$

  - Warning: μ_{xt} is **not** the long-run average! Mean at different t can in principle be different.

Example Mean Functions

- Moving Average: $x_t = (w_{t-1} + w_t + w_{t+1})/3$,

$$\begin{aligned} \mu_{xt} &= \mathbb{E}\left[x_t\right] \\ &= \mathbb{E}\left[(w_{t-1} + w_t + w_{t+1})/3\right] \\ &= (\mathbb{E}\left[w_{t-1}\right] + \mathbb{E}\left[w_t\right] + \mathbb{E}\left[w_{t+1}\right])/3 = 0 \end{aligned}$$

- Random walk with drift: $x_t = \delta \cdot t + \sum_{u=1}^{t} w_u$,

$$\mu_{xt} = \mathbb{E}\left[x_t\right] = \mathbb{E}\left[\delta \cdot t + \sum_{u=1}^{t} w_u\right] = \mathbb{E}\left[\delta \cdot t\right] + \mathbb{E}\left[\sum_{u=1}^{t} w_u\right] = \delta \cdot t.$$

- Signal plus noise: $x_t = 2\cos\left(2\pi t/50 + 0.6\pi\right) + w_t$,

$$\mu_{xt} = 2\cos\left(2\pi t/50 + 0.6\pi\right)$$

Autocovariance Function

- Definition:

$$\gamma_x(s,t) \equiv \mathrm{Cov}(x_s, x_y) = \mathbb{E}\left[(x_s - \mu_{xs})(x_t - \mu_{xt})\right]$$

- Example: $w_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_w^2)$,

$$\gamma_w(s,t) = \begin{cases} \sigma^2 & t = s, \\ 0 & t \neq s. \end{cases}$$

- Properties:
  - $\gamma_x(t,t) = \mathrm{Var}(x_t) = \mathbb{E}\left[(x_t - \mu_{xt})^2\right]$
  - Autocovariance is bilinear:

    $$y_t = ax_t + bz_t \Rightarrow \gamma_y(t,t) = a^2\gamma_x(t,t) + b^2\gamma_z(t,t)$$

    if every x_t and z_t are uncorrelated.
  - Cauchy-Schwartz inequality:

    $$\gamma_x(s,t) \leq \sqrt{\gamma_x(t,t)\gamma_x(s,s)}$$

  - Warning: $\gamma_x(t,t)$ is **not** the long-run variance! Variance at different t can in principle be different.

Autocorrelation Function

- Definition:

$$\rho_x(s,t) \equiv \mathrm{corr}(x_t, x_s) = \frac{\mathrm{Cov}(x_s, x_t)}{\sqrt{\mathrm{Var}(x_t)\mathrm{Var}(x_s)}} = \frac{\gamma_x(s,t)}{\sqrt{\gamma_x(t,t)\gamma_x(s,s)}}.$$

- Example: $w_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$,

$$\gamma_w(s,t) = \begin{cases} 1 & t = s, \\ 0 & t \neq s. \end{cases}$$

- Properties:
  - $\rho_x(t,t) = 1$
  - $-1 \leq \rho_x(s,t) \leq 1$
  - $|\rho_x(s,t)| = 1$ implies perfect linear relationship:

$$x_t = a \cdot x_s + b, \qquad \text{for some scalars } a,\ b$$

Example Autocovariance Functions

- Moving Average: $x_t = (w_{t-1} + w_t + w_{t+1})/3$,

$$\gamma_x(s,t) = \sigma_w^2 \begin{cases} 3/9 & t = s, \\ 2/9 & |t-s| = 1, \\ 1/9 & |t-s| = 2, \\ 0 & |t-s| > 2. \end{cases}$$

  Note: averaging/filtering introduces correlation.

- Random Walk with drift: $x_t = \delta \cdot t + \sum_{u=1}^{t} w_u$,

$$\gamma_x(s,t) = \min(s,t)\sigma_w^2.$$

- Signal plus noise: $x_t = 2\cos(2\pi t/50 + 0.6\pi) + w_t$,

$$\gamma_x(s,t) = \begin{cases} \sigma_w^2 & t = s, \\ 0 & t \neq s. \end{cases}$$

Stationarity

Stationarity

- *Strict stationarity*:
  - Definition: for all $t_1, \ldots, t_n$ and h,

$$\{x_{t_1}, \ldots, x_{t_n}\} =_D \{x_{t_1+h}, \ldots, x_{t_n+h}\},$$

  namely

$$F_{t_1,\ldots,t_n}(c_1, \ldots, c_n) = F_{t_{1+h},\ldots,t_{n+h}}(c_1, \ldots, c_n).$$

- *stationarity* (aka *weak* stationarity, mean-variance stationarity):
  - Definition: for all t, s,

$$\mu_{xt} = \mu_{xs} \qquad \gamma_x(s,t) = \gamma_x(0, |t-s|).$$

- **Strict stationarity implies weak stationarity**.
- Notation (abuse of):

$$h = t - s, \qquad \gamma(h) \equiv \gamma(0, |t-s|)$$

- **YES**: Moving average $x_t = (w_{t-1} + w_t + w_{t+1})/3$,

$$\mu_{xt} = 0, \qquad \gamma_x(s,t) = \sigma_w^2 \begin{cases} 3/9 & t = s, \\ 2/9 & |t-s| = 1, \\ 1/9 & |t-s| = 2, \\ 0 & |t-s| > 2. \end{cases}$$

- **NO**: Random walk with drift $x_t = \delta \cdot t + \sum_{u=1}^{t} w_u$,

$$\mu_{xt} = \delta \cdot t, \qquad \gamma_x(s,t) = \min(s,t)\sigma_w^2.$$

- **NO**: Signal plus noise $x_t = 2\cos\left(2\pi t/50 + 0.6\pi\right) + w_t$,

$$\mu_{xt} = 2\cos\left(2\pi t/50 + 0.6\pi\right), \qquad \gamma_x(h) = \begin{cases} \sigma_w^2 & h = 0, \\ 0 & h \neq 0. \end{cases}$$

Joint Stationarity & Cross-Covariance

- Definition: *Cross-covariance function* :

$$\gamma_{xy}(s,t) \equiv \mathrm{Cov}(x_s, y_t) = \mathbb{E}\left[(x_s - \mu_{xs})(y_t - \mu_{yt})\right].$$

  Note: $\gamma_{xx}(s,t) = \gamma_x(s,t)$.

- Definition: x and y are *jointly stationary* if each is stationary, and

$$\gamma_{xy}(s,t) = \gamma_{xy}(0, s-t)$$

  Notation: $\gamma_{xy}(h) = \gamma_{xy}(t+h, t)$.

- Example: $x_t = w_t + w_{t-1}$, $y_t = w_t - w_{t-1}$, (w_t) white noise.

$$\gamma_{xy}(s,t) = \sigma_w^2 \begin{cases} 0 & t = s, \\ 1 & s - t = 1, \\ -1 & s - t = -1, \\ 0 & |t-s| \geq 2. \end{cases}$$

Example: Cross-Covariance of Delayed Series

- (x_t) is stationary and zero mean. (y_t) is delayed, noisy version of (x_t):

$$y_t = Ax_{t-\ell} + w_t,$$

where (w_t) is white noise independent of (x_t). Then

$$\gamma_{yx}(h) = A \cdot \gamma_x(h - \ell) + \gamma_{wx}(h) = A \cdot \gamma_x(h - \ell)$$

- $\gamma_{yx}(h)$ peaks at $h = \ell$, takes value $A \cdot \gamma_x(0)$.
- The best linear prediction of $y_{t+\ell}$ given x_t is

$$\hat{y}_{t+\ell} = \frac{\gamma_{yx}(\ell)}{\gamma_x(0)} x_t = Ax_t.$$

Gaussian Process

- Definition: (x_t) is a *Gaussian process* if for any $\{t_1, \ldots, t_n\} \subset \mathbb{N}$, $(x_{t_1}, \ldots, x_{t_n})$ has a multivariate normal distribution.

Corollary

Every stationary Gaussian process is strictly stationary.

- Example: Gaussian random walk $x_t = x_{t-1} + w_t$
  - Is it Gaussian? **Yes**
  - Is it stationary? **No**
  - How about $y_t = x_{t+h} - x_t$, $h \in \mathbb{N}$? **Yes**

Linear and Stationary Gaussian Processes

- Definition: (stationary) *linear process* is the output of a linear time-invariant system

$$x_t = \mu + \sum_{h=-\infty}^{\infty} \psi_h w_{t-h}, \quad \sum_{h=-\infty}^{\infty} |\psi_h| < \infty.$$

Theorem (Cramèr)

Every stationary Gaussian process that is indeterministic is a linear process with Gaussian (w_t).

Examples of Linear Processes

- Example: Stationary linear processes:
  - White noise
  - Moving average
  - Non-explosive autoregressive
- Example: Non-Gaussian linear process: (U_t) is a sequence of Bernoulli trials (coin tosses $H \to -1, T \to 1$).

$$x_t = \mu + \sum_{h=-\infty}^{\infty} \psi_h U_{t-h}$$

Contrasting Example

(w_t) is white Gaussian noise.

$$y_t = \begin{cases} w_t & t \text{ is even} \\ \frac{w_{t-1}^2 - \sigma_w^2}{\sqrt{2}} & t \text{ is odd.} \end{cases}$$

Process (y_t) is:

- Pairwise uncorrelated **Yes**.
- Pairwise independent **No**.
- Strictly stationary **No**.
- Weakly stationary **Yes**.
- Gaussian **No**.

Estimating Auto- and Crosscovariance

Sample Auto/Cross Covariance/Correlation

- Definition: *sample mean*

$$\bar{x} \equiv \frac{1}{n} \sum_{t=1}^{n} x_t$$

- Definition: *sample autocovariance* function

$$\hat{\gamma}_x(h) \equiv \frac{1}{n} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x}).$$

- Definition: *sample cross-covariance* function

$$\hat{\gamma}_{xy}(h) \equiv \frac{1}{n} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(y_t - \bar{y}).$$

- Definition: *sample cross-correlation* function

$$\hat{\rho}_{xy}(h) \equiv \frac{\hat{\gamma}_{xy}(h)}{\sqrt{\hat{\gamma}_x(0)\hat{\gamma}_y(0)}}.$$

- Warning: mind the difference between *ensemble* averaging (many realizations) and *time* averaging (one realization).

Confidence Limits for Autocovariance

For white noise (w_t):

- Sample autocorrelation:

$$\mathrm{SE}\left[\hat{\rho}_w(h)\right] \approx \frac{1}{\sqrt{n}}.$$

- Sample cross-correlation:

$$\mathrm{SE}\left[\hat{\rho}_{wx}(h)\right] \approx \frac{1}{\sqrt{n}},$$

  where (x_t) is independent of (w_t).

(approximations assume $h \ll n$)

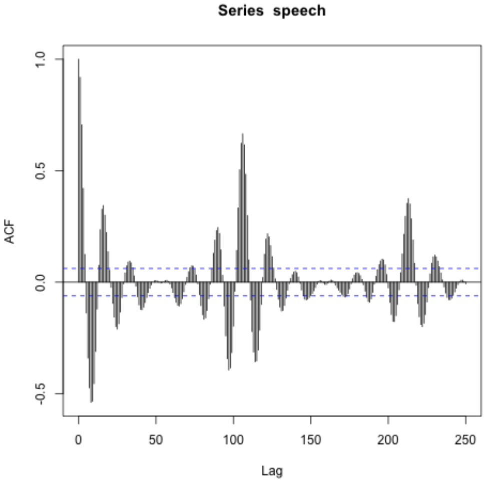
- From the CLT (Theorem A.7 and Property P1.1 in [Shumway & Stoffer]): for $h > 0$,

$$\Pr\left(|\hat{\rho}_w(h)| > 1.96/\sqrt{n}\right) \approx \Pr\left(|\mathcal{N}(0,1)| > 1.96\right) = 0.05.$$

$$\Pr\left(|\hat{\rho}_{wx}(h)| > 1.96/\sqrt{n}\right) \approx \Pr\left(|\mathcal{N}(0,1)| > 1.96\right) = 0.05.$$

ACF of Speech Data

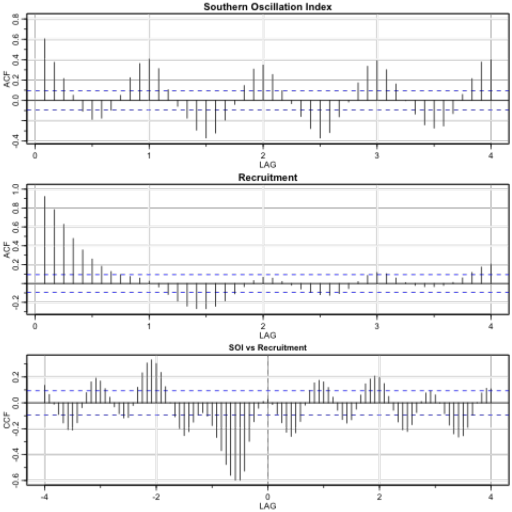
Example 1.27 in [Shumway & Stoffer]



Not a white noise!

ACF/CCF of SOI/Recruitment Data

Example 1.28 in [Shumway & Stoffer]



Multiple Testing Warning

- If $|\hat{\rho}_x(h)| > 1.96n^{-1/2}$ for *some* $h > 0$, can we determine that (x_t) is not a white noise?