Introduction to Time Series Analysis. Lecture 4.

Peter Bartlett

Last lecture:

1. Sample autocorrelation function
2. ACF and prediction
3. Properties of the ACF

Introduction to Time Series Analysis. Lecture 4.

Peter Bartlett

1. Review: ACF, sample ACF.
2. Properties of estimates of μ and ρ.
3. Convergence in mean square.

Mean, Autocovariance, Stationarity

A time series $\{X_t\}$ has **mean function** $\mu_t = \mathrm{E}[X_t]$
and **autocovariance function**

$$\begin{aligned}\gamma_X(t+h,t) &= \mathrm{Cov}(X_{t+h},X_t)\\ &= \mathrm{E}[(X_{t+h}-\mu_{t+h})(X_t-\mu_t)].\end{aligned}$$

It is **stationary** if both are independent of t.
Then we write $\gamma_X(h) = \gamma_X(h,0)$.
The **autocorrelation function (ACF)** is

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \mathrm{Corr}(X_{t+h},X_t).$$

Estimating the ACF: Sample ACF

For observations $x_1, \ldots, x_n$ of a time series,

the **sample mean** is $$\bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t.$$

The **sample autocovariance function** is

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}), \qquad \text{for } -n < h < n.$$

The **sample autocorrelation function** is

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}.$$

Properties of the autocovariance function

For the autocovariance function γ of a stationary time series $\{X_t\}$,

1. $\gamma(0) \geq 0$,
2. $|\gamma(h)| \leq \gamma(0)$,
3. $\gamma(h) = \gamma(-h)$,
4. γ is positive semidefinite.

Furthermore, any function $\gamma : \mathbb{Z} \to \mathbb{R}$ that satisfies (3) and (4) is the autocovariance of some stationary (Gaussian) time series.

Introduction to Time Series Analysis. Lecture 4.

1. Review: ACF, sample ACF.
2. Properties of estimates of μ and ρ.
3. Convergence in mean square.

Properties of the sample autocovariance function

The **sample autocovariance function**:

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}), \qquad \text{for } -n < h < n.$$

For any sequence $x_1, \ldots, x_n$, the sample autocovariance function $\hat{\gamma}$ satisfies

1. $\hat{\gamma}(h) = \hat{\gamma}(-h)$,
2. $\hat{\gamma}$ is positive semidefinite, and hence
3. $\hat{\gamma}(0) \geq 0$ and $|\hat{\gamma}(h)| \leq \hat{\gamma}(0)$.

Properties of the sample autocovariance function: psd

$$\hat{\Gamma}_n = \begin{pmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(n-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}(n-1) & \hat{\gamma}(n-2) & \cdots & \hat{\gamma}(0) \end{pmatrix}$$

$$= \frac{1}{n} M M', \qquad \text{(see next slide)}$$

$$\text{so } a'\hat{\Gamma}_n a = \frac{1}{n}(a'M)(M'a) = \frac{1}{n}\|M'a\|^2 \geq 0,$$

i.e., Γ_n is a covariance matrix. It is also important for forecasting.

Properties of the sample autocovariance function: psd

$$M = \begin{pmatrix} 0 & \cdots & 0 & 0 & \tilde{X}_1 & \tilde{X}_2 & \cdots & \tilde{X}_n \\ 0 & \cdots & 0 & \tilde{X}_1 & \tilde{X}_2 & \cdots & \tilde{X}_n & 0 \\ 0 & \cdots & \tilde{X}_1 & \tilde{X}_2 & \cdots & \tilde{X}_n & 0 & 0 \\ \vdots & & & & & & & \vdots \\ \tilde{X}_1 & \tilde{X}_2 & & \cdots & \tilde{X}_n & 0 & \cdots & 0 \end{pmatrix}.$$

and $\tilde{X}_t = X_t - \mu$.

Estimating μ

How good is $\bar{X}_n$ as an estimate of μ?

For a stationary process $\{X_t\}$, the sample average,

$$\bar{X}_n = \frac{1}{n}\left(X_1 + \cdots + X_n\right) \text{ satisfies}$$

$$\mathrm{E}(\bar{X}_n) = \mu, \qquad \text{(unbiased)}$$

$$\operatorname{var}(\bar{X}_n) = \frac{1}{n}\sum_{h=-n}^{n}\left(1 - \frac{|h|}{n}\right)\gamma(h).$$

Estimating the ACF: Sample ACF

To see why:

$$\begin{aligned}
\mathrm{var}(\bar{X}_n) &= \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)\left(\frac{1}{n}\sum_{j=1}^{n} X_j - \mu\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{E}(X_i - \mu)(X_j - \mu) \\
&= \frac{1}{n^2}\sum_{i,j} \gamma(i-j) \\
&= \frac{1}{n}\sum_{h=-(n-1)}^{n-1}\left(1 - \frac{|h|}{n}\right)\gamma(h).
\end{aligned}$$

Estimating μ

Since $$\mathrm{var}(\bar{X}_n) = \frac{1}{n} \sum_{h=-n}^{n} \left(1 - \frac{|h|}{n}\right) \gamma(h),$$

if $$\lim_{h \to \infty} \gamma(h) = 0, \ \mathrm{var}(\bar{X}_n) \to 0.$$

Estimating μ

Also, since $$\mathrm{var}(\bar{X}_n) = \frac{1}{n}\sum_{h=-n}^{n}\left(1 - \frac{|h|}{n}\right)\gamma(h),$$

if $$\sum_h |\gamma(h)| < \infty,\ n\,\mathrm{var}(\bar{X}_n) \to \sum_{h=-\infty}^{\infty}\gamma(h) = \sigma^2 \sum_{h=-\infty}^{\infty}\rho(h).$$

Compare this to the uncorrelated case....

Estimating μ

$$n\,\mathrm{var}(\bar{X}_n) \to \sigma^2 \sum_{h=-\infty}^{\infty} \rho(h).$$

i.e., instead of $\mathrm{var}(\bar{X}_n) \approx \dfrac{\sigma^2}{n}$, we have $\mathrm{var}(\bar{X}_n) \approx \dfrac{\sigma^2}{n/\tau}$,

with $\tau = \sum_h \rho(h)$. The effect of the correlation is a reduction of sample size from n to n/τ. (c.f. mixing time.)

Estimating μ: Asymptotic distribution

Why are we interested in asymptotic distributions?

- If we know the asymptotic distribution of $\bar{X}_n$, we can use it to construct hypothesis tests,
  e.g., is $\mu = 0$?
- Similarly for the asymptotic distribution of $\hat{\rho}(h)$,
  e.g., is $\rho(1) = 0$?

Notation: $X_n \sim AN(\mu_n, \sigma_n^2)$ means 'asymptotically normal':

$$\frac{X_n - \mu_n}{\sigma_n} \xrightarrow{d} Z, \text{ where } Z \sim N(0,1).$$

Estimating μ for a linear process: Asymptotically normal

Theorem (A.5) For a linear process $X_t = \mu + \sum_j \psi_j W_{t-j}$, if $\sum \psi_j \neq 0$, then

$$\bar{X}_n \sim AN\left(\mu_x, \frac{V}{n}\right),$$

$$\text{where} \qquad V = \sum_{h=-\infty}^{\infty} \gamma(h)$$

$$= \sigma_w^2 \left(\sum_{j=-\infty}^{\infty} \psi_j \right)^2.$$

($X \sim AN(\mu_n, \sigma_n)$ means $\sigma_n^{-1}(X_n - \mu_n) \xrightarrow{d} Z$.)

Estimating μ for a linear process

Recall: for a linear process $X_t = \mu + \sum_j \psi_j W_{t-j}$,

$$\gamma_X(h) = \sigma_w^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{h+j},$$

so

$$\lim_{n\to\infty} n \operatorname{var}(\bar{X}_n) = \lim_{n\to\infty} \sum_{h=-(n-1)}^{n-1} \left(1 - \frac{|h|}{n}\right) \gamma(h)$$

$$= \lim_{n\to\infty} \sigma_w^2 \sum_{j=-\infty}^{\infty} \psi_j \sum_{h=-(n-1)}^{n-1} \left(\psi_{j+h} - \frac{|h|}{n} \psi_{j+h}\right)$$

$$= \sigma_w^2 \left(\sum_{j=-\infty}^{\infty} \psi_j \right)^2 .$$

Estimating the ACF: Sample ACF for White Noise

Theorem For a white noise process W_t,

if $\mathrm{E}(W_t^4) < \infty$,

$$\begin{pmatrix} \hat{\rho}(1) \\ \vdots \\ \hat{\rho}(K) \end{pmatrix} \sim AN\left(0, \frac{1}{n} I\right).$$

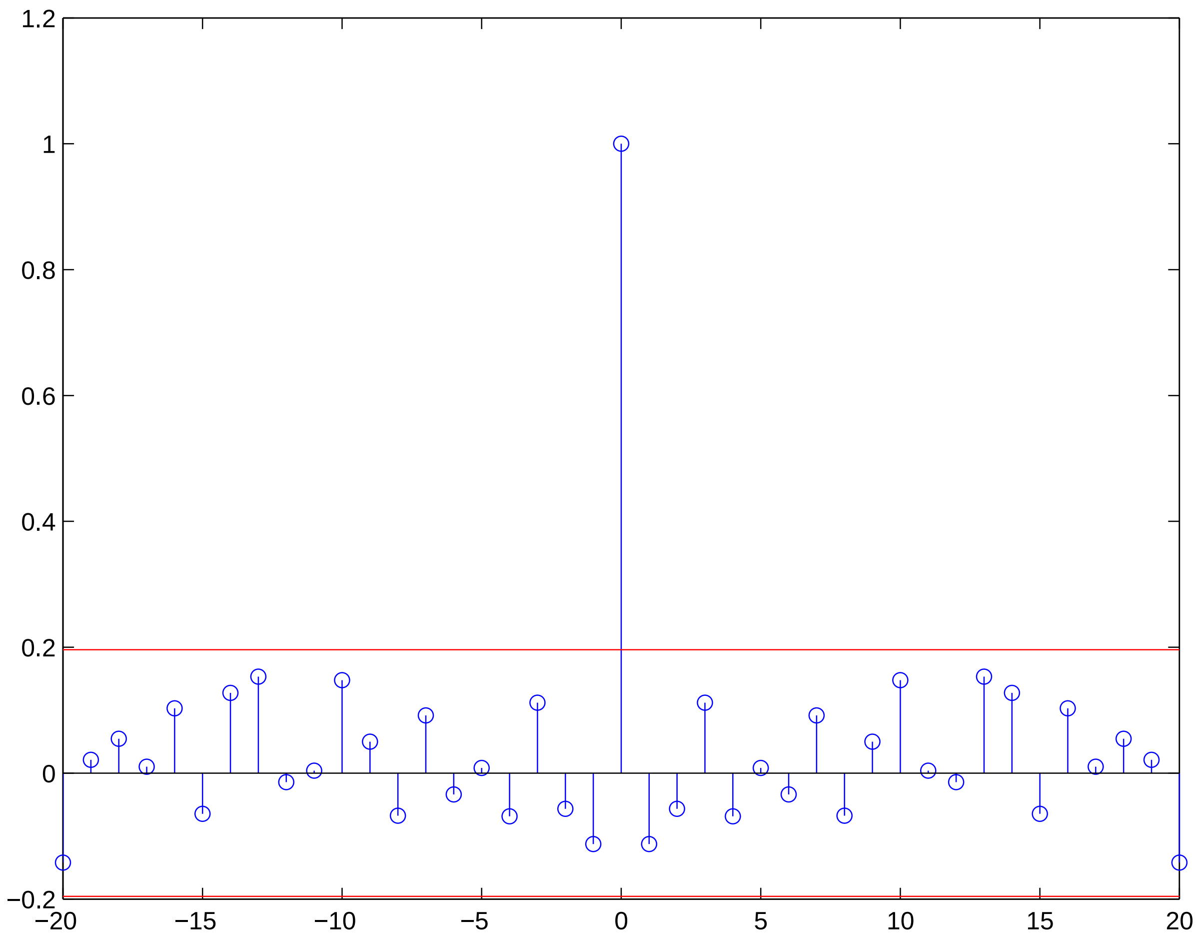
Sample ACF and testing for white noise

If $\{X_t\}$ is white noise, we expect no more than $\approx 5\%$ of the peaks of the sample ACF to satisfy

$$|\hat{\rho}(h)| > \frac{1.96}{\sqrt{n}}.$$

This is useful because we often want to introduce transformations that reduce a time series to white noise.

Sample ACF for white Gaussian (hence i.i.d.) noise



Estimating the ACF: Sample ACF

Theorem (A.7) For a linear process $X_t = \mu + \sum_j \psi_j W_{t-j}$, if $\mathrm{E}(W_t^4) < \infty$,

$$\begin{pmatrix} \hat{\rho}(1) \\ \vdots \\ \hat{\rho}(K) \end{pmatrix} \sim AN\left(\begin{pmatrix} \rho(1) \\ \vdots \\ \rho(K) \end{pmatrix}, \frac{1}{n}V \right),$$

$$\text{where } V_{i,j} = \sum_{h=1}^{\infty} \left(\rho(h+i) + \rho(h-i) - 2\rho(i)\rho(h)\right) \times \left(\rho(h+j) + \rho(h-j) - 2\rho(j)\rho(h)\right).$$

Notice: If $\rho(i) = 0$ for all $i \neq 0$, $V = I$.

Sample ACF for MA(1)

Recall: $\rho(0) = 1$, $\rho(\pm 1) = \frac{\theta}{1+\theta^2}$, and $\rho(h) = 0$ for $|h| > 1$. Thus,

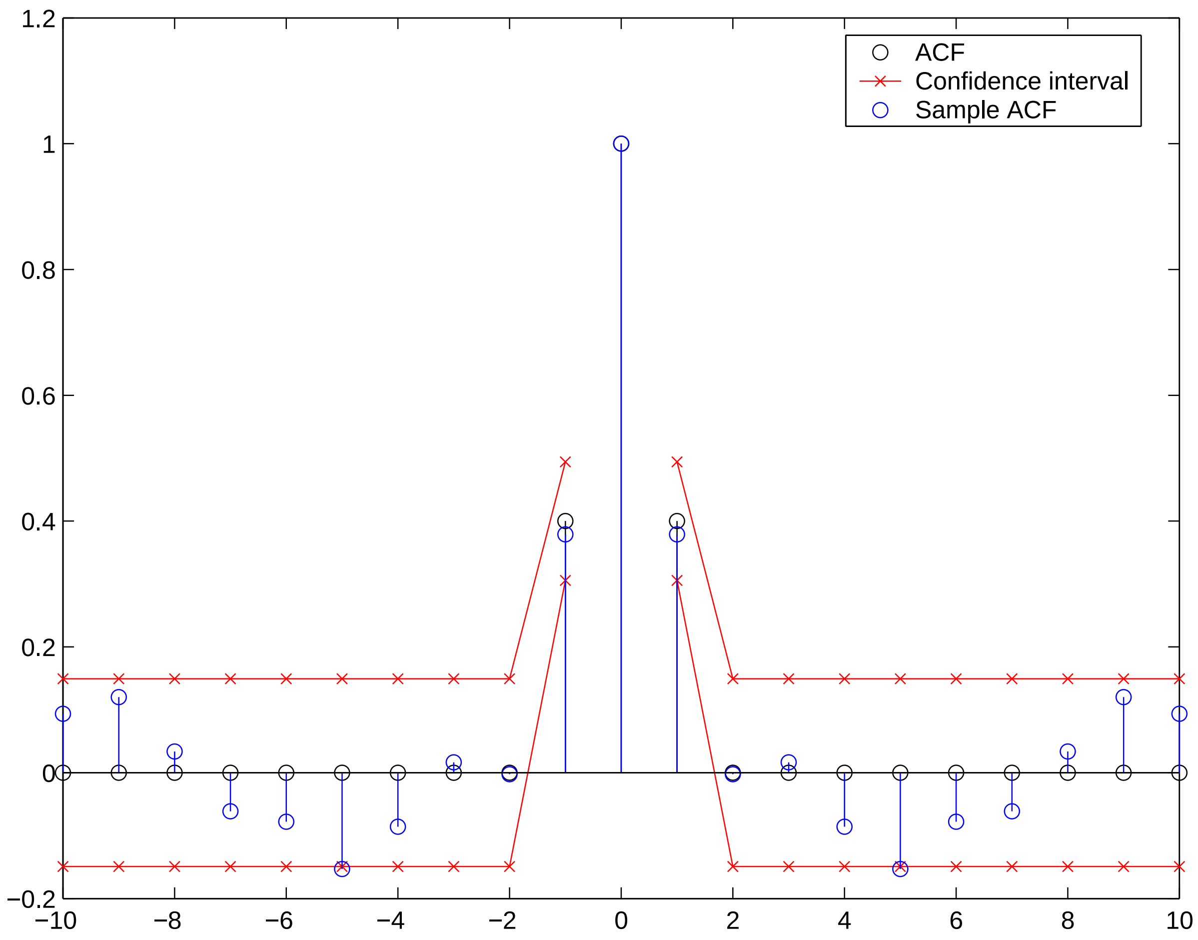
$$V_{1,1} = \sum_{h=1}^{\infty} \left(\rho(h+1) + \rho(h-1) - 2\rho(1)\rho(h)\right)^2 = (\rho(0) - 2\rho(1)^2)^2 + \rho(1)^2$$

$$V_{2,2} = \sum_{h=1}^{\infty} \left(\rho(h+2) + \rho(h-2) - 2\rho(2)\rho(h)\right)^2 = \sum_{h=-1}^{1} \rho(h)^2.$$

And if $\hat{\rho}$ is the sample ACF from a realization of this MA(1) process, then with probability 0.95,

$$|\hat{\rho}(h) - \rho(h)| \leq 1.96\sqrt{\frac{V_{hh}}{n}}.$$

Sample ACF for MA(1)



Introduction to Time Series Analysis. Lecture 4.

1. Review: ACF, sample ACF.
2. Properties of estimates of μ and ρ.
3. Convergence in mean square.

Convergence in Mean Square

- Recall the definition of a linear process:

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}$$

- What do we mean by these infinite sums of random variables? i.e., what is the 'limit' of a sequence of random variables?

- Many types of convergence:

  1. Convergence in distribution.
  2. Convergence in probability.
  3. Convergence in mean square.

Convergence in Mean Square

Definition: A sequence of random variables $S_1, S_2, \ldots$ **converges in mean square** if there is a random variable Y for which

$$\lim_{n \to \infty} \mathrm{E}(S_n - Y)^2 = 0$$

Example: Linear Processes

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}$$

Then if $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$,

$$(1) \qquad |X_t| < \infty \text{ a.s.}$$

$$(2) \qquad \sum_{j=-\infty}^{\infty} \psi_j W_{t-j} \text{ converges in mean square}$$

Example: Linear Processes (Details)

$$
\begin{aligned}
(1) \qquad P\left(|X_t| \geq \alpha\right) &\leq \frac{1}{\alpha}\mathrm{E}|X_t| && \text{(Markov's inequality)} \\
&\leq \frac{1}{\alpha}\sum_{j=-\infty}^{\infty} |\psi_j|\mathrm{E}|W_{t-j}| \\
&\leq \frac{\sigma}{\alpha}\sum_{j=-\infty}^{\infty} |\psi_j| && \text{(Jensen's inequality)} \\
&\to 0.
\end{aligned}
$$

Example: Linear Processes (Details)

For (2):

The **Riesz-Fisher Theorem** (Cauchy criterion):

S_n converges in mean square iff

$$\lim_{m,n\to\infty} \mathrm{E}(S_m - S_n)^2 = 0.$$

Example: Linear Processes (Details)

(2) $$S_n = \sum_{j=-n}^{n} \psi_j W_{t-j}$$ converges in mean square, since

$$\begin{aligned}
\mathrm{E}(S_m - S_n)^2 &= \mathrm{E}\left(\sum_{m \le |j| \le n} \psi_j W_{t-j} \right)^2 \\
&= \sum_{m \le |j| \le n} \psi_j^2 \sigma^2 \\
&\le \sigma^2 \left(\sum_{m \le |j| \le n} |\psi_j| \right)^2 \\
&\to 0.
\end{aligned}$$

Example: AR(1)

Let X_t be the stationary solution to $X_t - \phi X_{t-1} = W_t$, where $W_t \sim WN(0, \sigma^2)$.

If $|\phi| < 1$,

$$X_t = \sum_{j=0}^{\infty} \phi^j W_{t-j}$$

is a solution. The same argument as before shows that this infinite sum converges in mean square, since $|\phi| < 1$ implies $\sum_{j \geq 0} |\phi^j| < \infty$.

Example: AR(1)

Furthermore, X_t is the unique stationary solution: we can check that any other stationary solution Y_t is the mean square limit:

$$\begin{aligned}\lim_{n\to\infty} \mathrm{E}\left(Y_t - \sum_{i=0}^{n-1} \phi^i W_{t-i}\right)^2 &= \lim_{n\to\infty} \mathrm{E}(\phi^n Y_{t-n})^2 \\ &= 0.\end{aligned}$$

Example: AR(1)

Let X_t be the stationary solution to

$$X_t - \phi X_{t-1} = W_t,$$

where $W_t \sim WN(0, \sigma^2)$.

If $|\phi| < 1$,

$$X_t = \sum_{j=0}^{\infty} \phi^j W_{t-j}.$$

$\phi = 1$?

$\phi = -1$?

$|\phi| > 1$?

Introduction to Time Series Analysis. Lecture 4.

1. Properties of estimates of μ and ρ.
2. Convergence in mean square.