Properties of Sample mean and ACF

- So far we have ~~more~~ more or less assumed that the mean of the process $\{x_t\}$ we are looking at is 0.
- That may not be the case though. For a stationary process $\{x_t\}$ $E[x_t] = \mu \neq 0$ is possible. The μ won't depend on t, because of stationarity.
- This does not invalidate the developments so far. If $\mu \neq 0$ we can consider the process $\{x_t - \mu\}$ which will have mean zero, and ~~so~~ weak stationarity won't be affected.
- Given a time series $\{x_t\}$, one needs to estimate μ and the autocorrelation functions.

Estimation of μ.

- Suppose we have observation from a time series $x_1, x_2, \cdots, x_T$.
- The estimator of μ is given by the sample mean $\bar{x}_T$.

  ie
$$\hat{\mu} = \bar{x}_T = \frac{1}{T}(x_1 + x_2 + \cdots + x_T).$$

- By stationarity.
$$E[\hat{\mu}] = E[\bar{x}_T] = \mu.$$

  so $\hat{\mu}$ is unbiased for μ.

- $$E[(\bar{x}_T - \mu)^2] = \text{Var}(\bar{x}_T) = \frac{1}{T^2}\sum_{i=1}^{T}\sum_{j=1}^{T} \text{cov}(x_i, x_j)$$
$$= \frac{1}{T^2}\sum_{i-j=-T}^{T} (T - |i-j|)\,\gamma_x(i-j).$$
$$= \frac{1}{T^2}\sum_{h=-T}^{T} (T - |h|)\,\gamma_x(h)$$
$$= \frac{1}{T}\sum_{h=-T}^{T} \left(1 - \frac{|h|}{T}\right)\gamma_x(h).$$

- The question is how does $\text{Var}(\bar{x}_T)$ behave.

Var$(\bar{X}_T)$

- First of all note that.

$$T\,\mathrm{Var}(\bar{X}_T) = \sum_{h=-T}^{T}\left(1-\frac{|h|}{T}\right)\gamma_x(h) \leq \sum_{|h|\leq T} |\gamma_x(h)| .$$

- Now if $\gamma_x(h) \longrightarrow 0$ as $h \to \infty$, it can be shown that.

$$\lim_{T\to\infty} \mathrm{Var}(\bar{X}_T) \leq \lim_{T\to\infty} \frac{1}{T}\sum_{|h|\leq T} |\gamma_x(h)| = 0 .$$

- On the other hand if $\sum_{h=-\infty}^{\infty} |\gamma_x(h)| < \infty$,

$$\lim_{T\to\infty} T\,\mathrm{Var}(\bar{X}_T) = \sum_{h=-\infty}^{\infty} \gamma_x(h) .$$

- So we have the following result.

If $\{X_t\}$ is a stationary time series with mean μ and autocovariance function $\gamma_x(\cdot)$ the as $T\to\infty$.

$$\mathrm{Var}(\bar{X}_T) \longrightarrow 0 \quad \text{if} \quad \gamma_x(T) \longrightarrow 0 .$$

$$T\,\mathrm{Var}(\bar{X}_T) \longrightarrow \sum_{h=-\infty}^{\infty} \gamma_x(h) \quad \text{if} \quad \sum_{h=-\infty}^{\infty} |\gamma_x(h)| < \infty .$$

Confidence intervals for μ.

- We need to know the distribution or the approximate distribution of $\overline{X}_T$.

- If the time series is Gaussian.
$$\sqrt{T}\,(\overline{X}_T - \mu) \sim N\Big(0, \sum_{|h|\leq T}\Big(1-\frac{|h|}{T}\Big)\gamma_x(h)\Big).$$

- This formula gives an exact confidence interval if $\gamma_x(\cdot)$ are known and an approximate interval if $\gamma_x(\cdot)$ are estimated from the sample.

- For many time series, like the ARMA models. $\overline{X}_T$ is approximately normal with mean μ and variance $\frac{1}{T}\sum_{|h|<\infty}\gamma_x(h)$, for large T.

Confidence interval for μ.

- An approximate confidence interval for μ is then.

$$\left(\bar{X}_T - 1{\cdot}96\sqrt{v}/\sqrt{T},\ \bar{X}_T + 1{\cdot}96\sqrt{v}/T\right),$$

where $v = \sum_{h=-\infty}^{\infty} \gamma_x(h)$.

- Obviously v is not known in most cases and must be estimated from the data.

- A good estimate is given by.

$$\hat{v} = \sum_{|h|<\sqrt{n}} \left(1 - \frac{|h|}{n}\right) \hat{\gamma}_x(h).$$

Example.

- Let $\{X_t\}$ be an AR(1) model,

$$(X_t - \mu) = \phi (X_{t-1} - \mu) + Z_t \quad , \text{where } |\phi| < 1 \text{ and } Z_t \sim WN(0,\sigma^2)$$

- We know that $\gamma(h) = \dfrac{\phi^{|h|}\sigma^2}{1-\phi^2}$.

$$v = \sum_{h=-\infty}^{\infty} \gamma(h) \cancel{= 2\sum_{h=0}^{\infty}\gamma(h) = \frac{2\sigma^2}{1-\phi^2}\sum_{h=0}^{\infty}\phi^h}.$$

$$= \gamma(0) + 2\sum_{h=1}^{\infty}\gamma(h) = \left(1 + 2\sum_{h=1}^{\infty}\phi^h\right)\sigma^2/1-\phi^2$$

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

- So the 95% confidence interval for μ is given by,

$$\overline{X}_T \pm 1.96\,\sigma/(\sqrt{n}(1-\phi)).$$

- σ and ϕ has to be estimated from the data.

Estimation of $\gamma_x(\cdot)$ and $\rho_x(\cdot)$.

- A natural estimate of $\gamma_x(h)$ is given by.

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

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

- Both the estimators $\hat{\gamma}_x(h)$ and $\hat{\rho}_x(h)$ are biased even if we devide by $(n-h)$ instead of n. However, under general assumptions they are nearly unbiased.
- One advantage of deviding by n instead of $n-h$ is that the autocovariance matrix

$$\hat{\Gamma}_x(h) = \begin{bmatrix} \hat{\gamma}_x(0) & \hat{\gamma}_x(1) & \cdots & \hat{\gamma}_x(h-1) \\ \hat{\gamma}_x(1) & \hat{\gamma}_x(0) & \cdots & \hat{\gamma}_x(h-2) \\ \vdots & & & \\ \hat{\gamma}_x(h-1) & \cdots & \cdots & \hat{\gamma}_x(0) \end{bmatrix}$$

is n.n.d.

Distribution of $\hat{\rho}_x(h)$.

- In general even for simple time-serieses distribution of $\hat{\rho}_x(h)$ is intractible.

- However asymptotic results are known for several models.

- In particular, for linear models and ARMA models. it can be shown that

$$\hat{\rho}_x = (\hat{\rho}_x(1), \cdots, \hat{\rho}_x(h))^T \approx N(\rho_x, W/n) \text{ distribution.}$$

- Here $\rho_x = (\rho_x(1), \cdots, \rho_x(h))^T$ and W is the Covariance matrix.

- The formula for W is due to Bartlett. One can show that

$$\omega_{ij} = \sum_{k=1}^{\infty} \{\rho_x(k+i) + \rho_x(k-i) - 2\rho_x(i)\rho_x(k)\} \times \{\rho_x(k+j) + \rho_x(k-j) - 2\rho_x(j)\rho_x(k)\}.$$

Example.

- IID noise: If $\{X_t\}$ is iid noise clearly $\rho(h) = 0 \ \forall |h| > 0$. so clearly

$$w_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{o.w.} \end{cases}$$

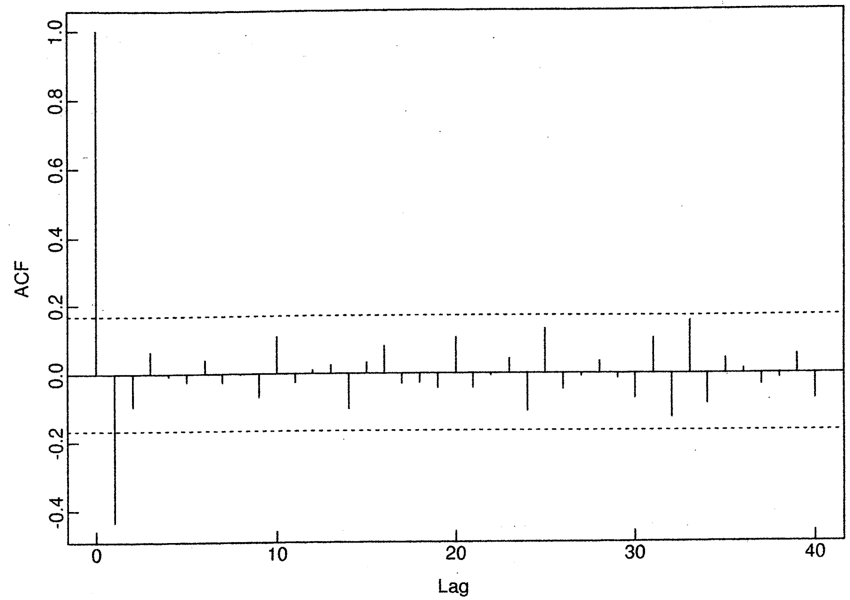
- MA (1).

If $\{X_t\}$ is a MA(1) process

$$X_t = Z_t + \theta Z_{t-1}, \ \{Z_t\} \sim WN(0, \sigma^2),$$

one can show that

$$w_{ii} = \begin{cases} 1 - 3\rho_x^2(1) + 4\rho_x^4(1) & \text{if } i = 1 \\ 1 + 2\rho^2(1) & \text{if } i > 1. \end{cases}$$



In the example we consider the case when $\theta = -.8$. $T = 200$.

The true $\rho(1) = -.8/1.64 = -.4878$,

$$\hat{\rho}(1) = -.4333 = -\frac{6.128}{\sqrt{n}}.$$

The bounds are

$$\pm 1.96 \frac{(1 + 2\hat{\rho}^2(1))^{1/2}}{\sqrt{T}}$$

Example.

- An AR(1) process: If we consider an AR(1) process

$$x_t = \phi x_{t-1} + z_t \text{, where } z_t \sim \text{IID noise}$$

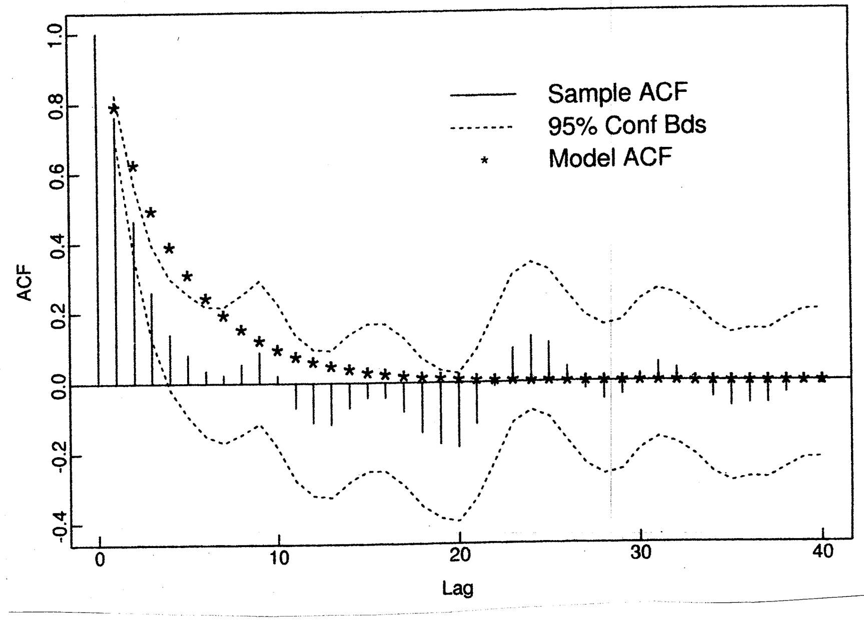
and $|\phi| < 1$.

We know that $\rho(h) = \phi^{|h|}$.

So

$$\omega_{ii} = \sum_{k=1}^{i} \phi^{2i}(\phi^{-k} - \phi^{k})^2 + \sum_{k=i+1}^{\infty} \phi^{2k}(\phi^{-i} - \phi^{i})^2$$

$$= \frac{(1-\phi^{2i})(1+\phi^2)}{(1-\phi^2)} - 2i\,\phi^{2i}.$$



$$\rho(i) = (.791)^i.$$

In the example the model assumed is

$$Y_t = .791\, Y_{t-1} + z_t.$$

Sample PACF.

- Recall that we defined PACF as a regression coefficient.
- Thus given a sample $x_1, \ldots, x_T$ it can be easily found out by solving a regression problem.
- For an AR(p) model the sample PACF values at lag larger than p are approximately distributed as a $N(0, 1/n)$ random variable.
- Thus roughly 95% of the sample PACF values beyond lag p should fall within the bounds $\pm 1.96/\sqrt{n}$.
- This is mainly used for diagnostics.

Example.

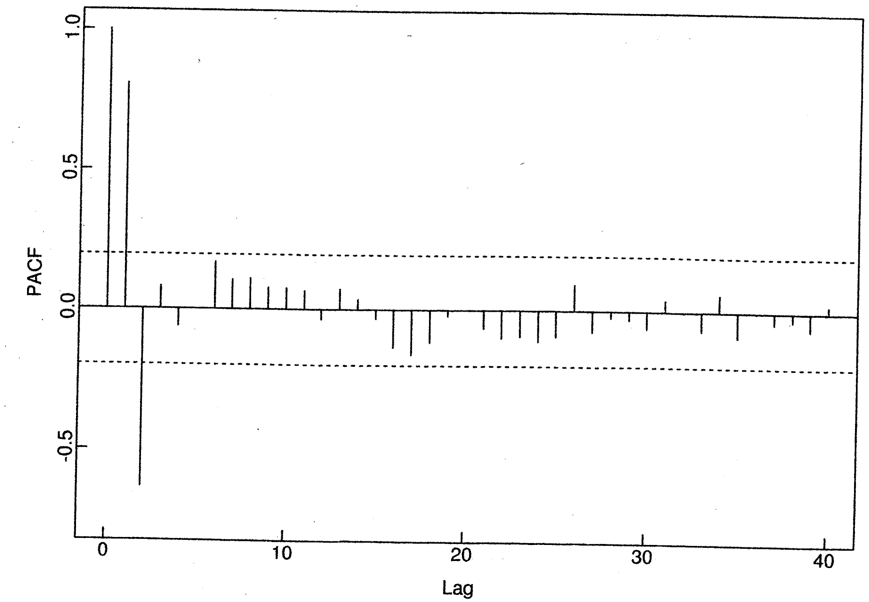
- If we observe that the sample PACF ~~are~~ are such that

  $|\hat{\alpha}(h)| > 1{\cdot}96/\sqrt{n}$ for $0 \leq h \leq p$

  and

  $|\hat{\alpha}(h)| < 1{\cdot}96/\sqrt{n}$ for $h > p$,

  it would indicate that the data comes from an $AR(p)$ process.



- The plot shows an $AR(2)$ process.

bounds $\pm 1{\cdot}96/\sqrt{100}$.

$$X_t - 1{\cdot}318\, X_{t-1} + {\cdot}634\, X_{t-2} = Z_t.$$

$T = 100$, $Z_t \sim WN(0, 289{\cdot}2)$.