Lecture 3

Stationary processes

Readings: Cryer & Chan Ch. 2 & Ch. 4.1-4.3; Brockwell & Davis Ch. 1.3-1.6; Shumway & Stoffer Ch. 1.2-1.6

MATH 8090 Time Series Analysis
August 31 & September 2, 2021

Whitney Huang
Clemson University

Agenda

1. Mean and Autocovaraince Functions
2. Stationarity
3. Some Examples of Stationary Processes
4. Estimation of Mean and Autocovariance Functions
5. Testing Temporal Dependence

Review: The Additive Decomposition

- The additive model for a time series $\{Y_t\}$ is

$$Y_t = \mu_t + s_t + \eta_t,$$

  where

  - μ_t is the trend component
  - s_t is the seasonal component
  - η_t is the random (noise) component with $\mathbb{E}(\eta_t) = 0$

- Standard procedure:

  (1) Estimate/remove the trend and seasonal components

  (2) Analyze the remainder, the residuals $\hat{\eta}_t = y_t - \hat{\mu}_t - \hat{s}_t$

- We will focus on (2) for the next few weeks

Time Series Models

- A time series model is a specification of the probabilistic distribution of a sequence of random variables (RVs) η_t

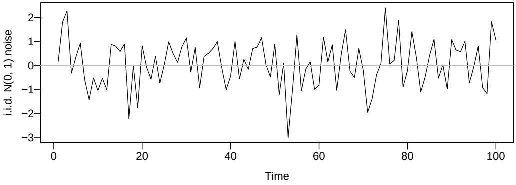
  (The observed time series is a realization of such a sequence of random variables)

- The simplest time series is i.i.d. (*independent and identically distributed*) noise
  - $\{\eta_t\}$ is a sequence of independent and identically distributed zero-mean (i.e., $\mathbb{E}(\eta_t) = 0, \forall t$) random variables $\Rightarrow$ no temporal dependence
  - It is of little value of using i.i.d. noise model to conduct forecast as there is no information from the past observations
  - **But**, we will use i.i.d. model as a building block to develop time series models that can accommodate time dependence

Example Realizations of i.i.d. Noise

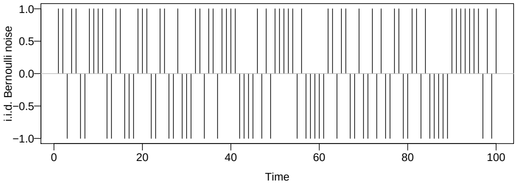
- Gaussian (normal) i.i.d. noise with mean 0 and variance $\sigma^2 > 0$

$$f(\eta_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{\eta_t^2}{2\sigma^2})$$



- Bernoulli i.i.d. noise with "success" probability

$$\mathbb{P}(\eta_t = 1) = p = 1 - \mathbb{P}(\eta_t = -1)$$



Means and Autocovaraince

A time series model could also be a specification of the means and autocovariances of the RVs

- The mean function of $\{\eta_t\}$ is

$$\mu_t = \mathbb{E}(\eta_t).$$

- μ_t is the population mean at time t, which can be computed as:

$$\mu_t = \begin{cases} \int_{-\infty}^{\infty} \eta_t f(\eta_t)\, d\eta_t & \text{when } \eta_t \text{ is a continuous RV;} \\ \sum_{-\infty}^{\infty} \eta_t p(\eta_t), & \text{when } \eta_t \text{ is a discrete RV,} \end{cases}$$

where $f(\cdot)$ and $p(\cdot)$ are the probability density function and probability mass function of η_t, respectively

Examples of Mean Functions

- **Example 1**: What is the mean function for $\{\eta_t\}$, an i.i.d. $\mathrm{N}(0, \sigma^2)$ process?

- **Example 2**: For each time point, let $Y_t = \beta_0 + \beta_1 t + \eta_t$ with β_0 and β_1 some constants and η_t is defined above. What is $\mu_Y(t)$?

Review: The Covariance Between Two RVs

- The covariance between the RVs X and Y is

$$\begin{aligned}\mathbb{Cov}(X,Y) &= \mathbb{E}\{(X-\mu_X)(Y-\mu_Y)\} \\ &= \mathbb{E}(XY) - \mu_X\mu_Y .\end{aligned}$$

  It is a measure of linear dependence between the two RVs. When $X = Y$ we have

$$\mathbb{Cov}(X,X) = \mathbb{Var}(X).$$

- For constants a, b, c, and RVs X, Y, Z:

$$\begin{aligned}\mathbb{Cov}(aX+bY+c,Z) &= \mathbb{Cov}(aX,Z)+\mathbb{Cov}(bY,Z) \\ &= a\mathbb{Cov}(X,Z)+b\mathbb{Cov}(Y,Z)\end{aligned}$$

  $\Rightarrow$

$$\begin{aligned}\mathbb{Var}(X+Y) &= \mathbb{Cov}(X,X)+\mathbb{Cov}(X,Y)+\mathbb{Cov}(Y,X)+\mathbb{Cov}(Y,Y) \\ &= \mathbb{Var}(X)+\mathbb{Var}(Y)+2\mathbb{Cov}(X,Y)\end{aligned}$$

Autocovariance Function

- The autocovariance function of $\{\eta_t\}$ is

$$\gamma(s,t) = \mathbb{C}\mathrm{ov}(\eta_s, \eta_t) = \mathbb{E}[(\eta_s - \mu_s)(\eta_t - \mu_t)]$$

  It measures the strength of linear dependence between two RVs η_s and η_t

- **Properties**:
  - $\gamma(s,t) = \gamma(t,s)$ for each s and t
  - When $s = t$ we have

    $$\gamma(t,t) = \mathbb{C}\mathrm{ov}(\eta_t, \eta_t) = \mathbb{C}\mathrm{ov}(\eta_t) = \sigma_t^2$$

    the value of the variance function at time t
  - $\gamma(s,t)$ is a non-negative definite function (will come back to this later)

Autocorrelation Function

- The autocorrelation function of $\{\eta_t\}$ is

$$\rho(s,t) = \mathbb{C}\mathrm{orr}(\eta_s, \eta_t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}}$$

  It measures the "scale invariant" linear association between η_s and η_t

- **Properties**:
  - $-1 \le \rho(s,t) \le 1$ for each s and t
  - $\rho(s,t) = \rho(t,s)$ for each s and t
  - $\rho(t,t) = 1$ for each t
  - $\rho(\cdot,\cdot)$ is a non-negative definite function

Stationarity

- We typically need "replicates" to estimate population quantities. For example, we use

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

  to be the estimate of μ_X, the population mean of the **single** RV, X
- However, in time series analysis, we have $n = 1$ (i.e., no replication) because we only have one realized value at each time point
- Stationarity means that some characteristic of $\{\eta_t\}$ does not depend on the time point, t, only on the "time lag" between time points **so that we can create "replicates"**

> Next, we will talk about strict stationarity and weak stationarity

Strictly Stationary Processes

- A time series, $\{\eta_t\}$, is strictly stationary if

$$[\eta_1, \eta_2, \cdots \eta_T] \stackrel{d}{=} [\eta_{1+h}, \eta_{2+h}, \cdots \eta_{T+h}],$$

  for all integers h and $T \geq 1 \Rightarrow$ the joint distribution are unaffected by time shifts

- Under such the strict stationarity
  - $\{\eta_t\}$ is identically distributed but not (necessarily) independent
  - When μ_t is finite, $\mu_t = \mu$ is independent of time t
  - When the variance function exists,

$$\gamma(s, t) = \gamma(s + h, t + h),$$

    for any s, t, and h

Weakly Stationary Processes

- $\{\eta_t\}$ is weakly stationary if
  - $\mathbb{E}(\eta_t) = \mu_t = \mu$
  - $\mathbb{Cov}(\eta_t, \eta_{t+h}) = \gamma(t, t+h) = \gamma(h)$, finite constant that can depend on h but not on t
- Other names for this type of stationarity include second-order, covariance, wide senese. The quantity h is called the lag
- Weak and strict stationarity
  - A strictly stationary process $\{\eta_t\}$ is also weakly stationary as long as μ is finite
  - Weak stationarity does not imply strict stationarity!

Autocovariance Function of Stationary Processes

The autocovariance function (ACVF) of a stationary process $\{\eta_t\}$ is defined to be

$$\begin{aligned}\gamma(h) &= \mathbb{Cov}(\eta_t, \eta_{t+h}) \\ &= \mathbb{E}[(\eta_t - \mu)(\eta_{t+h} - \mu)],\end{aligned}$$

which measures the lag-h time dependence

Properties of the ACVF:

- $\gamma(0) = \mathbb{Var}(\eta_t)$
- $\gamma(-h) = \gamma(h)$ for each h
- $\gamma(s-t)$ as a function of $(s-t)$ is non-negative definite

Autocorrelation Function of Stationary Processes

The autocorrelation function (ACF) of a stationary process $\{\eta_t\}$ is defined to be

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

which measures the “scale invariant” lag-h time dependence

Properties of the ACF:

- $-1 \leq \rho(h) \leq 1$ and $\rho(0) = 1$ for each h
- $\rho(-h) = \rho(h)$ for each h
- $\rho(s-t)$ as a function of $(s-t)$ is non-negative definite

The White Noise Process

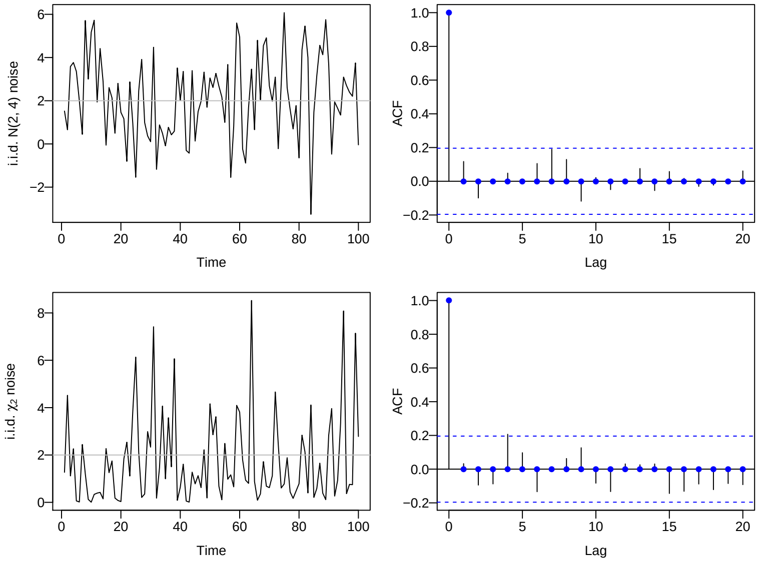
Let's assume $\mathbb{E}(\eta_t) = \mu$ and $\mathbb{Var}(\eta_t) = \sigma^2 < \infty$. $\{\eta_t\}$ is a white noise or $\mathrm{WN}(\mu, \sigma^2)$ process if

$$\gamma(h) = 0,$$

for $h \neq 0$

- $\{\eta_t\}$ is stationary
- However, distributions of η_t and η_{t+1} can be different!
- All i.i.d. noise with finite variance ($\sigma^2 < 0$) is white noise but the converse need not be true

Examples Realizations of White Noise Processes



The Moving Average Process of First Order (MA(1))

Let $\{Z_t\}$ be a $\mathrm{WN}(0, \sigma^2)$ process and θ be some constant $\in \mathbb{R}$. For each integer t, let

$$\eta_t = Z_t + \theta Z_{t-1}.$$

- The sequences of RVs $\{\eta_t\}$ is called the moving average process of order 1 or MA(1) process
- One can show that the MA(1) process $\{\eta_t\}$ is stationary

First-Order Moving Average Process: Mean Function

Need to show mean function is NOT a function of time t

$$
\begin{aligned}
\mathbb{E}[\eta_t] &= \mathbb{E}[Z_t + \theta Z_{t-1}] \\
&= \mathbb{E}[Z_t] + \theta \mathbb{E}[Z_{t-1}] \\
&= 0 + \theta \times 0 \\
&= 0, \quad \forall t
\end{aligned}
$$

☺

First-Order Moving Average Process: Covariance Function

Need to show the autovariance function $\gamma(\cdot,\cdot)$ is a function of time lag only

$$\begin{aligned}\gamma(t,t+h) &= \mathbb{Cov}(\eta_t,\eta_{t+h}) \\ &= \mathbb{Cov}(Z_t+\theta Z_{t-1}, Z_{t+h}+\theta Z_{t+h-1}) \\ &= \mathbb{Cov}(Z_t, Z_{t+h}) + \mathbb{Cov}(Z_t, \theta Z_{t+h-1}) \\ &+ \mathbb{Cov}(\theta Z_{t-1}, Z_{t+h}) + \mathbb{Cov}(\theta Z_{t-1}, \theta Z_{t+h-1})\end{aligned}$$

if $h=0,$ we have $\gamma(t,t+h)=\sigma^2+\theta^2\sigma^2=\sigma^2(1+\theta^2)$
if $h=\pm 1,$ we have $\gamma(t,t+h)=\theta\sigma^2$
if $|h|\geq 2,$ we have $\gamma(t,t+h)=0$

$\Rightarrow \gamma(t,t+h)$ only depends on h but not on t ☺

First-Order Moving Average Process: ACVF & ACF

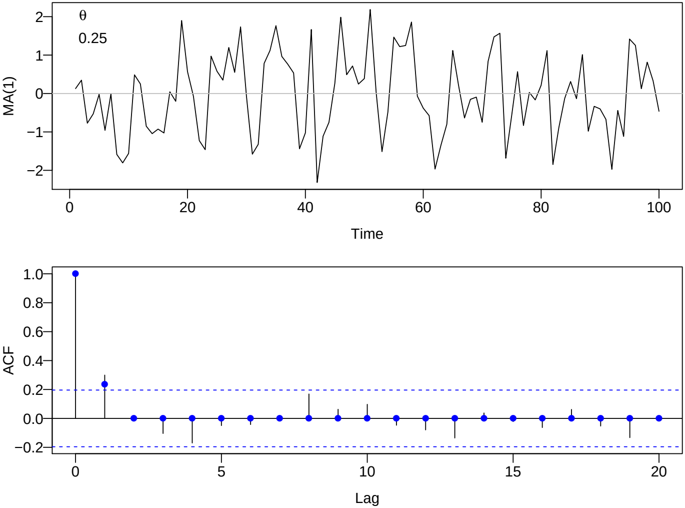
ACVF:

$$\gamma(h) = \begin{cases} \sigma^2(1+\theta^2) & h = 0; \\ \theta\sigma^2 & |h| = 1; \\ 0 & |h| \geq 2 \end{cases}$$

We can get **ACF** by dividing everything by $\gamma(0) = \sigma^2(1+\theta^2)$

$$\rho(h) = \begin{cases} 1 & h = 0; \\ \frac{\theta}{1+\theta^2} & |h| = 1; \\ 0 & |h| \geq 2. \end{cases}$$

Examples Realizations of MA(1) Processes



First-order autoregressive process, AR(1)

Let $\{Z_t\}$ be a $\mathrm{WN}(0, \sigma^2)$ process, and $-1 < \phi < 1$ be a constant. Let's assume $\{\eta_t\}$ is a stationary process with

$$\eta_t = \phi\eta_{t-1} + Z_t,$$

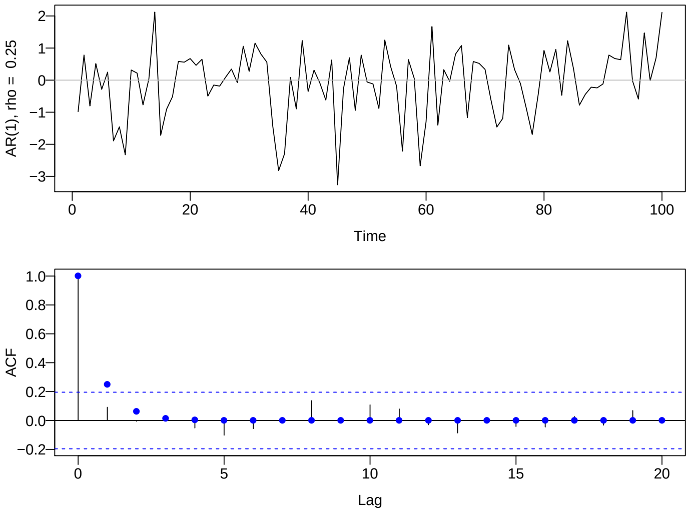
for each integer t, where η_s and Z_t are uncorrelated for each $s < t \Rightarrow$ future noise is uncorrelated with the current time point)

We will see later there is only one unique solution to this equation. Such a sequence $\{\eta_t\}$ of RVs is called an AR(1) process

Properties of the AR(1) process

Properties of the AR(1) process Cont'd

Examples Realizations of AR(1) Processes



The Random Walk Process

Let $\{Z_t\}$ be a $\mathrm{WN}(0, \sigma^2)$ process and for $t \geq 1$ definite

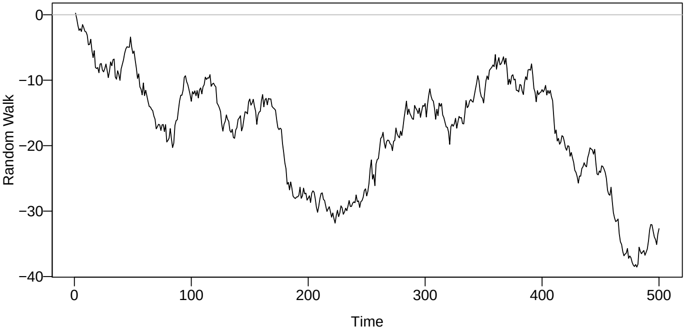
$$\eta_t = Z_1 + Z_2 + \cdots + Z_t = \sum_{s=1}^{t} Z_s.$$

- The sequence of RVs $\{\eta_t\}$ is called a random walk process
- **Special case**: If we have $\{Z_t\}$ such that for each t

$$\mathbb{P}(Z_t = z) = \begin{cases} \frac{1}{2}, & z = 1; \\ \frac{1}{2}, & z = -1, \end{cases}$$

  then $\{\eta_t\}$ is a simple symmetric random walk
- The random walk process is not stationary!

Example Realizations of Random Walk Processes



Gaussian Processes

$\{\eta_t\}$ is a Gaussian process (GP) if the joint distribution of any collection of the RVs has a multivariate normal (aka Gaussian) distribution

- The distribution of a GP is fully characterized by $\mu(\cdot)$, the mean function, and $\gamma(\cdot,\cdot)$, the autocovariance function. The joint probability density function of $\boldsymbol{\eta} = (\eta_1, \eta_2, \cdots, \eta_T)^T$ is

$$f(\boldsymbol{\eta}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{\eta} - \boldsymbol{\mu})\right),$$

  where $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_T)^T$ and the (i, j) element of the covariance matrix Σ is $\gamma(i, j)$

- If a GP $\{\eta_t\}$ is weakly stationary then the process is also strictly stationary

Estimating the Mean of Stationary Processes

Let $\{\eta_t\}$ be stationary with mean μ and ACVF $\gamma(s,t)$

- A natural estimator of μ is the sample mean

$$\bar{\eta} = \frac{1}{T}\sum_{t=1}^{T}\eta_t.$$

  $\bar{\eta}$ is an unbiased estimator of μ, i.e.

- Since $\{\eta_t\}$ is stationary, we have

$$\begin{aligned}\mathbb{Var}(\bar{\eta}) &= \frac{1}{T^2}\mathbb{Var}\left(\sum_{i=1}^{T}\eta_t\right)\\ &= \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbb{Cov}(\eta_s,\eta_t)\\ &= \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma(s-t)\end{aligned}$$

- **Exercise**: Show

$$\mathbb{Var}(\bar{\eta}) = \frac{1}{T}\sum_{h=-(T-1)}^{T-1}\left(1-\frac{|h|}{T}\right)\gamma(h)$$

AR(1) Example

Suppose $\{\eta_1, \eta_2, \eta_3\}$ is an AR(1) process with $|\phi| < 1$ and innovation variance σ. Show that the variance of $\bar{\eta}$ is $\frac{\sigma^2}{9}(3 + 4\phi + 2\phi^2)$

Solution:

The Sampling Distribution of $\bar{\eta}$

Let

$$v_T = \sum_{h=-(T-1)}^{(T-1)} \left(1 - \frac{|h|}{T}\right) \gamma(h)$$

- If $\{\eta_t\}$ is Gaussian we have

$$\sqrt{T}(\bar{\eta} - \mu) \sim \mathrm{N}(0, v_T)$$

- The result above is approximate for many non-Gaussian time series

- In practice we also need to estimate $\gamma(h)$ from the data

Confidence Intervals for μ

- If $\gamma(h) \to 0$ as $h \to \infty$ then

$$v = \lim_{T\to\infty} v_T = \sum_{h=-\infty}^{\infty} \gamma(h) \text{ exists.}$$

- Further, if $\{\eta_t\}$ is Gaussian and

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty,$$

then an approximate large-sample 95% CI for μ is given by

$$\left[\bar{\eta} - 1.96\sqrt{\frac{v}{T}}, \bar{\eta} + 1.96\sqrt{\frac{v}{T}}\right]$$

Strategies for Estimating v

1. Parametric:

   - Assume a parametric model $\gamma_{\boldsymbol{\theta}}(\cdot)$, and calculate $v = \sum_{h=-(T-1)}^{T-1} \left(1 - \frac{|h|}{T}\right) \gamma_{\hat{\boldsymbol{\theta}}}(h)$ based on the ACVF for that model
   - The standard error, v, will depend on the parameters $\boldsymbol{\theta}$ of the parametric model

- Nonparametric:

   - Estimate v by

$$\hat{v} = \sum \left(1 - \frac{|h|}{T}\right) \hat{\gamma}(h),$$

   where $\hat{\gamma}(\cdot)$ is an nonparametric estimate of ACVF

Examples of Parametric Forms for v

- i.i.d. Gaussian Noise: $v = \gamma(0) = \sigma \Rightarrow$ CI reduces to the classical case:

$$\left[\bar{\eta} - 1.96\sqrt{\frac{\sigma}{T}}, \bar{\eta} + 1.96\sqrt{\frac{\sigma}{T}}\right]$$

- MA(1) process: We have

$$\begin{aligned} v = \sum_{h=-\infty}^{\infty} \gamma(h) &= \gamma(-1) + \gamma(0) + \gamma(1) \\ &= \gamma(0) + 2\gamma(1) \\ &= \sigma^2(1 + \theta^2 + 2\theta) = \sigma^2(1 + \theta)^2 \end{aligned}$$

- **Exercise**: Show for an AR(1) process we have

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

An Estimate of $\gamma(\cdot)$

Goal: Want to estimate

$$\begin{aligned}\gamma(h) &= \mathbb{C}\mathrm{ov}(\eta_t, \eta_{t+h}) \\ &= \mathbb{E}\left[(\eta_t - \mu)(\eta_{t+h} - \mu)\right]\end{aligned}$$

using data $\{\eta_t\}_{t=1}^{T}$

- For $|h| < T$, consider $\hat{\gamma}(h) = \frac{1}{T}\sum_{t=1}^{T-|h|}(\eta_t - \bar{\eta})(\eta_{t+|h|} - \bar{\eta})$. We call $\hat{\gamma}(h)$ the sample ACVF
- The sample ACVF is a biased estimator of $\gamma(h)$, but, it is used as the **standard** estimate of $\gamma(h)$
- $\hat{\gamma}(h)$ are even and non-negative definite

The Sample Autocorrelation Function

- The sample autocorrelation function (ACF) is defined for $|h| < T$ by

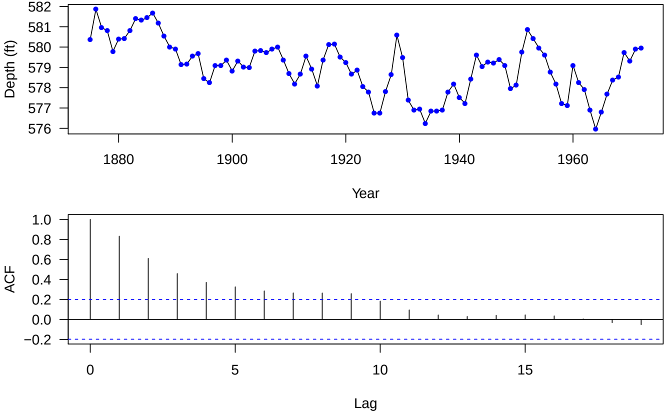
$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}.$$

- **Rule of thumb**: Box and Jenkins (1976) recommend using $\hat{\rho}(h)$ and $\hat{\gamma}(h)$ only for $\frac{|h|}{T} \leq \frac{1}{4}$ and $T \geq 50$

- This is because estimates $\hat{\rho}(h)$ and $\hat{\gamma}(h)$ are unstable for large $|h|$ as there will be no enough data points going into the estimator

Calculating the Sample ACF in R

- We use `acf` function to calculate the sample ACF
- Lake Huron Example



Asymptotic Distribution of the Sample ACF [Bartlett, 1946]

Let $\{\eta_t\}$ be a stationary process we suppose that the ACF

$$\boldsymbol{\rho} = (\rho(1), \rho(2), \cdots, \rho(k))^T$$

is estimated by

$$\hat{\boldsymbol{\rho}} = (\hat{\rho}(1), \hat{\rho}(2), \cdots, \hat{\rho}(k))^T$$

- For large T

  $$\hat{\boldsymbol{\rho}} \dot{\sim} \mathrm{N}_k(\boldsymbol{\rho}, \frac{1}{T}W),$$

  where N_k is the K-variate normal distribution and W is an $k \times k$ covariance matrix with (i, j) element defined by

  $$w_{ij} = \sum_{k=1}^{\infty} a_{ik} a_{jk},$$

  where $a_{ik} = \rho(k+i) + \rho(k-i) - 2\rho(k)\rho(i)$

Using the ACF as a Test for i.i.d. Noise

When $\{\eta_t\}$ is an i.i.d. process with finite variance, Bartlett's result simplifies for each $h \neq 0$

$$\hat{\rho}(h) \dot{\sim} \mathrm{N}(0, \frac{1}{h}).$$

This suggests a diagnostic for i.i.d. noise:

1. Plot the lag h versus the sample ACF $\hat{\rho}(h)$
2. Draw two horizontal lines at $\pm \frac{1.96}{T}$ (blue dashed lines in R)
3. About 95% of the $\{\hat{\rho}(h) : h = 1, 2, 3, \cdots\}$ should be within the lines if we have i.i.d. noise

The Portmanteau Test [Box and Pierce, 1970]for i.i.d. Noise

Suppose we wish to test:

$H_0 : \{\eta_1, \eta_2, \cdots, \eta_T\}$ is an i.i.d. noise sequence
$H_1 : H_0$ is false

- Under H_0,

$$\hat{\rho}(h) \dot{\sim} \mathrm{N}(0, \frac{1}{T}) \overset{d}{=} \frac{1}{\sqrt{n}}\mathrm{N}(0,1)$$

- Hence

$$Q = T\sum_{i=1}^{k}\hat{\rho}^2(h) \dot{\sim} \chi^2_{df=k}$$

- We reject H_0 if $Q > \chi^2_k(1-\alpha)$, the $1-\alpha$ quatile of the chi-squared distribution with k degrees of freedom

Ljung-Box Test [Ljung and Box, 1978]

Ljung and Box [1978] showed that

$$Q_{LB} = T(T-2)\sum_{h=1}^{k}\frac{\hat{\rho}^2(h)}{n-h} \dot{\sim} \chi^2_k.$$

The Ljung-Box test can be more powerful than the Portmanteau test

Both the Portmanteau Test (aka Box-Pierce test) and Ljung-Box test can be carried out in `R` using the function `Box.text`

Examples in R

```
> Box.test(rnorm(100), 20)

        Box-Pierce test

data:  rnorm(100)
X-squared = 12.197, df = 20, p-value = 0.9091

> Box.test(LakeHuron, 20)

        Box-Pierce test

data:  LakeHuron
X-squared = 182.43, df = 20, p-value < 2.2e-16

> Box.test(LakeHuron, 20, type = "Ljung")

        Box-Ljung test

data:  LakeHuron
X-squared = 192.6, df = 20, p-value < 2.2e-16
```