Advanced Methods in Biostatistics II

Lecture 9

November 21, 2017

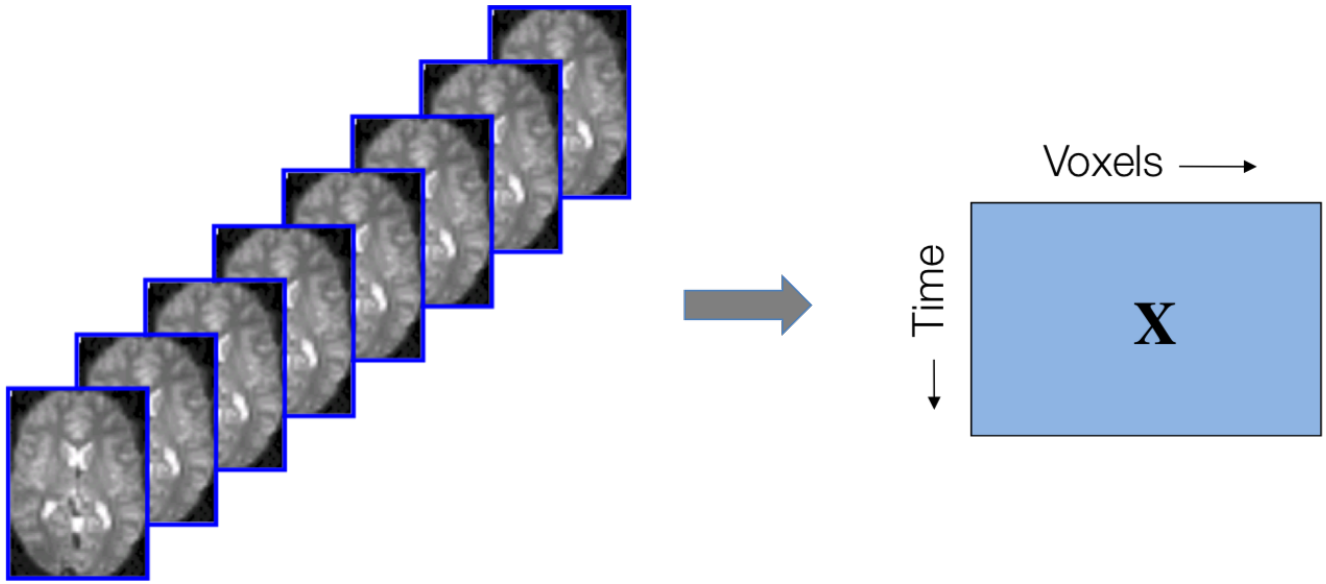
Principal component analysis

- Last time we discussed principal components analysis (PCA) and principal components regression (PCR).
- Suppose we use SVD to write: $\mathbf{X} = \mathbf{UDV}'$.
- Then we can express the principal components as follows:

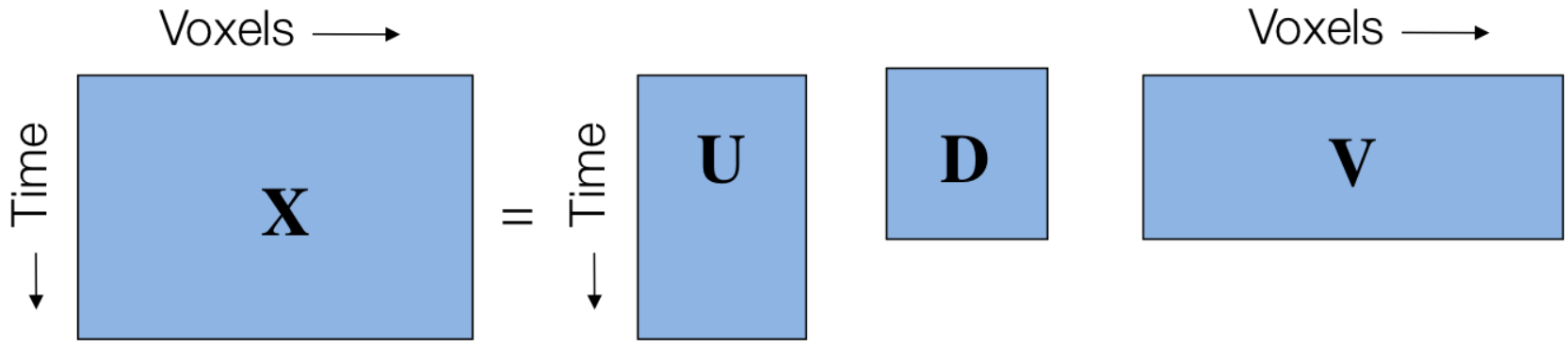
$$\begin{aligned} \mathbf{Z} &= \mathbf{XV} \\ &= \mathbf{UD}. \end{aligned}$$

Example

- Consider performing PCA on functional magnetic resonance imaging (fMRI) data.

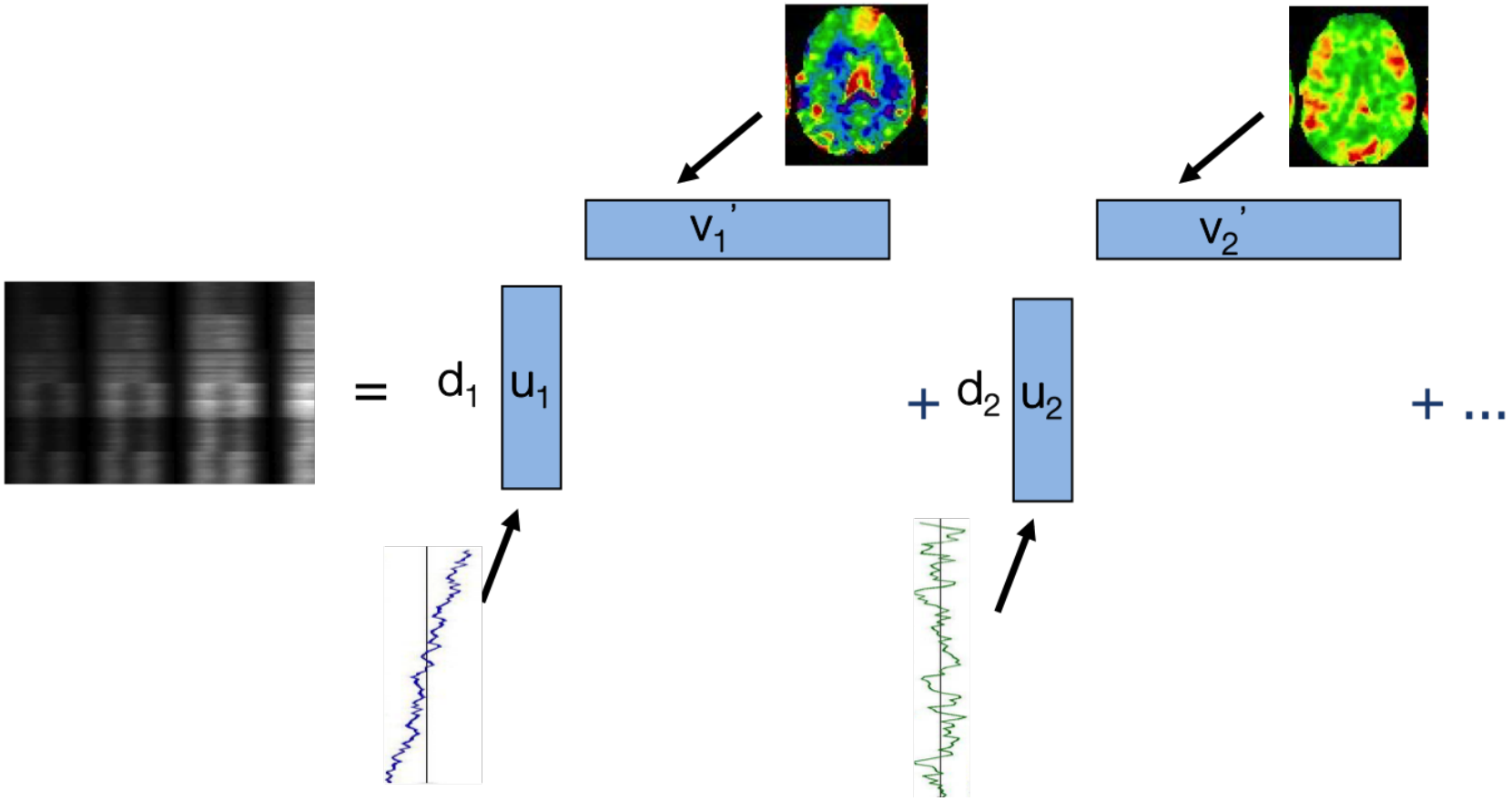


Example

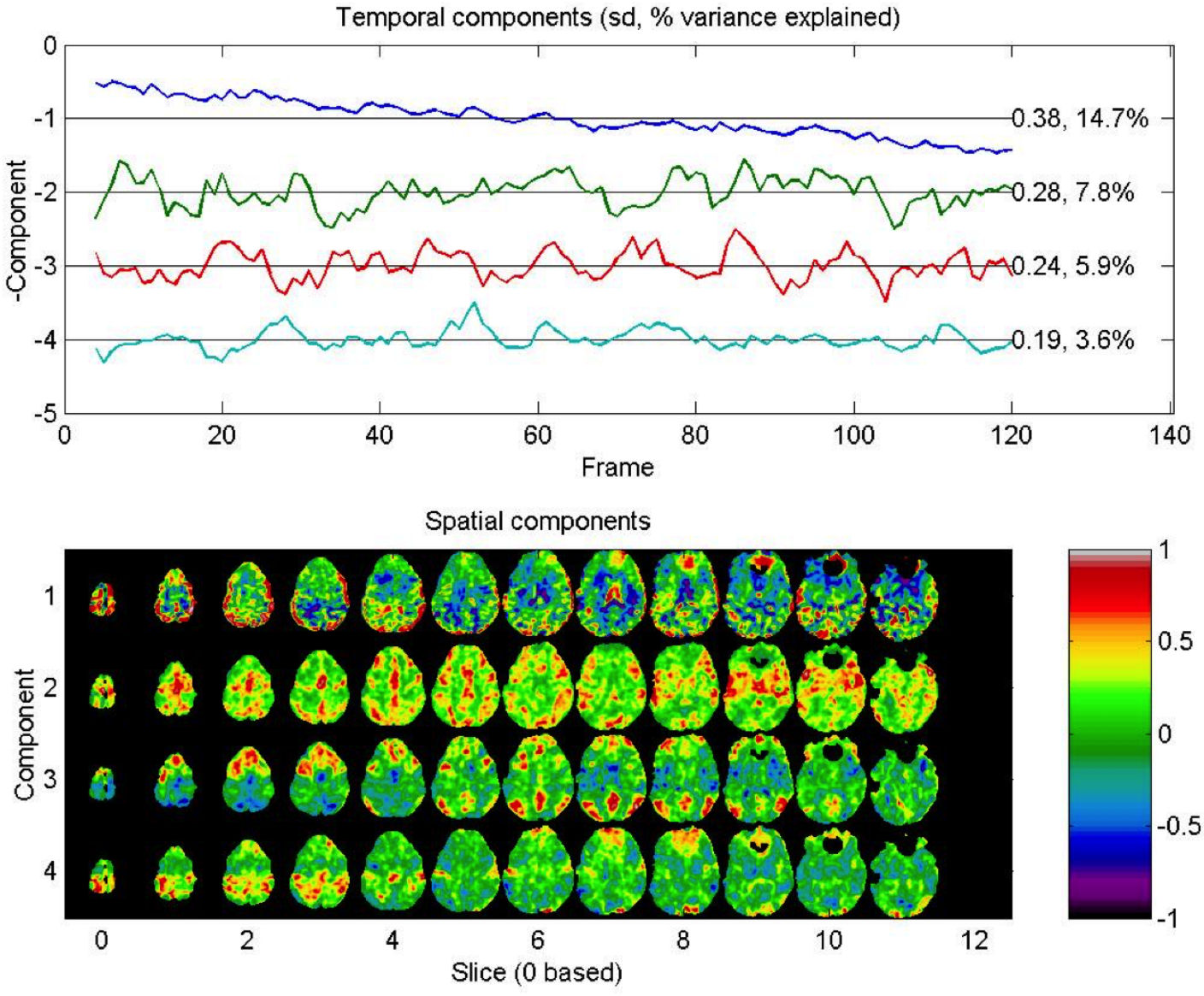


$$\begin{aligned} \mathbf{X} &= \mathbf{UDV}' \\ &= d_1\mathbf{u}_1\mathbf{v}_1' + d_2\mathbf{u}_2\mathbf{v}_2' + \ldots + d_p\mathbf{u}_p\mathbf{v}_p' \end{aligned}$$

Example



Example


Temporal components (sd, % variance explained)
0
-1
-2
-3
-4
-5
-Component
0.38, 14.7%
0.28, 7.8%
0.24, 5.9%
0.19, 3.6%
0
20
40
60
80
100
120
140
Frame
Spatial components
1
2
3
4
Component
0
2
4
6
8
10
12
Slice (0 based)
1
0.5
0
-0.5
-1

Principal component regression

- Principal components regression (PCR) uses $\mathbf{Z}$ instead of $\mathbf{X}$ as the explanatory variables in a linear model.
- Importantly, only the variables $\mathbf{z}_1, \ldots \mathbf{z}_m$ for some $m \leq p$ are typically used in the analysis.
- The $p - m$ components with the smallest eigenvalues are disgarded.
- By manually setting the projection onto the principal component directions with small eigenvalues equal to 0, dimension reduction is achieved.

Principal component regression

- Let $\mathbf{Z}_{(m)} = [\mathbf{z}_1, \ldots \mathbf{z}_m]$.
- Then, we can express the principal components regression model as follows:

$$\mathbf{y} = \mathbf{Z}_{(m)}\gamma_{(m)} + \varepsilon.$$

- Under this formulation,

$$\begin{aligned}\hat{\gamma}_{(m)} &= (\mathbf{Z}'_{(m)}\mathbf{Z}_{(m)})^{-1}\mathbf{Z}'_{(m)}\mathbf{y} \\ &= \mathbf{D}^{-2}_{(m)}\mathbf{Z}'_{(m)}\mathbf{y} \\ &= \mathbf{D}^{-1}_{(m)}\mathbf{U}'_{(m)}\mathbf{y}.\end{aligned}$$

Principal component regression

- The fitted values are given by:

$$\hat{\mathbf{y}}_{(m)} = \mathbf{H}_m \mathbf{y}$$

where

$$\begin{aligned} \mathbf{H}_m &= \mathbf{Z}_{(m)}(\mathbf{Z}'_{(m)}\mathbf{Z}_{(m)})^{-1}\mathbf{Z}'_{(m)} \\ &= \mathbf{U}_{(m)}\mathbf{U}'_{(m)} \end{aligned}$$

- Thus, we can write:

$$\begin{aligned} \hat{\mathbf{y}}_{(m)} &= \sum_{j=1}^{m} \mathbf{u}_j \mathbf{u}'_j \mathbf{y} \\ &= \sum_{j=1}^{m} \mathbf{u}_j \langle \mathbf{u}_j, \mathbf{y} \rangle. \end{aligned}$$

Ridge regression - revisited

- Recall that for ridge regression, we have:

$$\hat{\mathbf{y}}_r = \mathbf{H}_\lambda \mathbf{y}$$

where

$$\mathbf{H}_\lambda = \mathbf{U}\mathbf{W}\mathbf{U}'.$$

- Here $\mathbf{W}$ is a diagonal matrix whose elements are:

$$\frac{d_i^2}{d_i^2 + \lambda}$$

where d_i are the diagonal elements of $\mathbf{D}$ (i.e., the eigenvalues).

Ridge regression - revisited

- Thus, we can write:

$$\hat{\mathbf{y}}_r = \sum_{j=1}^{p} \frac{d_i^2}{d_i^2 + \lambda} \mathbf{u}_j \langle \mathbf{u}_j, \mathbf{y} \rangle.$$

- Ridge regression projects the vector $\mathbf{y}$ onto the principal component directions and then shrinks the projection on each direction.
- The amount of shrinkage depends on the variance of that principal component.
- In contrast, PCR sets directions with small variance equal to zero a priori.

High dimensional settings

- Most linear model techniques are designed for the low-dimensional setting (i.e., $n >> p$).
- However, increasingly we are faced with situations where $p > n$ (so-called small n, large p problems).
- Many traditional approaches are not appropriate in this setting, as they will overfit the data.

High dimensional settings

- Techniques such as ridge regression, the Lasso, and PCR are more appropriate.
- By either constraining the solution or performing data reduction, overfitting can be avoided.

Linear Models with Autocorrelated Errors

- When observations are measured at equally spaced time points and the error terms from adjacent time points are correlated, we say that the error term is serially correlated or autocorrelated.
- A major cause of autocorrelated error is the omission of one or more key explanatory variable in the model.
- If the time-ordered effects of the missing variable are positively correlated, then the error terms also tend to be positively correlated.
- This violates the usual assumption of independent errors made in the linear model.

Linear Models with Autocorrelated Errors

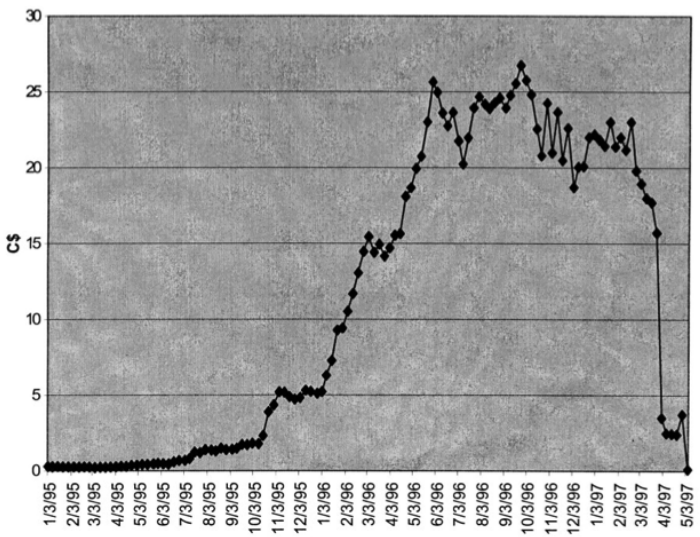
- Serial correlation will not affect the unbiasedness or consistency of least-squares estimators, but it does affect their efficiency.
- With positive serial correlation, the estimated standard errors will be smaller than the true standard errors.
- This will lead to an increased number of false positives.

Time series analysis

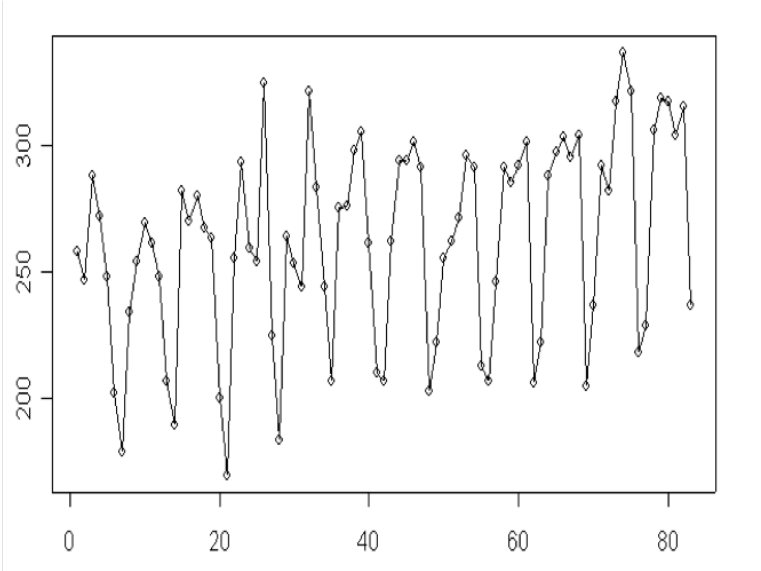
- To properly understand autocorrelated errors we begin with discussing methods from time series analysis.
- If a random variable X is indexed in time, the observations $\{X_t, t \in T\}$ is called a time series.
- A time series $X_t, t \in T$ can be regarded as a realization of a stochastic process.
- We are in particular interested in discrete equally spaced time series.

Examples

Stock prices for Bre-X in Canadian Dollars



The number of daily births in Quebec Canada between Monday Jan. 8, 1990 to Saturday March 31, 1990.



Time series analysis

- A complete description of a time series, observed as a collection of N random variables, is provided by its joint distribution function.
- Often these functions are not easily expressed and instead we focus on second order properties of the time series.

Second-order properties

1. The mean function of X_t is $\mu_X(t) = E(X_t)$.
2. The variance function of X_t is $\sigma_X^2(t) = E(X_t - \mu_X(t))$.
3. The autocovariance function of X_t is defined as:

$$\gamma_X(r, s) = \text{cov}(X_r, X_s) = E((X_r - E(X_r))(X_s - E(X_s)))$$

for $s, t \in T$.

4. The autocorrelation function of X_t is defined as:

$$\rho_X(r, s) = \frac{\gamma_X(r, s)}{\sqrt{\gamma_X(r, r)\gamma_X(s, s)}}.$$

Stationarity

- A central feature in the development of time series models is the assumption of some form of regularity over time in the time series.
- Often we assume that the behavior of a time series in an interval $[t, t + h)$ closely resembles that in the interval $[s, s + h)$.
- This implies there is a temporal homogeneity in the behavior of the series, called stationarity.

Strict stationarity

- A time series X_t is strictly stationary if all of its finite dimensional distributions are time invariant.
- This implies that

$$P(X_{i_1} < \alpha_1, \ldots, X_{i_k} < \alpha_k) = P(X_{h+i_1} < \alpha_1, \ldots, X_{h+i_k} < \alpha_k)$$

for any collection of time points $i_1 \ldots i_k$, any positive integer k and integer-valued lag h.

Weak stationarity

A time series X_t is weakly stationary if

1. $E|X_t|^2 < \infty$ for all t.
2. The mean function $\mu_X(t)$ does not depend on t.
3. The covariance function

$$\gamma_X(t, t+h)$$

is independent of t for all h.

Autocovariance and autocorrelation function

- If X_t is a weakly stationary time series then the autocovariance function (ACVF) at lag h can be written:

$$\gamma_X(h) = \text{cov}(X_{t+h}, X_t)$$

- When $h = 0$, we have that $\gamma_X(0) = \text{var}(X_t)$.

- Hence, the autocorrelation function (ACF) of X_t at lag h can be written:

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}.$$

Autocovariance and autocorrelation function

- For a stationary process X_t, the autocovariance and autocorrelation functions have the following properties:
  1. $\gamma_X(0) = \text{var}(X_t)$ and $\rho_X(0) = 1$.
  2. $|\gamma_X(k)| \leq \gamma_X(0)$ and $|\rho_X(k)| \leq 1$.
  3. $\gamma_X(k) = \gamma_X(-k)$ and $\rho_X(k) = \rho_X(-k)$.

Comments on stationarity

- A strictly stationary process with finite second moment is also weakly stationary.
- However, the converse is not necessarily true.
- An exception is that if X_t is a weakly stationary Gaussian process then it is also strictly stationary.
- Stationary processes play an important roll in the analysis of time series.

Comments on stationarity

- In reality, many observed time series are non-stationary.
- They evolve over time and have trends in their mean or variance.
- Transformations such as differencing or detrending are often used to make them stationary.

Time series models

- There exist a number of commonly used time series models.
- These include autoregressive (AR) processes, moving-average (MA) processes, and autoregressive moving average (ARMA) processes,
- We begin by discussing white noise processes which are the building block of the processes mentioned above.

White noise

- A sequence of uncorrelated random variables Z_t, each with mean 0 and variance σ^2, is called white noise.
- This is written: $Z_t \sim WN(0, \sigma^2)$.
- In Gaussian white noise, the Z_t are independent identically distributed (i.i.d.) normal random variables with mean 0 and variance σ^2.

White noise

- A white noise process Z_t has the following properties:

$$E(Z_t) = 0 \quad \forall t,$$

$$Var(Z_t) = \sigma^2 \quad \forall t$$

and

$$\rho(h) = \begin{cases} 1, & \text{if } h = 0 \\ 0, & \text{if } h \neq 0 \end{cases}$$

White noise

- A white noise process is weakly stationary (but not necessarily strictly stationary, since the Z_i's are not necessarily identically distributed).
- An i.i.d. sequence of random variables is white noise if it has a finite variance.
- White noise processes are the fundamental building blocks of weakly stationary processes and play a crucial role in time series analysis.

AR(p) process

- A time series X_t is an autoregressive process of order p, written AR(p):

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + Z_t$$

  where $Z_t \sim WN(0, \sigma^2)$ and $\phi_1, \phi_2, \ldots \phi_p$ are constants.

- An AR(p) process regresses X_t on its past values.

AR(1) process

- A special case is the AR(1) process:

$$X_t = \phi X_{t-1} + Z_t$$

where $Z_t \sim WN(0, \sigma^2)$ and $|\phi| < 1$.

AR(1) process

- The AR(1) process has the following properties:

$$E(X_t) = 0 \quad \forall t,$$

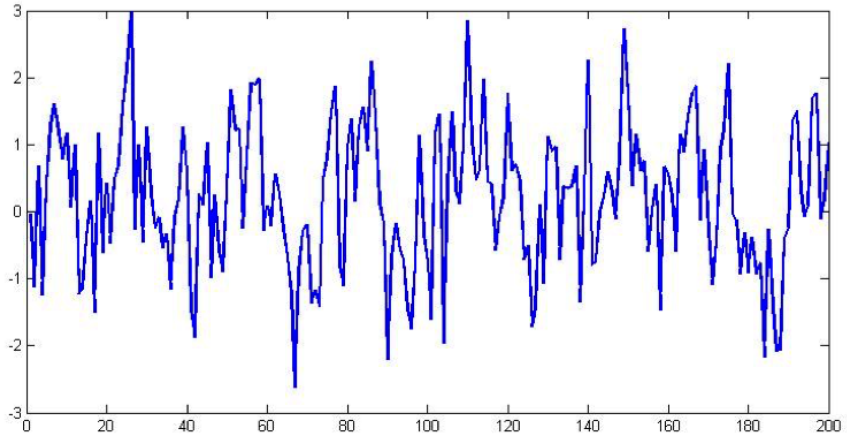
$$Var(X_t) = \frac{\sigma^2}{1-\phi^2} \quad \forall t$$

and

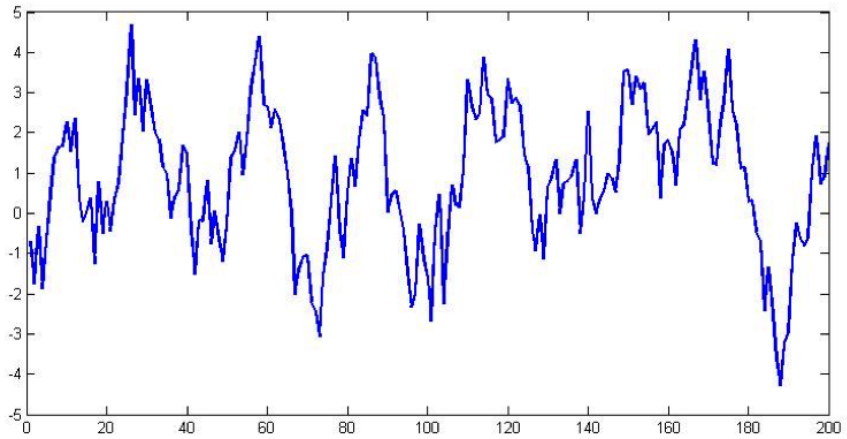
$$\rho(h) = \begin{cases} 1, & \text{if } h = 0 \\ \phi^{|h|}, & \text{if } h \neq 0 \end{cases}$$

Examples - AR(1)

ϕ=0.5



ϕ=0.9



MA(q) process

- A time series is a moving-average process of order q, written MA(q), if

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q}$$

where $Z_t \sim WN(0, \sigma^2)$ and $\theta_1, \theta_2, \ldots \theta_q$ are constants.

MA(1) process

- A special case is the MA(1) process:

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

where $Z_t \sim WN(0, \sigma^2)$ and θ are constants.

MA(1) process

- The MA(1) process has the following properties:

$$E(X_t) = 0 \quad \forall t,$$

$$Var(X_t) = (1 + \theta^2)\sigma^2 \quad \forall t$$

and

$$\rho(h) = \begin{cases} 1, & \text{if } h = 0 \\ \frac{\theta}{1+\theta^2}, & \text{if } h \pm 1 \\ 0 & \text{if } |h| > 2 \end{cases}$$

ARMA(p,q) model

- A time series is an autoregressive moving-average process, written ARMA(p,q), if

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} \\ &\quad + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q} \end{aligned}$$

where $Z_t \sim WN(0, \sigma^2)$ and $\phi_1, \phi_2, \ldots \phi_p, \theta_1, \theta_2, \ldots \theta_q$ are constants.