

# Zeitreihen Mathematische Modelle

Peter Büchel

HSLU I

Stoc: Block 13

# Mathematische Modelle für Zeitreihen

- Bisher: Zeitreihen als Beobachtungen von Daten eingeführt, die auf natürliche Weise chronologisch geordnet werden können
- Konzepte für Transformationen, Visualisierungen und Zerlegungen kennengelernt
- Berechnungen von Zeitreihen in **Python** studiert
- Schritt weiter: *Modellieren* von Zeitreihen

# Mathematische Konzepte für Zeitreihen

- Ziel der Zeitreihenanalyse: Mathematisches Modell zu entwickeln, das eine plausible Beschreibung der Versuchsdaten liefert
- Beschreibung des Charakters dieser scheinbar zufällig fluktuierenden Daten: Zeitreihen als Realisierung von zeitlich indexierten Zufallsvariablen

## Zeitreihen und diskrete stochastische Prozesse

Sei  $T$  eine Menge von Zeitpunkten, die gleichweit auseinander liegen

$$T = \{t_1, t_2, \dots\}$$

- ① Ein *diskreter stochastischer Prozess* ist eine Menge von Zufallsvariablen

$$\{X_1, X_2, \dots\}$$

Jede einzelne Zufallsvariable  $X_i$  hat eine eindimensionale Verteilungsfunktion  $F_i$  und kann zur Zeit  $t_i$  beobachtet werden

- ② Eine *Zeitreihe*

$$\{x_1, x_2, \dots\}$$

ist eine Realisierung eines diskreten stochastischen Prozesses  $\{X_1, X_2, \dots\}$

Wert  $x_i$  ist eine Realisierung der Zufallsvariable  $X_i$ , die zur Zeit  $t_i$  gemessen wird

- Wichtig: Unterscheidung

- ▶ Zeitreihe:

- Konkrete Beobachtung von Werten → mit Kleinbuchstaben bezeichnet

- ▶ Stochastischer Prozess:

- Theoretisches Konstrukt, der den zugrundeliegenden Mechanismus der Zeitreihe modelliert, der die Werte erzeugt → mit Grossbuchstaben bezeichnet

## Beispiel: Random Walk

- Person bewegt sich vom Ursprung in  $x$ -Richtung
- Bei jedem Schritt entscheidet die Person *zufällig*, ob sie 1 m nach links oder nach rechts geht
- Dies ist der einfachste Fall eines *Random Walk*
- Probabilistisches Modell für diesen Random Walk wäre
  - ▶ Wählen  $n$  unabhängige Bernoulli-Zufallsvariablen

$$D_1, \dots, D_n$$

die die Werte  $-1$  und  $1$  mit gleicher W'keit von  $p = 0.5$  annehmen

- ▶ Definieren Zufallsvariable

$$X_i = D_1 + \dots + D_i$$

für jedes  $i$  zwischen  $1$  und  $n$

- ▶ Dann ist

$$X_1, X_2, \dots$$

ein diskreter stochastischer Prozess der den Random Walk modelliert.

- Python-Code berechnet einen besonderen Fall dieses Prozesses, i.e. eine Zeitreihe  $\{x_1, x_2, \dots\}$ .

```
import matplotlib.pyplot as plt
import numpy as np

d = np.random.choice(a=[-1,1], size=10000, replace=True)

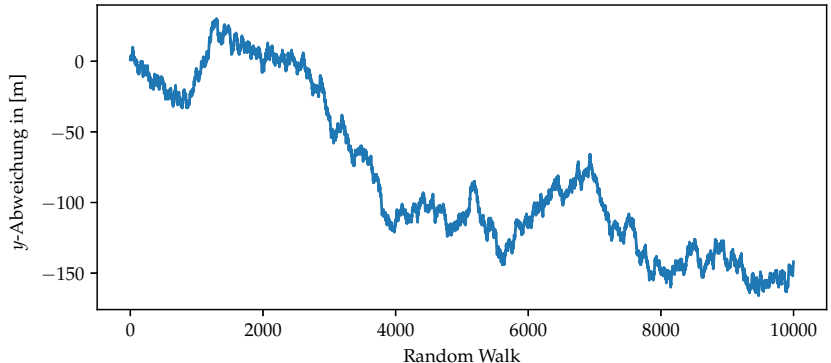
x = np.cumsum(d)

plt.plot(x)

plt.xlabel("Random Walk")
plt.ylabel("y-Abweichung in [m]")

plt.show()
```

- Plot:



- Bei jedem Durchlauf wird ein anderer Random Walk erzeugt



- Aus Definition des Prozesse: Folgende rekursive Definition äquivalent:

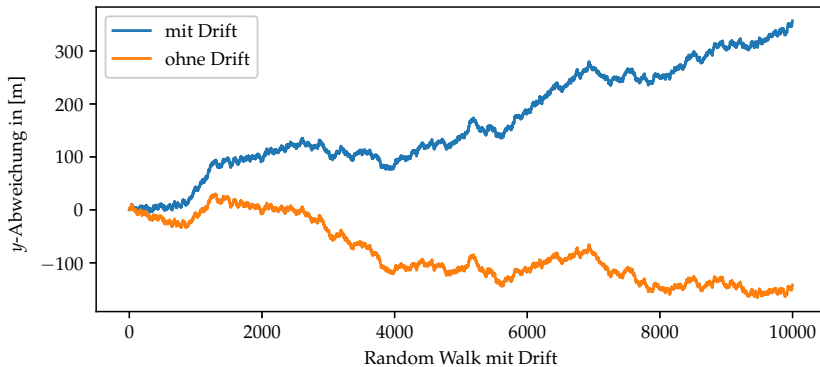
$$X_i = X_{i-1} + D_i, \quad X_0 = 0$$

- Zeitreihe mit einem *Drift*: Jedem Schritt wird eine fixe Konstante  $\delta$  zur Zeitreihe addiert:

$$Y_i = \delta + Y_{i-1} + D_i, \quad Y_0 = 0$$

- Folgende Abbildung: beobachtete Zeitreihe eines solchen Prozesses
- Random Walk mit Drift-Modellen wird verwendet um den Trend einer Zeitreihe zu modellieren

● Plot:



- Simulieren diesen Prozess mit einer `for`-Schleife

```
np.random.seed(35)

d = np.random.choice(a=[-1,1], size=10000, replace=True)

delta = 5*10**(-2)

x = np.cumsum(d)

y = np.zeros(10000)

for i in range(1,10000):
    y[i] = delta+y[i-1]+d[i]

plt.plot(y)
plt.plot(x)
plt.xlabel("Random Walk mit Drift")
plt.ylabel("y-Abweichung in [m]")

plt.show()
```

# Weisses Rauschen

- Eine Zeitreihe

$$\{x_1, x_2, \dots, x_n\}$$

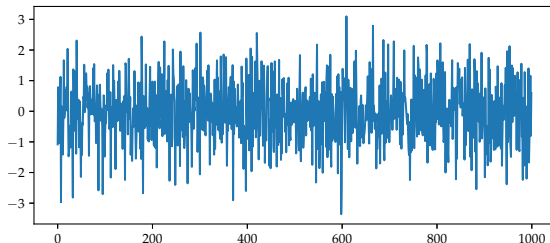
kann als *eine* Realisierung der multivariaten Zufallsvariablen

$$\{X_1, X_2, \dots, X_n\}$$

aufgefasst werden

- Modellierung und Vorhersagen für Zeitreihen kommt dementsprechend der Analyse der Daten von *einer* Beobachtung gleich, was ohne weitere Annahmen über die Zeitreihe unmöglich ist
- Beispiel, die ohne diese Annahmen auskommt und somit nicht vorhersehbar ist: *weisses Rauschen* (*white noise*)

- Plot:



- Prozess des weissen Rauschens: unabhängige, gleich verteilte Zufallsvariablen

$$\{W_1, W_2, \dots, W_n\}$$

- Alle  $W_i$  Mittelwert 0 und Varianz  $\sigma^2$  hat

- Code:

```
w = np.random.normal(size=1000)

plt.plot(w)

plt.show()
```

- Zufallsvariablen  $W_i$  zusätzlich normalverteilt → *Gauss'sches weissen Rauschen*
- Diese Modelle beschreiben das Rauschen bei Ingenieurproblemen
- Begriff *weiss*: Analogie zum weissen Licht
- Deutet an, dass alle möglichen periodischen Oszillationen in der Zeitreihe mit gleicher Stärke vorhanden sind
- Beobachtungen in einem Prozess des weissen Rauschens sind unkorreliert und können mit den gewöhnlichen statistischen Methoden

## Weisse Rauschen, dass seriell korreliert ist

- Wenden *sliding window filter* an auf den Prozess des weissen Rauschens:

$$\{W_1, W_2, \dots, W_n\}$$

- Erhalten einen *moving average*-Prozess
- Wählen insbesondere ein Fenster der Länge 3:

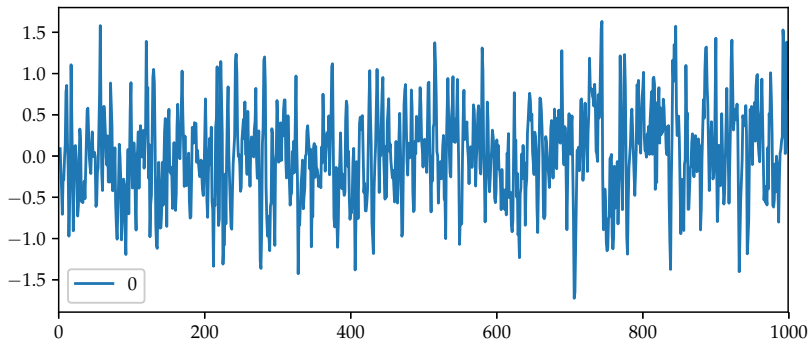
$$V_i = \frac{1}{3}(W_{i-1} + W_i + W_{i+1})$$

- Wählen

$$V_1 = W_1 \quad \text{und} \quad V_2 = 0.5(W_1 + W_2)$$

- Resultierender Prozess ist glatter  $\rightarrow$  Oszillationen höherer Ordnung werden ausgeglättet

## ● Plot:



## ● Code:

```
w = DataFrame(np.random.normal(size=1000))  
  
w.rolling(window=3).mean().plot()  
  
plt.show()
```



# Autoregressive Zeitreihen

- Viele Beispiele von Anwendungsproblemen, wie akkustische Zeitreihen in der Sprachanalyse, enthalten dominanten oszillierende Komponenten, die sinusförmiges Verhalten aufweisen
- Ein mögliches Beispiel um solche quasiperiodischen Daten zu erzeugen, sind autoregressive Zeitreihen

## Beispiel:

- Betrachten wieder einen Prozess des weissen Rauschens

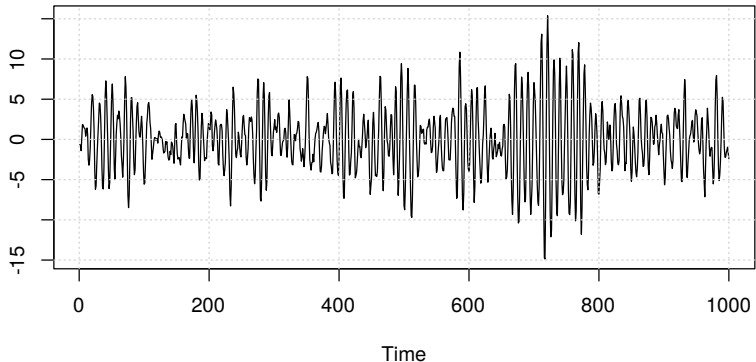
$$\{W_1, W_2, \dots, W_n\}$$

- Definieren dann rekursive die folgende Reihe

$$X_i = 1.5X_{i-1} - 0.9X_{i-2} + W_i$$

- Wert für Zeitpunkt  $i$  modelliert als Linearkombination der letzten beiden Werte addiert mit einer zufälligen Komponente
- So ein Prozess wird *autoregressiv* genannt
- Definition der Anfangsbedingungen sind subtil, da der ganze Prozess stark von diesen abhängt
- Werden vorläufig die Frage der Anfangsbedingungen ignorieren

- Plot:



- Realisierung des autoregressiven Prozesses oben
- Oszillierende Verhalten kommt deutlich zum Vorschein

- Beispiele oben: Motivierten den Gebrauch von verschiedenen Kombinationen von Zufallsvariablen zur Erzeugung von Zeitreihe mit dem wir Anwendungsprobleme nachahmen
- Wichtig: Statistisches Verhalten solcher Modelle verstehen, um deren Genauigkeit abzuschätzen
- In Definition eines diskreten stochastischen Prozess  $\{X_1, X_2, \dots\}$  wurde die Existenz einer Verteilungsfunktion  $F_i(x)$  für alle Beobachtungen  $X_i$  in diesem Prozess postuliert, also

$$P(X_i \leq x) = F_i(x)$$

- Kenntnis der einzelnen Verteilungen reicht aber nicht, um das serielle Verhalten eines Prozesses zu verstehen, da die Beobachtungen gegenseitig voneinander *abhängen*
- Bekannt: Vollständige probabilistische Struktur eines solchen Prozesses durch die *gemeinsame Verteilung aller endlichen Ansammlungen*  $\{X_{i_1}, \dots, X_{i_n}\}$  aller Beobachtungen gegeben ist
- Müssen als eine Funktionen  $F$  finden, so dass

$$P(X_{i_1} \leq x_1, \dots, X_{i_n} \leq x_n) = F(x_1, \dots, x_n)$$

für alle möglichen Indizes  $i_1, \dots, i_n$

- Praxis: Keine solche multivariate Verteilungen notwendig
- Meiste Information in diesen gemeinsamen Verteilungen kann durch Mittelwerte, Varianz und Kovarianz beschrieben werden

# Mass für die Unabhängigkeit

- Definieren zuerst die ersten und zweiten *Momente* um den ganzen Prozess zu analysieren
- Beginnen mit der Mittelwertsfolge:

## Mittelwertsfolge

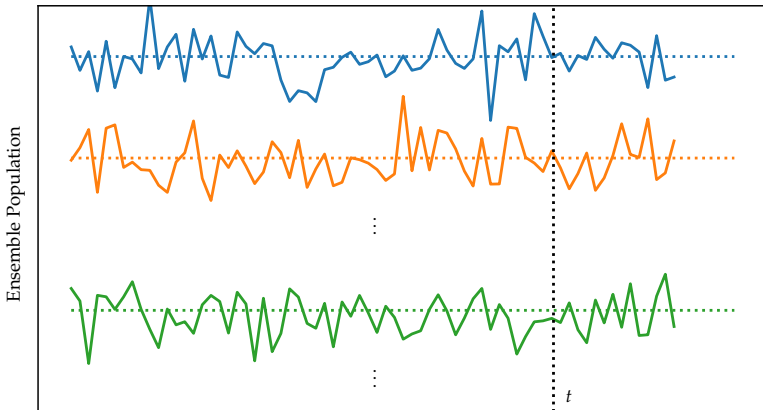
Die Mittelwertsfolge

$$\{\mu(1), \mu(2), \dots\}$$

(oder Mittelwertsfunktion) eines diskreten stochastischen Prozess  $\{X_1, X_2, \dots\}$  ist definiert durch die Folge der Mittelwerte:

$$\mu(i) = E[X_i]$$

● Skizze:



# Beispiele

- Berechnen Mittelwertsfolgen für einige Beispiele aus Abschnitt vorher
- Falls  $W_i$  ein Prozess des weissen Rauschens bezeichnet, dann

$$E[X_i] = 0 \quad \text{für alle } i \geq 1$$

- Nehmen Mittelwerte in diesem Prozess, so ändert sich folglich nichts am Mittelwert
- Mittelwertsfolgen in einem moving average Prozess ist 0



- Ist  $X_i$  ein Random Walk mit Drift, also  $X_0 = 0$ :

$$X_i = \delta + X_{i-1} + W_i$$

- Dann gilt:

$$E[X_1] = \delta + E[X_0] + E[W_1] = \delta$$

$$E[X_2] = \delta + E[X_1] + E[W_2] = 2\delta$$

$$E[X_3] = \delta + E[X_2] + E[W_3] = 3\delta$$

$\vdots$

- Das bedeutet, dass

$$\mu(i) = i\delta$$

# Repetition: Empirische Kovarianz und Korrelation

- Definition:

## Empirische Kovarianz und Korrelation

Für Stichproben  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  lautet die *empirische Kovarianz*:

$$\text{Cov}_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

Falls  $x = y$ , so gilt

$$\text{Cov}_{xx} = \frac{\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})}{n - 1}$$

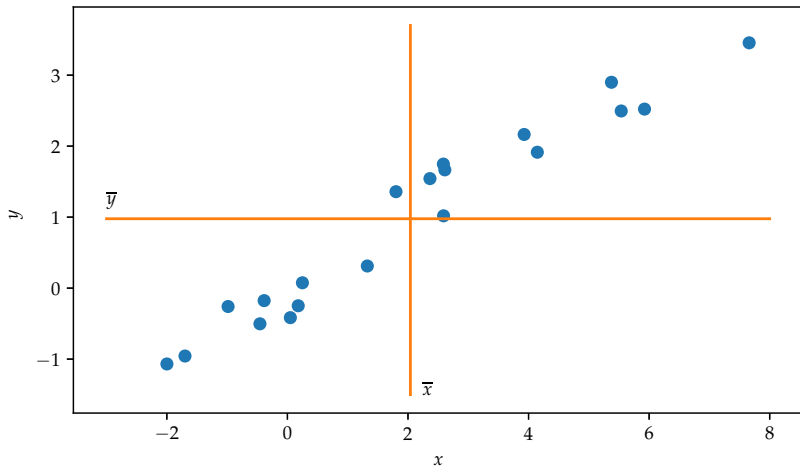
und dies ist gerade die *empirische Varianz* von  $x$ :

$$\text{Cov}_{xx} = \text{Var}_x = s_x^2$$

wobei  $s_x^2$  die empirische Varianz bezeichnet.

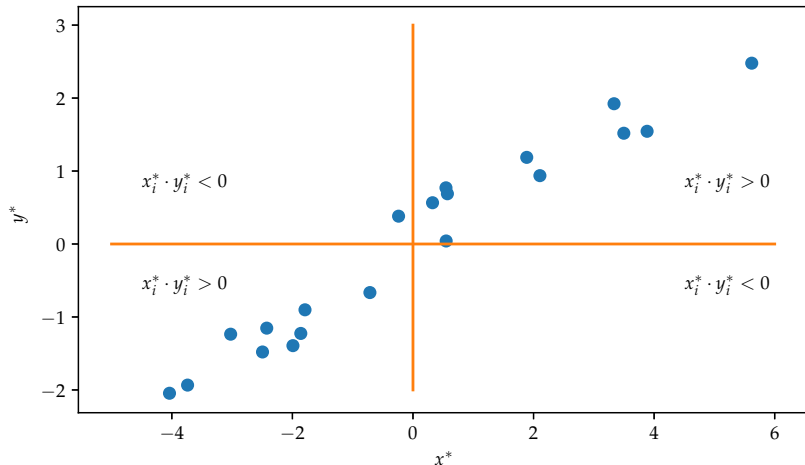
# Kovarianz und linearer Zusammenhang

- Beispiel: Punkte folgen mehr oder weniger Geraden



- Subtrahieren von den  $x$ -Koordinaten den Mittelwert  $\bar{x}$  und von den  $y$ -Koordinaten den Mittelwert  $\bar{y}$

- Abbildung:



- Empirische Kovarianz für diese Punkte lautet nun

$$\text{Cov}_{x^*y^*} = \frac{\sum_{i=1}^n x_i^* y_i^*}{n - 1}$$

- Im Zähler werden Produkte  $x_i^* y_i^*$  aufaddiert
- Im I. und III. Quadranten sind diese Produkte positiv, im II. und IV. Quadranten negativ
- Beispiel oben: Punkte praktisch alle im I. und III. Quadranten
- $\text{Cov}_{x^*y^*}$  sicher positiv:

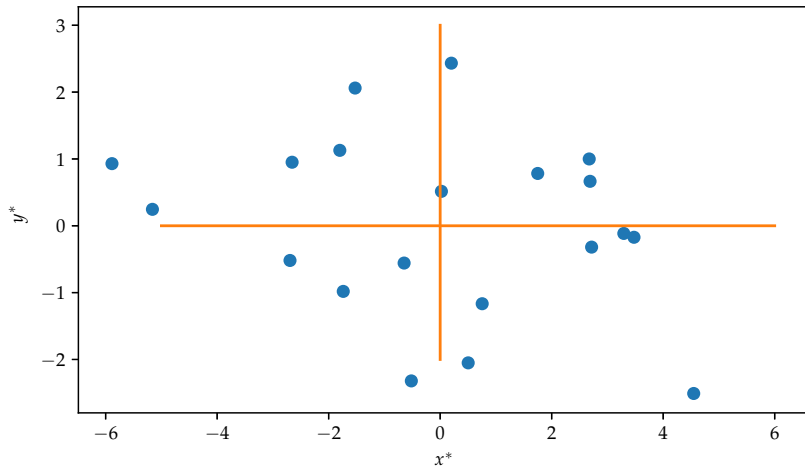
$$\text{Cov}_{x^*y^*} > 0$$

- Liegen die Punkte eher auf einer fallenden Geraden, so liegen die Punkte meistens im II. und IV. Quadranten
- Der Wert von  $\text{Cov}_{x^*y^*}$  wird dann sicher negativ:

$$\text{Cov}_{x^*y^*} < 0$$

# Kein linearer Zusammenhang

- Skizze:

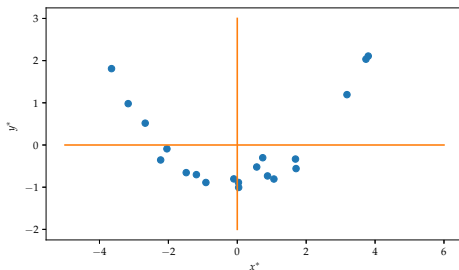


- Hälfte aller Punkte im I. und III. Quadranten (Produkte positiv)
- Andere Hälfte im II. und IV. Quadranten (Produkte negativ)
- Produkte betragsmässig ähnlich
- Produkte  $x_i^* y_i^*$  über alle Punkte aufaddiert heben sich in etwa auf

$$\text{Cov}_{x^* y^*} \approx 0$$

# Quadratischer Zusammenhang

- Abbildung:



- Beträge der Produkte links und rechts von der  $y$ -Achse heben sich auf
- Es gilt

$$\text{Cov}_{x^*y^*} \approx 0$$

- Kovarianz erkennt also nur *lineare* Zusammenhänge.



# Empirische Korrelation

- Definition:

## Empirische Korrelation

Die empirische Korrelation  $r$  für die Koordinatenpaare  $(x_i, y_i)$  ist wie folgt definiert:

$$r_{xy} = \frac{\text{Cov}_{xy}}{s_x \cdot s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1) \cdot s_x \cdot s_y}$$

wobei  $s_x$  und  $s_y$  die empirischen Standardabweichungen von den Stichproben  $x_i$  und  $y_i$  bezeichnen.

- Ist  $r_{xy} = 1$ , so liegen alle Punkte auf einer steigenden Geraden
- Für  $r_{xy} = -1$  liegen die Punkte alle auf einer fallenden Geraden.

# Autokovarianz und Autokorrelation

- Beginnen mit Kovarianz von Beobachtungen innerhalb eines *einzelnen* Prozesses

## Autokovarianz und Autokorrelation

Sei  $\{X_1, X_2, \dots\}$  ein diskreter stochastischer Prozess

- 1 Die *Autokovarianz*  $\gamma_X$  ist definiert durch

$$\gamma_X(i, j) = \text{Cov}(X_i, X_j) = E[(X_i - \mu(i))(X_j - \mu(j))]$$

- 2 Die *Autokorrelation*  $\rho_X$  ist definiert durch

$$\rho_X(i, j) = \frac{\gamma_X(i, j)}{\sqrt{\gamma_X(i, i)\gamma_X(j, j)}}$$

- Falls Kontext klar  $\rightarrow X$  weglassen
- Wichtige Eigenschaft für Autokovarianz und Autokorrelation: Symmetrie

$$\gamma(i, j) = \gamma(j, i)$$

- Autokovarianz misst *lineare Abhängigkeit* von zwei Punkten im selben Prozess beobachtet zu verschiedenen Zeitpunkten
- Falls die Zeitreihe sehr glatt  $\rightarrow$  Autokovarianz gross, auch wenn  $i$  und  $j$  weit auseinander liegen
- Beachte:

$$\gamma(i, j) = 0$$

Heisst nur, dass  $X_i$  und  $X_j$  nicht linear abhängig sind, sie können aber trotzdem nicht linear verknüpft sein

- Für  $i = j$  wird die Autokovarianz zur Varianz von  $X_i$

- Autokorrelation kann im gleichen Sinne beschrieben werden, aber normalisiert:

$$\rho(i, j) \in [-1, 1]$$

- Gibt es einen linearen Zusammenhang zwischen  $X_i$  und  $X_j$ , dann ist

$$\rho(X_i, X_j) = \pm 1$$

- Genauer: Falls

$$X_i = \beta_0 + \beta_1 X_j$$

dann ist die Autokorrelation 1 falls  $\beta_1 > 0$ , ansonsten  $-1$

- Autokorrelation: Grobes Mass an, wie die Reihe zur Zeit  $i$  durch den Wert der Reihe zur Zeit  $j$  vorhergesagt werden kann

# Beispiele: Autokovarianz und die Autokorrelation

- Prozess des weissen Rauschens hat Autokovarianzfunktion

$$\gamma(i, j) = \begin{cases} 0 & \text{falls } i \neq j \\ \sigma^2 & \text{falls } i = j \end{cases}$$

- Entsprechend ist die Autokorrelation 1 falls  $i = j$  und 0 sonst.

- Autokovarianz drei Punkte moving average Prozesses

- Aus der Definition der Autokovarianz ist klar, dass

$$\gamma(i, j) = \text{Cov}(X_i, X_j) = \text{Cov}\left(\frac{1}{3}(W_{i-1} + W_i + W_{i+1}), \frac{1}{3}(W_{j-1} + W_j + W_{j+1})\right)$$

- Falls  $i = j$ , dann

$$\begin{aligned}\text{Cov}(X_i, X_i) &= \frac{1}{9} \text{Cov}(W_{i-1} + W_i + W_{i+1}, W_{i-1} + W_i + W_{i+1}) \\ &= \frac{1}{9} (\text{Cov}(W_{i-1}, W_{i-1}) + \text{Cov}(W_i, W_i) + \text{Cov}(W_{i+1}, W_{i+1})) \\ &= \frac{3\sigma^2}{9}\end{aligned}$$

- Dies folgt aus der Tatsache, dass  $W_i$ ,  $W_{i-1}$  und  $W_{i+1}$  gegenseitig unkorreliert sind

- Analog für  $i + 1 = j$ :

$$\begin{aligned}\text{Cov}(X_i, X_{i+1}) &= \frac{1}{9} \text{Cov}(W_{i-1} + W_i + W_{i+1}, W_i + W_{i+1} + W_{i+2}) \\ &= \frac{1}{9} (\text{Cov}(W_i, W_i) + \text{Cov}(W_{i+1}, W_{i+1})) \\ &= \frac{2\sigma^2}{9}\end{aligned}$$

- Zusammenfassend

$$\gamma(i, j) = \begin{cases} \frac{3\sigma^2}{9} & \text{falls } i = j \\ \frac{2\sigma^2}{9} & \text{falls } |i - j| = 1 \\ \frac{\sigma^2}{9} & \text{falls } |i - j| = 2 \\ 0 & \text{sonst} \end{cases}$$

- Glättung des weissen Rauschens führt auf eine nichtriviale Autokovarianzstruktur
- Bemerkenswert: Autokovarianz hängt nur vom Abstand der Beobachtungen ab, aber nicht von deren Wert

- Autokorrelation:

$$\rho(i,j) = \frac{\gamma(i,j)}{\sqrt{\gamma(i,i)\gamma(j,j)}} = \frac{\gamma(i,j)}{\gamma(i,i)}$$

- Erhalten

$$\rho(i,j) = \begin{cases} 1 & \text{falls } i = j \\ \frac{2}{3} & \text{falls } |i - j| = 1 \\ \frac{1}{3} & \text{falls } |i - j| = 2 \\ 0 & \text{sonst} \end{cases}$$



# Graphische Darstellung

- Funktion `plot_acf`

```
import matplotlib.pyplot as plt
import pandas as pd
from pandas import DataFrame
from pandas import Series
import numpy as np

from statsmodels.graphics.tsaplots import plot_acf
from statsmodels.tsa.stattools import acf

w = DataFrame(np.random.normal(size=1000))

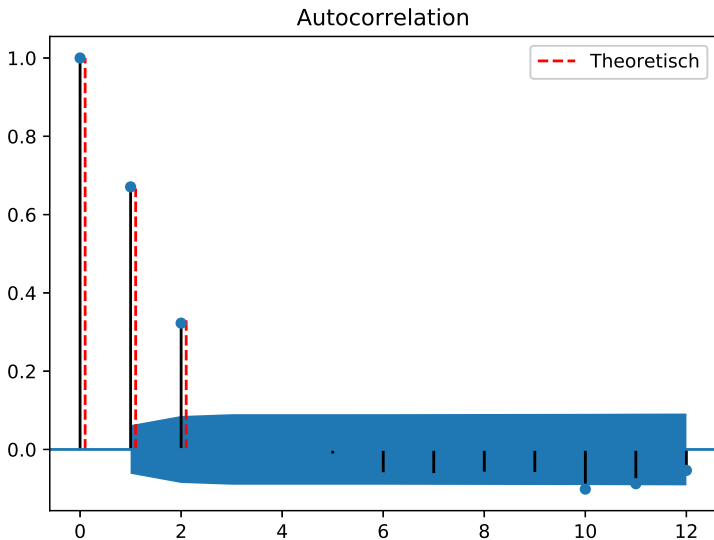
MA = DataFrame(w.rolling(window=3).mean()).dropna()

plot_acf(MA, lags=12, c="C1")

plt.vlines(x=2.1, ymin=0, ymax=1/3, color="red", linestyle='--', label=
plt.vlines(x=1.1, ymin=0, ymax=2/3, color="red", linestyle='--')
plt.vlines(x=0.1, ymin=0, ymax=1, color="red", linestyle='--')

plt.legend()
```

● Skizze:



- Autokovarianz des Random Walk
- Erinnerung: Random Walk Prozess  $X_i$  definiert als Summe von unabhängigen Bernoulli Zufallsvariablen

$$X_i = D_1 + \cdots + D_i$$

jede mit W'keit  $p = 0.5$

- Varianz für jedes  $D_i$  is

$$\sigma^2 = p(1 - p) = 0.25$$

- Damit:

$$\gamma(i, j) = \text{Cov} \left( \sum_{k=0}^i D_k, \sum_{l=0}^j D_l \right) = \min(i, j) \sigma^2$$

- Autokovarianz des Random Walks hängt nicht nur vom Unterschied der Beobachtungen, sondern auch von den Zeitpunkten  $i$  und  $j$

- Insbesondere ist die Varianz der Prozesses zur Zeit  $i$

$$\text{Var}(X_i) = i\sigma^2$$

und nimmt somit mit der Zeit zu

- Autokorrelationsfunktion des Random Walk einfach berechnen:

$$\rho(i, j) = \frac{\gamma(i, j)}{\sqrt{\gamma(i, i)\gamma(j, j)}} = \frac{\min(i, j)}{\sqrt{i \cdot j}}$$

# Stationarität

- Strikte Stationarität: Definition siehe Skript
- Für Anwendungen oft ungeeignet
- Schwächere Form der Stationarität

## Schwache Stationarität

Ein stochastischer Prozess  $X_i$  heisst *schwach stationär* falls

- 1 die Mittelwertsfolgen  $\mu_X(i)$  konstant ist und nicht vom Zeitindex  $i$  abhängt und
- 2 die Autokovarianzfolgen  $\gamma_X(i, j)$  hängt von  $i$  und  $j$  nur durch die Differenz  $|i - j|$  ab.

- Jede strikt stationäre Zeitreihe ist auch schwach stationär
- Die Umkehrung ist allgemein nicht wahr
- Man kann zeigen: Für Gauss'sche Prozesse (jede endliche Auswahl der Zufallsvariablen im Prozess hat gemeinsame Normalverteilung) dass dann die beiden Begriffe der Stationarität äquivalent sind
- Autokovarianz/-korrelation für (schwache) Stationarität hängt nur vom *Zeitunterschied (lag)*  $h = i - j$  abhängt
- Folgen als Funktionen von  $h$  selbst betrachten:

$$\gamma(h) = \gamma(i, i + h)$$

$$\rho(h) = \rho(i, i + h)$$

- Offensichtlich gilt

$$\gamma(h) = \gamma(-h)$$

so dass wir nur Werte  $h = 0, 1, \dots$  betrachten müssen.

# Beispiel

- Betrachten den drei Punkte moving average Prozesses
- Klar, dass die Mittelwertsfunktion

$$\mu(i) = \mu = 0$$

konstant ist

- Autokovarianz hängt nur vom Zeitunterschied ab:

$$\gamma(h) = \begin{cases} \frac{3\sigma^2}{9} & \text{falls } h = 0 \\ \frac{2\sigma^2}{9} & \text{falls } |h| = 1 \\ \frac{\sigma^2}{9} & \text{falls } |h| = 2 \\ 0 & \text{else.} \end{cases}$$

- Prozess des moving average schwach stationär