

Aufgabenblatt 5 – Lösungen

Aufgabe 1 (Zeitreihendekomposition)

- a) Was bedeutet Zeitreihendekomposition und warum interessieren wir uns dafür?

Zeitreihendekomposition meint die Zerlegung einer Zeitreihe in verschiedene Komponenten, zum Beispiel Trend, Saisonalität, Zyklen und Restkomponente gemäß dem klassischen Komponentenmodell. Die Zerlegung soll zu einem guten Verständnis der einzelnen Komponenten führen, mit dem Ziel, bessere Vorhersagen für die Gesamtzeitreihe zu treffen.

- b) Bitte beschreiben Sie, was die R Funktion `decompose()` tut und wenden Sie die Funktion auf den ‘Geburten in Deutschland’ Datensatz an (siehe Aufgabenblatt 4).

Gemäß der Hilfeseite `?decompose` zerlegt diese Funktion eine Zeitreihe in eine Trend-, Saison- und Restkomponente. Sie hat folgende Argumente:

- Das Argument `x` ist ein `ts` Object, also eine mit `ts()` erstellte Zeitreihe.
- Mit `type` kann zwischen einem additiven oder multiplikativen Modell gewählt werden (wir beschränken uns in der Vorlesung auf die additive Variante).
- Dem Argument `filter` kann ein numerischer Vektor mit Gewichten für ein heuristisches Glättungsverfahren übergeben werden.

Die Funktion führt die folgenden Schritte durch:

1. Bestimme den Trend durch heuristische Glättung. Die Fensterbreite entspricht der Frequenz (mit Modifikation im Falle einer geraden Frequenz).
2. Bestimme eine fixe Saisonkomponente als Durchschnittswert in den Saisons, nachdem der Trend abgezogen wurde.
3. Die Restkomponente ist dann einfach noch das, was von der Zeitreihe übrig bleibt, nachdem Trend und Saisonkomponente abgezogen wurden.

Siehe R Code für die Anwendung auf den ‘Geburten in Deutschland’ Datensatz.

- c) Implementieren Sie selbst eine R Funktion, mit der die Schritte aus b) durchgeführt werden können und vergleichen Sie das Ergebnis Ihrer Funktion mit `decompose()`.

Siehe R Code.

Aufgabe 2 (Stationäre Prozesse)

a) Was ist ein stationärer Prozess und warum interessieren wir uns dafür?

Ein stationärer (stochastischer) Prozess ist eine Folge Zufallsvariablen $(y_t)_{t \in T}$ mit

1. konstantem Erwartungswert über die Zeit,
2. konstanter Varianz über die Zeit,
3. und Kovarianz zwischen y_t und y_{t+k} , die nur von k und nicht von t abhängt.

Wir benutzen stochastische Prozesse für die Modellierung von Zyklen. *Stationäre* Prozesse bieten ‘Stabilität’, damit können wir etwas über die Zyklen lernen.

b) Im Folgenden bezeichnet $\epsilon_t \sim WN(0, \sigma^2)$ unabhängiges weißes Rauschen, $t \in \mathbb{N}$ ist wie gewohnt der Zeitindex und $a, b \in \mathbb{R}$ sind beliebige aber feste Konstanten. Welche der folgenden Prozesse sind stationär?

- $y_t = a\epsilon_t + b$,
- $y_t = \epsilon_t + b\epsilon_{t-2}$,
- $y_t = t + b\epsilon_0$,
- $y_t = ay_{t-1} + \epsilon_t$, wobei hier $|a| < 1$ und $y_0 = 0$ gilt.

Wir prüfen die drei unter a) aufgeführten Eigenschaften für die vier Prozesse:

1. Der erste Prozess ist stationär: Erwartungswert und Varianz sind b bzw. $a^2\sigma^2$ (also zeitunabhängig), die Kovarianz zwischen y_t und y_{t+k} ist 0 (für $k \geq 1$).
2. Der zweite Prozess ist auch stationär: $\mathbb{E}(y_t) = 0$, $\text{Var}(y_t) = (1 + b^2)\sigma^2$, $\text{Cov}(y_t, y_{t+1}) = 0$, $\text{Cov}(y_t, y_{t+2}) = b\sigma^2$, $\text{Cov}(y_t, y_{t+k}) = 0$ für $k > 2$.
3. Der dritte Prozess ist *nicht* stationär, denn $\mathbb{E}(y_t) = t$ (zeitabhängig).
4. Der vierte Prozess ist ebenfalls *nicht* stationär. Wir können $y_t = ay_{t-1} + \epsilon_t$ als $y_t = \sum_{i=0}^t a^i \epsilon_{t-i} + a^t y_0$ umschreiben und sehen, dass $\text{Var}(y_t) = \sum_{i=0}^t a^{2i} \sigma^2$ (zeitabhängig). Beachte aber, dass der Prozess stationär *wäre*, wenn y_0 nicht deterministisch, sondern eine (spezielle) Zufallsvariable ist (siehe Vorlesung).

c) Bitte simulieren Sie jeweils eine Trajektorie der Prozesse aus b), wobei Sie normalverteiltes weißes Rauschen, $a = -0.5$ und $b = 0.1$ verwenden.

Siehe R Code.

Aufgabe 3 (Korrelogramme)

- a) Was ist ein Korrelogramm und warum interessieren wir uns dafür?

Ein Korrelogramm ist die Visualisierung der Funktion $k \rightarrow \text{Cov}(y_t, y_{t+k})/\text{Var}(y_t)$, die den ‘Lag’ k auf die Autokorrelationskoeffizienten abbildet. Damit kann ein stationärer Prozess identifiziert werden. Für eine empirische Zeitreihe werden die Autokorrelationskoeffizienten geschätzt (siehe Vorlesung), das ergibt ein *empirisches* Korrelogramm.

- b) Bitte erstellen Sie empirische Korrelogramme für die in 2c) simulierten Trajektorien. Welche Besonderheiten fallen auf?

Siehe R Code für die empirischen Korrelogramme. Geplottet werden die empirisch geschätzten Autokorrelationskoeffizienten beginnend ab ‘Lag’ 0 (dieser Wert ist immer 1). Zusätzlich sehen wir blau gestrichelte Linien, die (standardmäßig) ein 95% Konfidenzintervall für die geschätzten Autokorrelationen um die 0 liefern (mehr Details hierzu gibt es bald in der Vorlesung).

- Der erste Prozess weist keine signifikanten Autokorrelationen auf.
- Gleiches gilt für den zweiten Prozess, aber hier wird eine signifikante Autokorrelation im zweiten Lag deutlich, wenn T oder b erhöht werden.
- Der dritte Prozess hat eine deutliche Autokorrelation über die Zeit.
- Der vierte Prozess hat eine oszillierende (da a negativ ist) Autokorrelation, die geometrisch abklingend ist. Das wird deutlicher, wenn a (betragsmäßig) oder T größer gewählt werden.

- c) Häufig betrachtet man auch die *partielle* Autokorrelationsfunktion. Wo liegt der Unterschied und wie kann sie geschätzt werden?

Die *partielle* Autokorrelationsfunktion visualisiert die Autokorrelationen, die sich ergeben, wenn die Autokorrelationen früherer ‘Lags’ herausgerechnet werden. Damit bieten sie einen besseren Überblick. Beachte, dass die Autokorrelation und die partielle Autokorrelation für den ‘Lag’ 1 identisch sind. Die partiellen Autokorrelationen können per Regression geschätzt werden (siehe Vorlesung) oder mit dem sogenannten Yule-Walker Schätzer.