

Lineare Vorhersagen für Zeitreihen

Philipp [REDACTED] - Matrikel-Nr.: [REDACTED]



Universität [REDACTED]

Fakultät für Mathematik, Informatik und Naturwissenschaften

Fachbereich Mathematik

Abschlussarbeit zur Erlangung des Grades Bachelor of Science

Erstgutachterin: Prof. Dr. [REDACTED]

Zweitgutachter: Prof. Dr. [REDACTED]

Abgabedatum: [REDACTED] 2024

Inhaltsverzeichnis

1	Einleitung und Beispiele	1
2	Komponentenzerlegung	2
2.1	Bestimmung Trendkomponente	5
2.2	Bestimmung Saisonkomponente	6
3	Lineare Vorhersagen	6
3.1	Allgemeine Lösung	8
3.2	Stationäre Lösung	12
3.3	1-Schritt-Vorhersage	13
3.4	h-Schritt-Vorhersage	21
4	Schätzung aus Beobachtung	22
4.1	Autokovarianzfunktion	22
4.2	Verteilung	26
	Literaturverzeichnis	28
	Datenverzeichnis	28

1 Einleitung und Beispiele

Lit.: [3], [5], [6]. In vielen Bereichen des Lebens ist die Erfassung und Analyse bestimmter Daten von essentieller Bedeutung. So ermittelt bspw. das [Statistische Bundesamt](#) [9] monatlich die Anzahl der Verkehrstoten in Deutschland. Die [Frankfurter Börse](#) [8] protokolliert täglich den Endstand des Deutschen Aktienindex, während das [Umweltbundesamt](#) [10] jährlich die Treibhausgas-Emissionen in Deutschland überwacht.

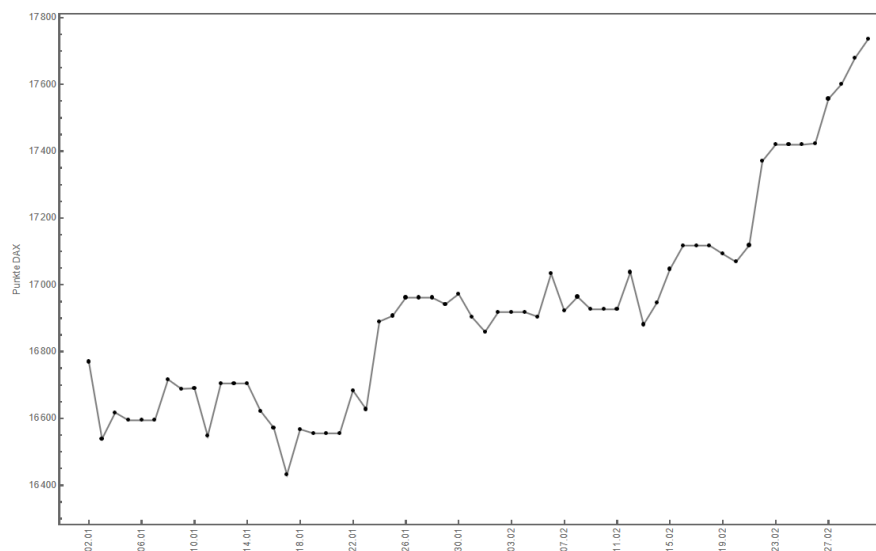


Abbildung 1: Täglicher Endstand des DAX zwischen 02.01.24 und 01.03.24

Wann immer eine messbare Größe in festen zeitlichen Abständen ermittelt wird, spricht man beim erhaltenen Datensatz von einer *Zeitreihe*. Besonders interessant bei der Analyse von Zeitreihen ist deren Vorhersage (vgl. Abbildung 2). Das heißt die Frage nach der zukünftigen Entwicklung der entsprechenden Größe, basierend auf einem Satz bereits gemachter Messungen (sog. *Beobachtung*).

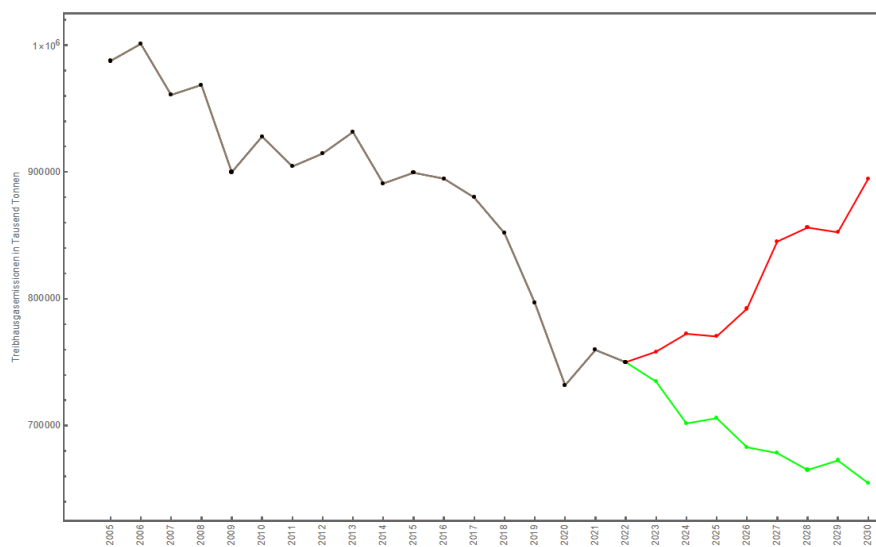


Abbildung 2: Jährliche Treibhausgas-Emissionen seit 2005 mit möglichen Entwicklungen

Die vorliegende Arbeit zielt darauf ab, mathematische Methoden zu beschreiben, die es ermöglichen, auf der Grundlage einer gemachten Beobachtung Zeitreihen effizient

vorherzusagen. Im ersten Schritt machen wir uns dazu klar, welche deterministischen Komponenten Zeitreihen enthalten können und wie sich diese geeignet entfernen lassen. Dies bildet die Grundlage, um uns anschließend auf die Vorhersage sogenannter *stationärer* Zeitreihen zu fokussieren. Hierbei setzen wir zunächst Methoden der Approximation ein, um eine allgemeine Lösung des *linearen* Vorhersageproblems zu erarbeiten. Anschließend werden wir uns darum kümmern, ein effizientes Berechnungsverfahren der gefundenen Lösung zu entwickeln, was den wesentlichen Bestandteil dieser Arbeit darstellt. Schlussendlich betrachten wir noch geeignete Schätzer, um das beschriebene Verfahren nur auf einer gemachten Beobachtung der Zeitreihe basieren zu lassen sowie die Möglichkeit zu haben, über die Sicherheit der gemachten Beobachtung auszusagen.

2 Komponentenerlegung

Lit.: [3],[5],[6]. Aus den eingangs gezeigten Beispielen wird deutlich: Zeitreihen sind *regelmäßig indizierte Daten einer bestimmten Größe*. Regelmäßigkeit können wir durch Indizierung mit den ganzen Zahlen $\mathbb{Z}$, die Daten durch Zufallsvariablen X_t , und die Größe selbst durch ihre Wertemenge S beschreiben. Formell also wie folgt.

Definition 2.1 (Zeitreihe) Eine Zeitreihe $X := X_T$ ist eine Familie von Zufallsvariablen $X_t \in S$ mit $t \in T = \mathbb{Z}$ oder $T = \mathbb{N}$ auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$.

- Dabei ist $X_T := (X_t : t \in T)$, $(S, \mathcal{S})$ ein Messraum, $X_t \in S$ meint $X_t : \Omega \rightarrow S$, und wir vereinbaren die Notation $X \subset S$ für $X \in S^T := \{X : T \rightarrow S\}$.

Für reale Zeitreihen gilt zumeist $T = \mathbb{N} := \{1, 2, \dots\}$ und $S = \mathbb{R}$; kurz also $X_{\mathbb{N}} \subset \mathbb{R}$. Betrachten wir nun die Zeitreihe der Verkehrstoten in Deutschland pro Monat, um ein Gefühl dafür zu bekommen, welche deterministischen Komponenten in Zeitreihen zu erwarten sind. Interessant, aber nicht weiter relevant, an diesen Daten ist, dass man deutlich die Corona-Zeit (2020-2022) als Rückgang der Verkehrstoten erkennen kann.

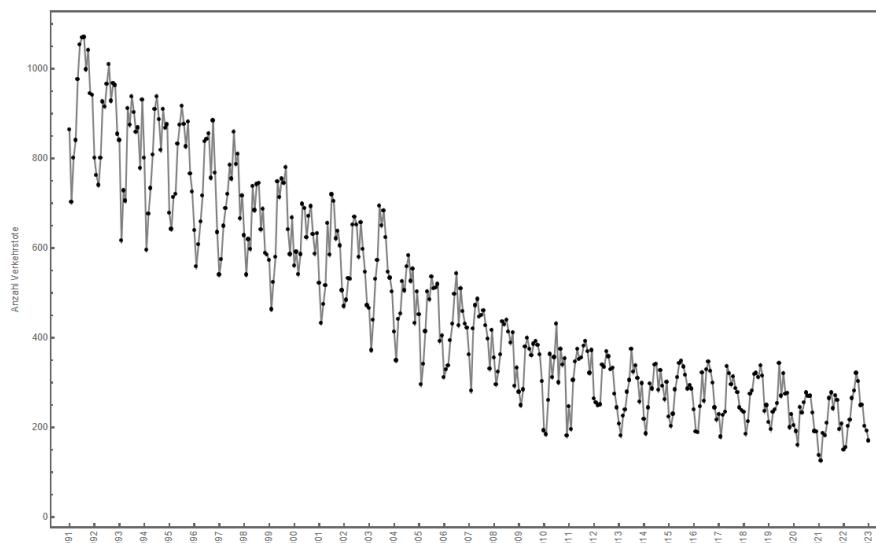


Abbildung 3: Fortwährend sinkende Anzahl Verkehrstoter

Deutlich zu erkennen ist ein gewisser, deterministisch zu begründender, abfallender Trend. Nicht nur sind Fahrzeuge in den letzten Jahrzehnten immer sicherer geworden, auch könnte man davon ausgehen, dass sich bei den Verkehrsteilnehmern ein höheres Risikobewusstsein und damit verbunden eine größere Vorsicht entwickelt haben. Entfernt man diesen Trend, wie im nächsten Abschnitt gezeigt, verbleibt folgende Zeitreihe.

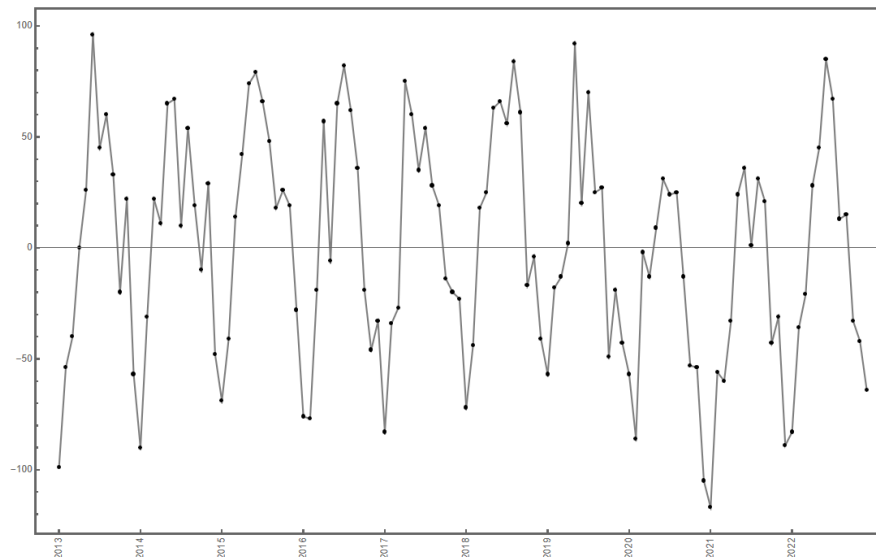


Abbildung 4: Zeitreihe der Verkehrstoten ab 2013 (trendbereinigt)

Trotz Entfernen des abfallenden Trends bleibt eine gewisse periodische Saisonalität in den Daten erkennbar. In den Wintermonaten ist die Anzahl der Verkehrstoten auf einem geringeren Niveau als in den Sommermonaten. Auch dies lässt sich wieder deterministisch begründen. Schlechte Witterungsverhältnisse in den Wintermonaten führen zu schlechter Sicht und rutschigen Straßen, wodurch man davon ausgehen könnte, dass während dieser Zeit die Verkehrsteilnehmer vorsichtiger fahren, was wiederum zu weniger Unfällen führt.

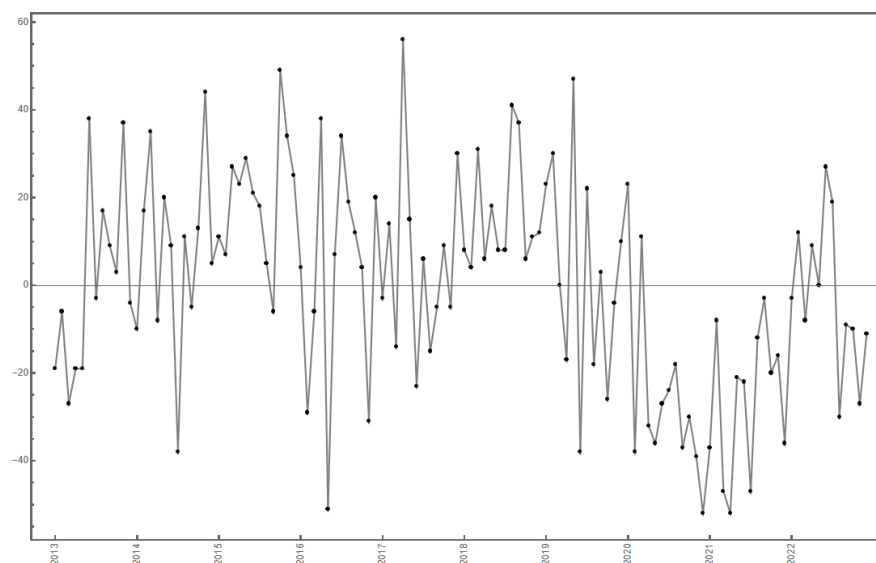


Abbildung 5: Zeitreihe der Verkehrstoten ab 2013 (trend- und saisonbereinigt)

Entfernt man nun auch jene Saisonalität, wie im übernächsten Abschnitt gezeigt, verbleibt eine Zeitreihe (vgl. Abbildung 5), die nur noch aus den zufälligen Schwingungen der betrachteten Daten besteht (*Stationarität*). Dank diesen Überlegungen liegt es also nahe, einer Zeitreihe X drei Komponenten zuzuschreiben. Einen deterministischen Trend m , eine deterministische Saison s , und eine Zufallskomponente Z . Wir halten dies in folgender Definition fest.

Definition 2.2 (Komponentenzerlegung) Sei $X \subset S$ eine Zeitreihe und seien $m, s \subset S$ deterministisch. Dann nennen wir die Darstellung $X = Z + m + s$ Komponentenzerlegung von X .

- Dabei sind $Z := Z_T$, $m := m_T$, und $s := s_T$ für $X = X_T$ sowie die Addition komponentenweise zwischen abzählbar indizierten Familien zu verstehen.
- Weiter soll Z zentriert sein, das heißt $E(Z_t) = 0$ für alle $t \in T$, s periodisch, das heißt es gibt ein $d \in \mathbb{N}$ sodass $s_t = s_{t+d}$ für alle $t \in T$, und m nicht periodisch.
- Z heißt Zufalls-, m heißt (det.) Trend-, und s heißt (det.) Saisonkomponente.

Abschließend in diesem Kapitel wollen wir noch klären, was genau mit 'stationär' gemeint ist. Dazu orientieren wir uns wieder am Beispiel der Verkehrstoten. Betrachtet man Abbildung 5, so erkennt man, dass diese Zeitreihe um ein konstantes Niveau μ herum schwingt (hierbei ist $\mu = 0$). Formell können wir dies durch einen konstanten Erwartungswert aller beteiligten Zufallsvariablen $E(Z_t) = \mu$ formulieren.

Ferner fällt auf, dass benachbarte Datenpunkte ungefähr auf gleicher Höhe liegen und diese Wirkung mit steigendem zeitlichen Abstand nachlässt. Insbesondere ist dabei nicht relevant, welchen Datenpunkt wir initial betrachtet haben. Formell können wir dies durch eine nur vom zeitlichen Abstand zweier Zufallsvariablen abhängige Kovarianz $\text{Cov}(Z_t, Z_{t+h}) = \text{Cov}(Z_0, Z_h)$ beschreiben.

Diese Punkte folgen auch aus folgender Überlegung. Ist eine Beobachtung 'stationär', so gehen wir davon aus, dass diese repräsentativ für die Verteilung der Zeitreihe ist. Formell können wir dies durch Translations-Invarianz $Z_I \stackrel{D}{=} Z_{I+h}$ aller endlichen gemeinsamen Verteilungen beschreiben. Wir definieren also zwei Begriffe wie folgt.

Definition 2.3 (Stationarität) Eine Zeitreihe Z mit $\text{Var}(Z_t) < \infty$ für alle $t \in \mathbb{Z}$ heißt stationär, falls sie einem der folgenden beiden Punkte genügt.

- Z heißt stark stationär, falls für alle endlichen $I \subset \mathbb{Z}$ sowie $h \in \mathbb{Z}$ gilt:

$$Z_I \stackrel{D}{=} Z_{I+h}$$

- Z heißt schwach stationär, falls mit $\mu := E(Z_0)$ sowie $\gamma(h) := \text{Cov}(Z_0, Z_h)$ für $h \in \mathbb{Z}$ gilt:

$$E(Z_t) = \mu \quad \text{und} \quad \text{Cov}(Z_t, Z_{t+h}) = \gamma(h)$$

Die Funktion γ heißt Autokovarianzfunktion von Z .

Man überlegt sich leicht, dass starke Stationarität einer Zeitreihe sofort auch schwache Stationarität jener Zeitreihe impliziert, jedoch im Allgemeinen nicht umgekehrt.

Sei Z eine stark stationäre Zeitreihe. Dann gilt insbesondere

$$Z_t \stackrel{\mathcal{D}}{=} Z_0 \quad \text{und} \quad (Z_t, Z_{t+h}) \stackrel{\mathcal{D}}{=} (Z_0, Z_h)$$

für alle $t, h \in \mathbb{Z}$. Da Erwartungswert und Kovarianz aber nur von der Verteilung der entsprechenden Zufallsvariablen abhängen, folgt also sofort:

$$E(Z_t) = \mu \quad \text{und} \quad \text{Cov}(Z_t, Z_{t+h}) = \gamma(h)$$

Also ist Z auch schwach stationär.

2.1 Bestimmung Trendkomponente

Lit.: [-]. In diesem Abschnitt stellen wir eine Möglichkeit vor, die Trendkomponente m einer Zeitreihe $X = Z + m + s$ zu bestimmen. Sei dazu eine Beobachtung X_B mit Indexmenge $B = \{b_1, \dots, b_n\}$ mit $b_k \leq b_{k+1}$ gegeben. Die wesentliche Idee ist es, alle Steigungen zwischen benachbarten Punkten wie folgt geeignet gewichtet zu mitteln

$$\Delta := \frac{1}{n-1} \sum_{k=1}^{n-1} \lambda_k \left(\frac{X_{b_{k+1}} - X_{b_k}}{b_{k+1} - b_k} \right)$$

und darauf basierend eine Trendgerade mit dieser Steigung zu erstellen. Um eine Gerade mit vorgegebener Steigung eindeutig festzulegen, benötigt man einen Punkt, durch den diese verlaufen soll. Hierfür eignet sich das Mittel ξ der letzten $r \in \mathbb{N}$ Beobachtungen. Mit den Gewichten $\lambda_k := q^{n-k-1}$ für ein $q \in (0, 1)$ berücksichtigt man dabei aktuelle Werte stärker als weiter zurückliegende.

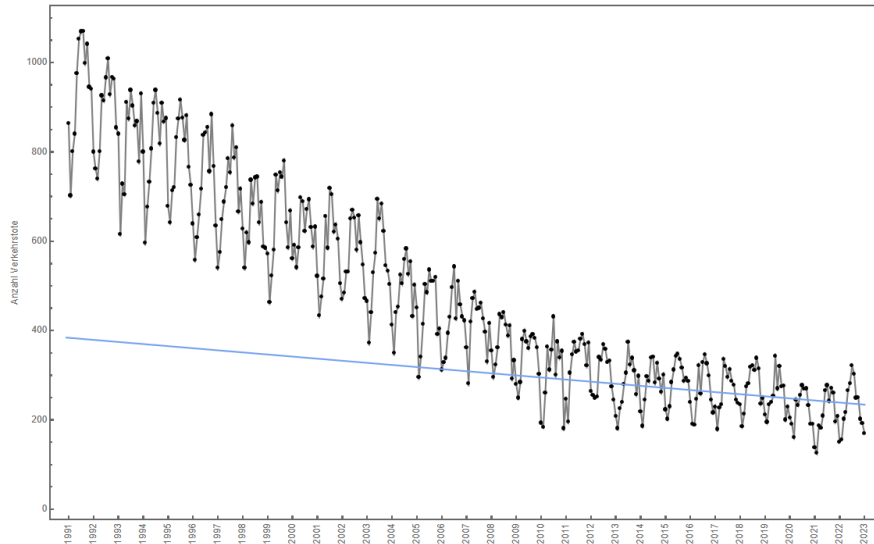


Abbildung 6: Anzahl Verkehrstoter (schwarz) mit abfallendem Trend (blau)

Deutlich zu erkennen ist, dass diese Methode bei sich ändernden Trends (hier stärkerer Abfall bis 2010 und schwächerer Abfall ab 2010) lediglich bei aktuellen Werten eine gute Beschreibung des Trends liefert. Dafür lässt sich der Trend $m = m_T$ mit

$$m_t := \xi + \Delta \cdot (t - n)$$

direkt und ohne großen Rechenaufwand aus der Beobachtung X_B bestimmen. Der konkrete Trend in Abbildung 6 wurde mit den Parametern $q = 99/100$ und $r = 12$ bestimmt und liefert ab 2010 wie gewünscht trendbereinigte Daten (vgl. Abbildung 4).

2.2 Bestimmung Saisonkomponente

Lit.: [5]. In diesem Abschnitt stellen wir eine Möglichkeit vor, die Saisonkomponente s einer Zeitreihe $X = Z + m + s$ zu bestimmen. Sei dazu eine trendbereinigte Beobachtung X_B mit Indexmenge $B = \{b_1, \dots, b_n\}$ mit $b_k \leq b_{k+1}$ gegeben. Die wesentliche Idee ist es, alle Werte im Abstand einer Periode d mit $s_t = s_{t+d}$ zu mitteln

$$s_t := \frac{1}{\chi(t)} \sum_{k \in \mathbb{Z}} X_{t+kd} I\{t + kd \in B\} \quad \text{mit} \quad \chi(t) := |(t + d\mathbb{Z}) \cap B|$$

wobei $s = s_T$ gilt. Im Beispiel der Verkehrstoten eignet sich eine Periode von $d = 12$ Monaten, also genau einem Jahr. Die Saison dieser Zeitreihe hat damit folgende Gestalt.

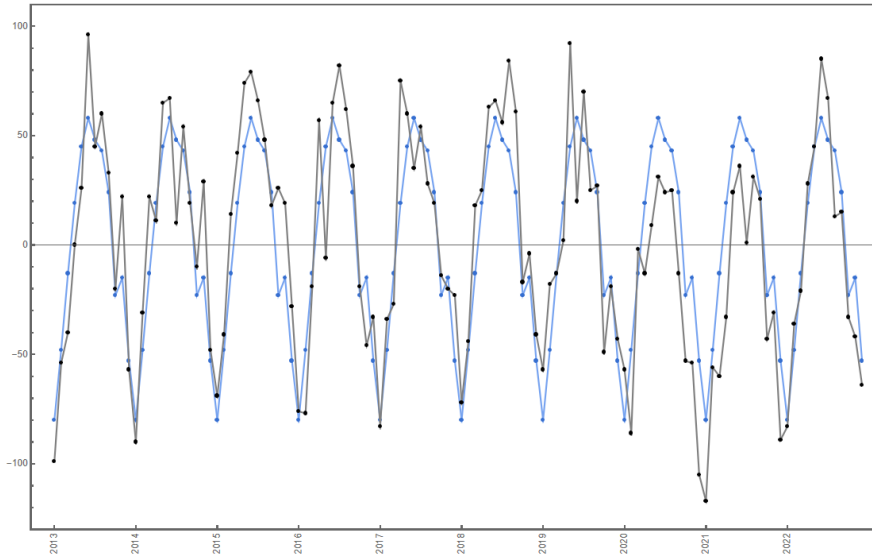


Abbildung 7: Zeitreihe der Verkehrstoten ab 2013 (trendbereinigt, schwarz) mit Saison (blau)

Die Differenzen der schwarzen und blauen Daten entsprechen genau den zufälligen Abweichungen vom gewöhnlichen Saisonwert und ergeben Abbildung 5.

3 Lineare Vorhersagen

Lit.: [1], [4], [5]. Nun widmen wir uns dem wesentlichen Teil dieser Arbeit, der Vorhersage von Zeitreihen. Zu einer gegebenen Beobachtung Z_B mit $B = \{b_1, \dots, b_n\} \subset \mathbb{Z}$ einer stationären Zeitreihe Z soll die beste lineare Vorhersage $\hat{Z}_h$ zu einem Zeitpunkt $h \in \mathbb{Z} \setminus B$ gemacht werden. Hierbei bedeutet 'linear', dass wir unsere Vorhersage auf die Form einer Linearkombination der gemachten Beobachtung einschränken, das heißt

$$\hat{Z}_h = \phi_{b_1} Z_{b_1} + \dots + \phi_{b_n} Z_{b_n}$$

für Koeffizienten $\phi_b \in \mathbb{R}$ mit $b \in B$. Wie bereits eingangs erwähnt, wollen wir Methoden der Approximation verwenden, um dieses Problem zu lösen und um gleichzeitig zu klären, was genau mit 'bester' Vorhersage gemeint ist.

Lemma 3.1 Sei $Z \subset \mathbb{R}$ eine (schwach) stationäre und zentrierte Zeitreihe. Dann gilt bereits $Z \subset L_2 := L_2(\Omega, \mathcal{A}, P)$. Wobei $(\Omega, \mathcal{A}, P)$ der zu Z gehörende Wahrscheinlichkeitsraum ist.

┌ Es reicht zu zeigen, dass $\|Z_t\|_2 < \infty$ gilt. Betrachte daher:

$$\begin{aligned}\|Z_t\|_2^2 &= \int Z_t^2 dP = E(Z_t^2) \\ &= \text{Var}(Z_t) - E(Z_t)^2 \stackrel{*}{=} \text{Var}(Z_t)\end{aligned}$$

Bei $\star$ ging ein, dass Z zentriert, also $E(Z_t) = 0$, ist. Ferner gilt $\text{Var}(Z_t) < \infty$ per Definition von Stationarität. Insgesamt also in der Tat $\|Z_t\|_2 < \infty$. ┐

Mit dem bekannten Schätzer $\hat{\mu} = \sum_b Z_b/n$ für $\mu = E(Z_0)$ können wir jede Zeitreihe Z zentrieren. Daher gehen wir im Folgenden immer von $\mu = 0$ aus. Der Raum L_2 ist ein euklidischer *Hilbertraum* (vollständiger Vektorraum mit Skalarprodukt), wobei das *Skalarprodukt* hierbei definiert ist als:

$$\langle X, Y \rangle := \int XY dP \quad \text{für } X, Y \in L_2$$

Man sieht sofort, dass die 2-Norm die von $\langle -, - \rangle$ induzierte Norm $\|X\|_2 = \langle X, X \rangle^{1/2}$ ist. Es gilt: Gegeben einem Hilbertraum F und einer abgeschlossenen und konvexen Teilmenge $S \subset F$, existiert zu jedem $f \in F$ eine eindeutige Bestapproximation $s^* \in S$ an f . Diese Aussage soll uns als roter Faden dienen.

Lemma 3.2 Sei F ein normierter Vektorraum und $S \subset F$ ein endlich-dimensionaler Untervektorraum. Dann ist S bereits abgeschlossen und konvex.

┌ Sei $(b_1, \dots, b_n)$ eine Basis von S . Dann ist S abgeschlossen, denn für $f \in F$ und $s_m \in S$ mit $s_m \rightarrow f$ ist auch der Grenzwert f bereits in S enthalten:

$$\begin{aligned}f &= \lim_{m \rightarrow \infty} s_m = \lim_{m \rightarrow \infty} \sum_{k \leq n} \alpha_{m,k} b_k \\ &= \sum_{k \leq n} \underbrace{\lim_{m \rightarrow \infty} \alpha_{m,k}}_{=: \alpha_k} b_k = \sum_{k \leq n} \alpha_k b_k \in S\end{aligned}$$

Wobei $\alpha_{m,k} \rightarrow \alpha_k$, wegen obiger Darstellung, durch $s_m \rightarrow f$ garantiert ist. Die Konvexität von S folgt mit $s, l \in S$ sowie $\lambda \in (0, 1)$ durch folgende Betrachtung:

$$\begin{aligned}\lambda s + (1 - \lambda)l &= \lambda \sum_{k \leq n} s_k b_k + (1 - \lambda) \sum_{k \leq n} l_k b_k \\ &= \sum_{k \leq n} (\lambda s_k + (1 - \lambda)l_k) b_k \in S\end{aligned}$$

Mit Hilfe dieser Lemmata können wir unsere Situation nun im Sinne der Approximation auffassen. Wähle dazu $F = L_2$ als den Hilbertraum und $S = \text{span}(Z_B)$ als endlich-dimensionalen Untervektorraum davon. Dann gibt es (wie wir noch sehen werden) zu jedem $h \in \mathbb{Z} \setminus B$ eine eindeutige Bestapproximation $s^* = \hat{Z}_h$ aus S an $f = Z_h$ aus F . Ferner ist damit auch geklärt, was wir mit 'beste' lineare Vorhersage meinen.

Definition 3.3 (Vorhersageproblem) Das lineare Vorhersageproblem $[Z_B|h]$ besteht darin, die Bestapproximation $\hat{Z}_h \in \text{span}(Z_B)$ an $Z_h \in L_2$ zu finden, das heißt:

$$\|\hat{Z}_h - Z_h\|_2 \leq \|\hat{Z}'_h - Z_h\|_2 \quad \forall \hat{Z}'_h \in \text{span}(Z_B)$$

Bemerkenswert ist dabei, dass, obwohl wir typischerweise einen Wert in der Zukunft vorhersagen wollen, beispielsweise $[Z_{[n]}|n+1]$ mit $[n] := \{1, \dots, n\}$, dies dank dieser Formulierung gar keine Rolle spielt. Statt über zeitliche Abstände zwischen den Werten denken wir einfach über abstrakte räumliche Abstände zwischen diesen nach. Insbesondere können wir so auch problemlos Werte in der Vergangenheit 'vorhersagen', beispielsweise $[Z_{[n]}|0]$.

3.1 Allgemeine Lösung

Lit.: [1], [4], [5]. Nachdem wir nun geklärt haben, was genau unter der besten linearen Vorhersage zu verstehen ist, soll es in diesem Abschnitt darum gehen, jene Lösung $\hat{Z}_h$ von $[Z_B|h]$ zu bestimmen. Dazu machen wir zunächst wieder einen Schritt zurück und betrachten die allgemeine Situation mit F und S . Hierbei leiten wir eine wesentliche Aussage über Minimalargumente konvexer Funktionale her, welche wir immer weiter konkretisieren und schlussendlich auf unsere exakte Situation anwenden werden.

Definition 3.4 (Gâteaux-Ableitung) Sei F ein normierter Vektorraum und $\varphi : F \rightarrow \mathbb{R}$ ein Funktional. Dann nennen wir folgenden Ausdruck, falls existent, die Gâteaux-Ableitung von φ bei x in Richtung v mit $x, v \in F$.

$$\partial\varphi(x; v) := \lim_{h \searrow 0} \frac{\varphi(x + hv) - \varphi(x)}{h}$$

Definition 3.5 (konvexes Funktional) Sei F ein normierter Raum und $\varphi : F \rightarrow \mathbb{R}$ ein Funktional. Dann nennen wir φ konvex, falls für alle $x, y \in F$ und $\lambda \in (0, 1)$ gilt:

$$\varphi(\lambda x + (1 - \lambda)y) \leq \lambda\varphi(x) + (1 - \lambda)\varphi(y)$$

Der eingeführte Ableitungsbegriff $\partial\varphi(x; v)$ stellt eine Verallgemeinerung der Richtungsableitung auf beliebige normierte Räume F dar. Insbesondere existiert er für alle konvexen Funktionale $\varphi : F \rightarrow \mathbb{R}$ an allen Stellen $x \in F$ in alle Richtungen $v \in F$.

Damit gewappnet wollen wir uns nun die angekündigte Aussage über Minimalargumente konvexer Funktionale überlegen. Betrachte dazu $F = \mathbb{R}^2$ und $\varphi(x, y) = x^2 + y^2$.

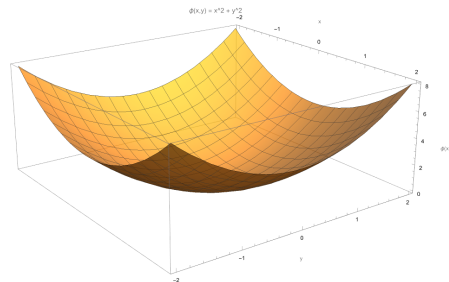


Abbildung 8: konvexes Funktional $\varphi(x, y) = x^2 + y^2$ mit Minimalargument $(0, 0)$

Anhand dieses Beispiels erkennen wir, dass das Minimalargument $s^* = (0, 0)$ die einzige Stelle im Graphen von φ ist, an welcher der Graph eine positive Steigung $\partial\varphi(s^*; s - s^*)$ in Richtung aller anderen Punkte $s \in \mathbb{R}^2$ aufweist. Dies führt auf folgenden Satz.

Satz 3.6 (Minimalargument) *Sei F ein normierter Vektorraum, $s^* \in S \subset F$ konvex, und $\varphi : F \rightarrow \mathbb{R}$ ein konvexes Funktional. Dann gilt $\varphi(s^*) = \inf_{s \in S} \varphi(s)$ genau dann wenn:*

$$\partial\varphi(s^*; s - s^*) \geq 0 \quad \forall s \in S$$

Sei $D_{x;v}(h) := \frac{1}{h}(\varphi(x + hv) - \varphi(x))$ wie in Definition 3.4, sodass

$$\partial\varphi(x; v) = \lim_{h \searrow 0} D_{x;v}(h)$$

gilt. Sei nun $s^* \in S$ mit $\varphi(s^*) = \inf_{s \in S} \varphi(s)$ das Minimalargument von φ . Da S konvex ist, gilt für beliebiges $s \in S$ und $h \in (0, 1)$:

$$s^* + h(s - s^*) = hs + (1 - h)s^* \in S$$

Nach Voraussetzung kann der Wert von φ angewendet auf beliebige Elemente in S nicht kleiner als $\varphi(s^*)$ werden, insbesondere also $\varphi(s^* + h(s - s^*)) \geq \varphi(s^*)$. Durch Umstellen sieht man daraus weiter $D_{s^*; s - s^*}(h) \geq 0$ für $h > 0$, und damit:

$$\partial\varphi(s^*; s - s^*) = \lim_{h \searrow 0} D_{s^*; s - s^*}(h) \geq 0$$

Sei nun umgekehrt $s^* \in S$ mit der Eigenschaft $\partial\varphi(s^*; s - s^*) \geq 0$ für alle $s \in S$ gegeben. Man sieht leicht Monotonie ($\#$) von $D_{x;v}$ ein und erhält damit weiter:

$$\begin{aligned} 0 \leq \partial\varphi(s^*; s - s^*) &= \lim_{h \searrow 0} D_{s^*; s - s^*}(h) \\ &\stackrel{\#}{\leq} D_{s^*; s - s^*}(1) = \varphi(s) - \varphi(s^*) \end{aligned}$$

Durch Umstellen sieht man folglich $\varphi(s^*) \leq \varphi(s)$ für alle $s \in S$ wie behauptet. ┐

Wie eingangs erläutert wollen wir dieses allgemeine Resultat nun schrittweise konkretisieren. Dazu erinnern wir uns an das zu lösende Vorhersageproblem $[Z_B|h]$. Die entsprechende Lösung $\hat{Z}_h \in \text{span}(Z_B)$ war als Lösung folgender Gleichung definiert:

$$\underbrace{\|\hat{Z}_h - Z_h\|_2}_{=\varphi_{Z_h}(\hat{Z}_h)} \leq \underbrace{\|\hat{Z}'_h - Z_h\|_2}_{=\varphi_{Z_h}(\hat{Z}'_h)}$$

Dies entspricht aber gerade der Beschreibung des Minimalarguments eines Funktionals $\varphi_f : (F, \|\cdot\|) \rightarrow \mathbb{R}$ mit $\varphi_f(u) := \|u - f\|$ für $f \in F$. Konkret $F = L_2$ und $f = Z_h$. Man rechnet leicht nach, dass dieses Funktional konvex ist.

Seien $u, v \in F$ und $\lambda \in (0, 1)$, dann gilt nach Einfügen einer Null:

$$\begin{aligned} \varphi_f(\lambda u + (1 - \lambda)v) &= \|\lambda u + (1 - \lambda)v - f + \lambda f - \lambda f\| \\ &= \|\lambda(u - f) + (1 - \lambda)(v - f)\| \end{aligned}$$

Mit der Dreiecksungleichung $(\star)$ der Norm erhält man weiter:

$$\begin{aligned}\varphi_f(\lambda u + (1 - \lambda)v) &\stackrel{\star}{\leq} \lambda \|u - f\| + (1 - \lambda) \|v - f\| \\ &= \lambda \varphi_f(u) + (1 - \lambda) \varphi_f(v)\end{aligned}$$

Damit können wir Satz 3.6 auf $\varphi_f(u) = \|u - f\|$ anwenden und erhalten folgenden Satz, den man auch Kolmogoroff-Kriterium nennt.

Satz 3.7 (Kolmogoroff-Kriterium) *Sei F ein normierter Raum, $s^* \in S \subset F$ konvex, und $f \in F$ fix. Dann gilt $\|s^* - f\| = \inf_{s \in S} \|s - f\|$ genau dann wenn:*

$$\partial \|s^* - f; s - s^*\| := \partial \varphi_f(s^*; s - s^*) \geq 0 \quad \forall s \in S$$

Dieser Satz nähert sich der Situation im Problem $[Z_B|h]$ durch Verwendung der Norm schon etwas an, jedoch können wir die Aussage noch weiter konkretisieren. Da $F = L_2$ euklidisch ist, das heißt, es gilt $\|u\| = \langle u, u \rangle^{1/2}$ für alle $u \in F$, können wir die entsprechende Normableitung wie folgt ausrechnen.

Sei $(F, \|\cdot\|)$ euklidisch und $\varphi(x) := \|x\| = \langle x, x \rangle^{1/2}$. Mit $u, v \in F$ gilt nach Kettenregel (analog zur Kettenregel für gewöhnliche Ableitungen) einerseits:

$$\partial \varphi^2(u; v) = 2\|u\| + \partial \varphi(u; v)$$

Und durch Rechnen mit der Definition 3.4 außerdem:

$$\begin{aligned}\partial \varphi^2(u; v) &= \lim_{h \searrow 0} \frac{1}{h} (\|u + hv\|^2 - \|u\|^2) \\ &= \lim_{h \searrow 0} (\|u\|^2 + 2h\langle u, v \rangle + h^2\|v\|^2 - \|u\|^2) \\ &= \lim_{h \searrow 0} \frac{1}{h} (2h\langle u, v \rangle + h^2\|v\|^2) = 2\langle u, v \rangle\end{aligned}$$

Gleichsetzen beider Ausdrücke liefert $2\|u\| + \partial \varphi(u; v) = 2\langle u, v \rangle$. Umstellen liefert:

$$\partial \varphi(u; v) = \frac{\langle u, v \rangle}{\|u\|}$$

Damit können wir Satz 3.7 weiter konkretisieren. Dazu stellt man auch fest, dass die Aussage $\|s^* - f\| = \inf_{s \in S} \|s - f\|$ gerade bedeutet, dass s^* die Bestapproximation an f aus S ist.

Satz 3.8 (Kolmogoroff-Kriterium) *Sei F ein euklidischer Raum, $s^* \in S \subset F$ konvex, und $f \in F$ fix. Dann ist s^* die Bestapproximation an f aus S genau dann wenn:*

$$\langle s^* - f, s - s^* \rangle \geq 0 \quad \forall s \in S$$

Ist der Approximationsraum S sogar ein Untervektorraum von F , so vereinfacht sich obiges Kriterium sogar noch weiter. Aus $S - s^* = S$ folgt weiter:

$$\langle s^* - f, s - s^* \rangle \geq 0 \quad \forall s \in S \quad \Leftrightarrow \quad \langle s^* - f, s \rangle = 0$$

Dies bedeutet $s^* - f \perp S$, also ist s^* gerade die orthogonale Projektion von f auf S . Bevor wir uns wieder unserem konkreten Problem $[Z_B|h]$ zuwenden, definieren wir zunächst den Projektionsoperator, welcher uns anschließend als Lösung fungieren wird.

Definition 3.9 (Projektionsoperator) Sei F ein euklidischer Raum und $S \subset F$. Dann nennen wir einen Operator $\mathcal{P} : F \rightarrow S$ orthogonale Projektion, falls gilt:

- (1) $\mathcal{P}|_S = \text{id}$ 'Projektion'
- (2) $\mathcal{P}^* = \mathcal{P}$ 'Orthogonalität'

Dabei ist $\mathcal{P}^*$ der zu $\mathcal{P}$ adjungierte Operator.

Unter Eigenschaft (1) wird Eigenschaft (2) durch $\mathcal{P}f - f \perp S$ nahegelegt.

$$\begin{aligned} \lceil & \mathcal{P}f - f \perp S \\ & \Leftrightarrow \langle \mathcal{P}f - f, s \rangle = 0 \\ & \Leftrightarrow \langle \mathcal{P}f, s \rangle = \langle f, s \rangle \\ & \Leftrightarrow \langle \mathcal{P}f, s \rangle = \langle f, \mathcal{P}s \rangle \end{aligned} \quad \rfloor$$

Ist S wieder ein endlich-dimensionaler Untervektorraum, so lässt sich $\mathcal{P}f$ konkret als Linearkombination der Basisvektoren darstellen.

Satz 3.10 (Darstellung Projektion) Sei F ein euklidischer Vektorraum und $S \subset F$ ein endlich-dimensionaler Untervektorraum mit Orthonormalbasis $\{\beta_1, \dots, \beta_n\}$, das heißt $\langle \beta_i, \beta_j \rangle = I\{i = j\}$. Dann gilt für alle $f \in F$:

$$\mathcal{P}f = \sum_{k \leq n} \langle f, \beta_k \rangle \beta_k$$

Da $\mathcal{P}f \in S$ landet und S ein Vektorraum mit Basis $\{\beta_k\}_{k \leq n}$ ist, können wir $\mathcal{P}f$ auf jeden Fall wie folgt darstellen (mit $\lambda_k \in \mathbb{R}$):

$$\mathcal{P}f = \sum_{k \leq n} \lambda_k \beta_k$$

Wir müssen also nur herausfinden, was λ_k ist. Dazu betrachten wir:

$$\begin{aligned} \langle \mathcal{P}f, \beta_j \rangle &= \left\langle \sum_k \lambda_k \beta_k, \beta_j \right\rangle \\ &= \sum_k \lambda_k \underbrace{\langle \beta_k, \beta_j \rangle}_{=I\{k=j\}} = \lambda_j \end{aligned}$$

Mit den Eigenschaften der orthogonalen Projektion folgt abschließend:

$$\lambda_j = \langle \mathcal{P}f, \beta_j \rangle \stackrel{(2)}{=} \langle f, \mathcal{P}\beta_j \rangle \stackrel{(1)}{=} \langle f, \beta_j \rangle$$

Beachte, dass uns diese Darstellung Eindeutigkeit gibt. Nun können wir also bereits eine allgemeine Lösung angeben. Es gilt

$$\hat{Z}_h = \mathcal{P}Z_h \text{ löst } [Z_B|h]$$

wobei $\mathcal{P} : L_2 \rightarrow \text{span}(Z_B)$ wie oben sei. Da es immer möglich ist, eine Orthonormalbasis eines endlich-dimensionalen Vektorraums zu finden (z.B. mit dem Gram-Schmidt-Verfahren), ist dieses Ergebnis dank Satz 3.10 sogar berechenbar.

3.2 Stationäre Lösung

Lit.: [1], [5]. In diesem Abschnitt erläutern wir, wie sich die allgemeine Lösung $\hat{Z}_h = \mathcal{P}Z_h$ von $[Z_B|h]$ im Fall einer stationären und zentrierten Zeitreihe Z bestimmen lässt. Dazu versuchen wir, die Koeffizienten ϕ_b von $\mathcal{P}Z_h$ direkt zu bestimmen, ohne vorher eine Orthonormalbasis von $\text{span}(Z_B)$ finden zu müssen. Betrachte also den Ansatz:

$$\hat{Z}_h = \mathcal{P}Z_h = \sum_{b \in B} \phi_b Z_b$$

Wir hatten uns bereits nach der Definition des Projektionsoperators überlegt, dass $\langle f, s \rangle = \langle \mathcal{P}f, s \rangle$ gilt für $\mathcal{P} : F \rightarrow S$ und $f \in F$ sowie $s \in S$. In unserer konkreten Situation mit $F = L_2$, $S = \text{span}(Z_B)$, $f = Z_h$, folgt für $s = Z_\beta$ mit $\beta \in B$ also:

$$\langle Z_h, Z_\beta \rangle = \langle \hat{Z}_h, Z_\beta \rangle = \sum_{b \in B} \phi_b \langle Z_b, Z_\beta \rangle$$

Ist die Zeitreihe Z zentriert, lassen sich die auftauchenden Skalarprodukte nur durch die Autokovarianzfunktion γ der Zeitreihe bestimmen.

Seien $t, s \in \mathbb{Z}$ zwei beliebige Zeitpunkte. Man berechnet

$$\begin{aligned} \langle Z_t, Z_s \rangle &= \int Z_t Z_s dP = E(Z_t Z_s) \\ &= \text{Cov}(Z_t, Z_s) + E(Z_t)E(Z_s) \\ &\stackrel{\star_1}{=} \text{Cov}(Z_t, Z_s) \stackrel{\star_2}{=} \gamma(|t - s|) \end{aligned}$$

wobei bei $\star_1$ die Zentriertheit ($E(Z_t) = 0$) und bei $\star_2$ die Stationarität ($\text{Cov}(Z_t, Z_s)$ nur vom Abstand $|t - s|$ abhängig) von Z einging.

Es gilt also $\langle Z_h, Z_\beta \rangle = \gamma(|h - \beta|)$ und analog $\langle Z_b, Z_\beta \rangle = \gamma(|b - \beta|)$. Damit lässt sich obige Gleichung also wie folgt darstellen.

$$\gamma(|h - \beta|) = \sum_{b \in B} \phi_b \gamma(|b - \beta|)$$

Insgesamt gibt es n verschiedene solche Gleichungen, wenn $B = \{b_1, \dots, b_n\}$ gilt. Als lineares Gleichungssystem ausgedrückt erhält man:

$$\underbrace{\begin{pmatrix} \gamma(|h - b_1|) \\ \vdots \\ \gamma(|h - b_n|) \end{pmatrix}}_{=: \gamma^{(h)}} = \underbrace{\left(\gamma(|b_i - b_j|) \right)_{i,j \leq n}}_{=: \Gamma_B} \cdot \underbrace{\begin{pmatrix} \phi_{b_1} \\ \vdots \\ \phi_{b_n} \end{pmatrix}}_{=: \phi^{(h)}}$$

Die gesuchten Koeffizienten $\phi^{(h)}$ sind also Lösung des obigen Gleichungssystems und lösen das Vorhersageproblem im Fall einer stationären Zeitreihe.

$$\begin{aligned} \hat{Z}_h &= \langle \phi^{(h)}, Z_B \rangle \text{ löst } [Z_B|h] \\ \text{wobei } \Gamma_B \phi^{(h)} &= \gamma^{(h)} \text{ gilt} \end{aligned}$$

Hierbei ist $\langle x, y \rangle = \sum_k x_k y_k$ das Standardskalarprodukt mit $x, y \in \mathbb{R}^n$. Gelegentlich ist im Folgenden implizit zu verstehen, ob wir dieses Skalarprodukt oder jenes für Zufallsvariablen $X, Y \in L_2$ wie auf Seite 7 definiert meinen.

3.3 1-Schritt-Vorhersage

Lit.: [1], [5]. Nachdem wir im vorherigen Abschnitt die Lösung des Problems $[Z_B|h]$ im Fall einer stationären Zeitreihe Z auf das Lösen des linearen Gleichungssystems $\Gamma_B \phi^{(h)} = \gamma^{(h)}$ reduziert haben, sind wir nun an einer effizienten Berechnungsmöglichkeit dieser Lösung interessiert. Lineare Gleichungssysteme mit n Unbekannten lassen sich typischerweise nur mit Laufzeit $\mathcal{O}(n^3)$ lösen, was nicht besonders effizient ist.

Daher konkretisieren wir uns auf das Problem $[Z_{[n]}|n+1]$, das heißt auf die 1-Schritt-Vorhersage (1SV) bei einer gemachten Beobachtung von $Z_1, \dots, Z_n$ (zur Erinnerung: $[n] = \{1, \dots, n\}$). Wir definieren die orthogonalen Projektionen $\mathcal{P}_n : L_2 \rightarrow S_n$ wobei $S_n := \text{span}(Z_{[n]})$ ist. Mit den bereits bestimmten Resultaten gilt also:

$$\hat{Z}_{n+1} = \mathcal{P}_n Z_{n+1} = \langle \phi^{(n+1)}, Z_{[n]} \rangle$$

Wir nutzen die Gelegenheit und klären kurz die Form aller relevanten Objekte. Der Koeffizientenvektor $\phi^{(n+1)}$ hat n unbekannte Einträge. Der Kovarianzvektor $\gamma^{(n+1)}$ hat n bekannte Einträge in Form der Werte $\gamma(n), \dots, \gamma(1)$ wie folgt.

$$\phi^{(n+1)} = \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_n \end{pmatrix}, \quad \gamma^{(n+1)} = \begin{pmatrix} \gamma(n) \\ \vdots \\ \gamma(1) \end{pmatrix}$$

Ebenso ist es lohnenswert, sich die Gestalt der Kovarianzmatrix $\Gamma_{[n]}$ klar zu machen.

$$\Gamma_{[n]} := \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma(1) \\ \gamma(n-1) & \cdots & \gamma(1) & \gamma(0) \end{pmatrix}$$

Die Kovarianzmatrix $\Gamma_{[n]}$ hat auf ihren Diagonalen immer den gleichen Wert stehen. Dies wird uns später noch sehr nützlich werden.

Nun können wir uns dem Ziel dieses Abschnittes widmen. Wurde bereits die 1-Schritt-Vorhersage $[Z_{[n-1]}|n]$ berechnet, so stehen die Koeffizienten $\phi^{(n)}$ dieser zur Verfügung. Anstatt also kostspielig ein Gleichungssystem zu lösen, könnte man versuchen, die Koeffizienten $\phi^{(n+1)}$ der Vorhersage $[Z_{[n]}|n+1]$ rekursiv aus $\phi^{(n)}$ zu berechnen.

Die dazu passende Idee ist es, den Vorhersageraum S_n so zu zerlegen, dass die Vorhersage $[Z_{[n]}|n+1]$ in mehrere Vorhersagen der Form $[Z_{[n-1]}|n]$ (d.h. unter Ausnutzung von Stationarität ineinander umwandelbar s.u.) zerfällt. Der natürlich wirkende Ansatz

$$S_n = S_{n-1} \oplus V_n \text{ mit } V_n = \text{span}(Z_n - \mathcal{P}_{n-1}Z_n)$$

welcher die neue Beobachtung Z_n orthogonal von S_n abspaltet, führt leider nicht zum Erfolg, denn damit gilt:

$$\mathcal{P}_n Z_{n+1} = \mathcal{P}_{n-1} Z_{n+1} + (\dots)$$

Es taucht also eine 2-Schritt-Vorhersage auf, für welche uns keine passenden Koeffizienten vorliegen. Auf Grund der Stationarität von Z hängen die Koeffizienten $\phi^{(n+1)}$ nur von der Autokovarianzfunktion γ und diese nur vom Abstand der Z_t ab. In diesem Sinne sind folgende 1-Schritt-Vorhersagen äquivalent.

$$[Z_1, \dots, Z_{n-1}|n] \leftrightarrow [Z_2, \dots, Z_n|n+1] \leftrightarrow [Z_2, \dots, Z_n|1]$$

Bei gegebener Beobachtung $Z_{[n]} = (Z_1, \dots, Z_n)$ spalten wir also nicht Z_n sondern Z_1 ab und betrachten anschließend die Vorhersagen $[Z_2, \dots, Z_n|n+1]$ sowie $[Z_2, \dots, Z_n|1]$. Seien dazu die orthogonalen Projektionen $\mathcal{U}_n : L_2 \rightarrow U_n := \text{span}(Z_2, \dots, Z_n)$ sowie $\mathcal{V}_n : L_2 \rightarrow V_n := \text{span}(Z_1 - \mathcal{U}_n Z_1)$ definiert. Damit zerfällt der Raum S_n wie folgt.

Satz 3.11 (Zerlegung Vorhersageraum) Seien S_n , U_n , V_n wie gehabt, dann gilt die orthogonale Zerlegung $S_n = U_n \oplus V_n$.

┌ Sind X ein Vektorraum und $Y, Z \subset X$ Untervektorräume, so gilt $X = Y \oplus Z$ genau dann wenn $Y \perp Z$ und wenn es für alle $x \in X$ eindeutige $y \in Y$, $z \in Z$ gibt, sodass $x = y + z$ gilt. Für unsere Situation bleibt also nur letzteres zu zeigen, denn per Konstruktion sind U_n , V_n Untervektorräume vom Vektorraum S_n und $U_n \perp V_n$ (Definition als Span und orthogonale Abspaltung). Für ein $S \in S_n$ gilt:

$$\begin{aligned} S &= \sum_{k=2}^n \phi_k Z_k + \phi_1 Z_1 + \phi_1 \mathcal{U}_n Z_1 - \phi_1 \mathcal{U}_n Z_1 \\ &= \sum_{k=2}^n \phi_k Z_k + \phi_1 \sum_{k=2}^n \lambda_k Z_k + \phi_1 (Z_1 - \mathcal{U}_n Z_1) \\ &= \underbrace{\sum_{k=2}^n (\phi_k + \phi_1 \lambda_k) Z_k}_{\in U_n} + \underbrace{\phi_1 (Z_1 - \mathcal{U}_n Z_1)}_{\in V_n} \in U_n + V_n \end{aligned}$$

Die Eindeutigkeit dieser Darstellung folgt aus der Eindeutigkeit der Darstellung als Linearkombination sowie der Eindeutigkeit der Projektion. ┐

Mit Hilfe dieser Zerlegung können wir nun die Vorhersage $\hat{Z}_{n+1}$ des Problems $[Z_{[n]}|n+1]$ aufschlüsseln, da die Projektion $\mathcal{P}_n$ auf $S_n = U_n \oplus V_n$ entsprechend in $\mathcal{U}_n + \mathcal{V}_n$ zerfällt.

$$\begin{aligned}\hat{Z}_{n+1} &= \mathcal{P}_n Z_{n+1} \\ &= \mathcal{U}_n Z_{n+1} + \mathcal{V}_n Z_{n+1} \\ &= \mathcal{U}_n Z_{n+1} + \alpha_n (Z_1 - \mathcal{U}_n Z_1)\end{aligned}$$

Wobei sich α_n entsprechend der Projektionsdarstellung in Satz 3.10 zu

$$\alpha_n = \frac{\langle Z_{n+1}, Z_1 - \mathcal{U}_n Z_1 \rangle}{\|Z_1 - \mathcal{U}_n Z_1\|^2}$$

ergibt. Diese Darstellung beinhaltet nun genau die Lösungen der gewünschten 1-Schritt-Vorhersagen. Wir wollen nun zeigen, dass sich diese unter Ausnutzung der Stationarität der Zeitreihe Z durch die bereits bekannten Koeffizienten $\phi^{(n)}$ bestimmen lassen.

Satz 3.12 *Sei Z eine (schwach) stationäre Zeitreihe und $\hat{Z}_n = \langle \phi^{(n)}, Z_{[n-1]} \rangle$ die Lösung von $[Z_{[n-1]}|n]$. Dann gilt für die Lösung $\mathcal{U}_n Z_{n+1}$ der Vorhersage $[Z_2, \dots, Z_n|n+1]$:*

$$\mathcal{U}_n Z_{n+1} = \sum_{k=2}^n \phi_{k-1}^{(n)} Z_k$$

Wir beweisen die Aussage, indem wir das Vorhersageproblem $[Z_B|h]$ mit $B = \{2, \dots, n\}$ und $h = n+1$ wie in Abschnitt 3.2 lösen. Das heißt, es gilt $\mathcal{U}_n Z_{n+1} = \hat{Z}_h = \langle \psi^{(h)}, Z_B \rangle$ mit $\Gamma_B \psi^{(h)} = \eta^{(h)}$. Der Kovarianzvektor $\eta^{(h)}$ hat dabei die Form

$$\eta^{(h)} \stackrel{h=n+1}{=} \begin{pmatrix} \gamma((n+1)-2) \\ \vdots \\ \gamma((n+1)-n) \end{pmatrix} = \begin{pmatrix} \gamma(n-1) \\ \vdots \\ \gamma(1) \end{pmatrix} = \gamma^{(n)}$$

des Kovarianzvektors $\gamma^{(n)}$ der Vorhersage $[Z_{n-1}|n]$ und die Kovarianzmatrix Γ_B hat dabei ebenfalls die Form

$$\begin{aligned}\Gamma_B &= (\gamma(|s-t|))_{s,t=2,\dots,n} \\ &= \begin{pmatrix} \gamma(0) & \cdots & \gamma(n-2) \\ \cdots & \ddots & \cdots \\ \gamma(n-2) & \cdots & \gamma(0) \end{pmatrix} \\ &= (\gamma(s-t))_{s,t=1,\dots,n-1} = \Gamma_{[n-1]}\end{aligned}$$

der Kovarianzmatrix $\Gamma_{[n-1]}$ der Vorhersage $[Z_{[n-1]}|n]$. Damit gilt aber weiter:

$$\psi^{(h)} = \Gamma_B^{-1} \eta^{(h)} = \Gamma_{[n-1]}^{-1} \gamma^{(n)} = \phi^{(n)}$$

Und somit $\hat{Z}_h = \langle \phi^{(n)}, Z_B \rangle = \sum_{k=2}^n \phi_{k-1}^{(n)} Z_k$ wie behauptet. Die Stationarität der Zeitreihe Z ging dabei in der Anwendung der Ergebnisse aus Abschnitt 3.2 ein. ┐

Im Anschluss wollen wir auch die Projektion $\mathcal{U}_n Z_1$ mit Hilfe der bereits bekannten Koeffizienten $\phi^{(n)}$ darstellen. Dies wird die eingangs erwähnte Stelle, an der uns die spezielle Form der Kovarianzmatrix Γ_B hilfreich ist. Betrachte das folgende Lemma.

Lemma 3.13 Seien $A \in \mathbb{R}^{n \times n}$ und $x, \bar{x} \in \mathbb{R}^n$ wie folgt:

$$A = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ a_2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_2 \\ a_n & \cdots & a_2 & a_1 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \bar{x} := \begin{pmatrix} x_n \\ \vdots \\ x_1 \end{pmatrix}$$

Wir nennen $\bar{x}$ den Kehrtvektor von x . Für $y := Ax$ gilt dann $\bar{y} = A\bar{x}$.

Wir beweisen dieses Lemma mittels Induktion über n . Für $n = 1$ ist die Aussage trivial, da dann der Kehrtvektor $\bar{x} = x$ bzw. $\bar{y} = y$ erfüllt. Für $n > 1$ betrachten wir (zur besseren Übersicht drucken wir Matrizen im Folgenden fett):

$$\mathbf{A}_n := \begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ a_2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_2 \\ a_n & \cdots & a_2 & a_1 \end{pmatrix} \stackrel{\#}{=} \left(\begin{array}{c|c} \mathbf{A}_{n-1} & \begin{smallmatrix} | \\ a^{(n)} \end{smallmatrix} \\ \hline a_n \cdots a_2 & | \end{array} \right) \stackrel{*}{=} \left(\begin{array}{c|c} \begin{smallmatrix} | \\ a^{(n)} \end{smallmatrix} & \begin{smallmatrix} a_2 \cdots a_n \\ \hline \mathbf{A}_{n-1} \end{smallmatrix} \end{array} \right)$$

Wobei die Vektoren $x^{(n)} := (x_1 \cdots x_n)^T \stackrel{\#}{=} \begin{pmatrix} x^{(n-1)} \\ x_n \end{pmatrix}$ mit $\overline{x^{(n)}} \stackrel{*}{=} \begin{pmatrix} x_n \\ \overline{x^{(n-1)}} \end{pmatrix}$ seien.

Zu zeigen ist nun, dass $\overline{y^{(n)}} = \mathbf{A}_n \overline{x^{(n)}}$ mit $y^{(n)} := \mathbf{A}_n x^{(n)}$ aus $\overline{y^{(n-1)}} = \mathbf{A}_{n-1} \overline{x^{(n-1)}}$ folgt. Dazu betrachten wir komponentenweise ($1 < k < n$) den Vektor $\overline{y^{(n)}}$:

$$\begin{aligned} (\overline{y^{(n)}})_k &= (y^{(n)})_{n-k+1} = (A_n x^{(n)})_{n-k+1} \\ &\stackrel{\#}{=} \underbrace{(\mathbf{A}_{n-1} x^{(n-1)})_{n-k+1}}_{=(y^{(n-1)})_{n-k+1} = (\overline{y^{(n-1)}})_k} + \underbrace{(a^{(n)})_{n-k+1}}_{=(a^{(n)})_k} x_n \\ &\stackrel{IV}{=} (\mathbf{A}_{n-1} \overline{x^{(n-1)}})_k + (a^{(n)})_k x_n \stackrel{*}{=} (\mathbf{A}_n \overline{x^{(n)}})_k \end{aligned}$$

Die Fälle $k = 1$ und $k = n$ sieht man sofort. Insgesamt folgt die Behauptung. ┐

Gewappnet mit dieser Aussage können wir nun folgenden Satz zeigen.

Satz 3.14 Sei Z eine (schwache) stationäre Zeitreihe und $\hat{Z}_n = \langle \phi^{(n)}, Z_{[n-1]} \rangle$ die Lösung von $[Z_{[n-1]}|n]$. Dann gilt für die Lösung $\mathcal{U}_n Z_1$ der Vorhersage $[Z_2, \dots, Z_n|1]$:

$$\mathcal{U}_n Z_1 = \sum_{k=2}^n \overline{\phi_{k-1}^{(n)}} Z_k$$

Wir beweisen die Aussage analog zum Beweis von Satz 3.12. Betrachte daher das Vorhersageproblem $[Z_B|h]$ mit $B = \{2, \dots, n\}$ und $h = 1$. Dann gilt $\mathcal{U}_n Z_1 = \hat{Z}_1 = \langle \psi^{(h)}, Z_B \rangle$ mit $\Gamma_B \psi^{(h)} = \eta^{(h)}$. Der Kovarianzvektor $\eta^{(h)}$ hat dabei die Form

$$\eta^{(h)} \stackrel{h=1}{=} \begin{pmatrix} \gamma(2-1) \\ \vdots \\ \gamma(n-1) \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(n-1) \end{pmatrix} = \overline{\gamma^{(n)}}$$

des Kehrtkovarianzvektors $\overline{\gamma^{(n)}}$ und für die Kovarianzmatrix gilt $\Gamma_B = \Gamma_{[n-1]}$ komplett analog zum Beweis von Satz 3.12. Mit obiger Gleichung gilt also

$$\overline{\gamma^{(n)}} = \eta^{(1)} = \Gamma_{[n-1]} \psi^{(1)}$$

und mit Lemma 3.13 auch $\overline{\gamma^{(n)}} = \Gamma_{[n-1]} \overline{\phi^{(n)}}$ sodass wir insgesamt erhalten:

$$\begin{aligned} \Gamma_{[n-1]} \psi^{(1)} &= \Gamma_{[n-1]} \overline{\phi^{(n)}} & | \Gamma_{[n-1]}^{-1} \\ \Leftrightarrow \psi^{(1)} &= \overline{\phi^{(n)}} \end{aligned}$$

Wobei die Invertierbarkeit von $\Gamma_{[n-1]}$ durch die eindeutige Lösbarkeit von $\gamma^{(n)} = \Gamma_{[n-1]} \phi^{(n)}$ durch $\mathcal{P}_{n-1} Z_n$ folgt (vgl. Darstellung Projektionsoperator). ┐

An dieser Stelle wollen wir ein kurzes Zwischenfazit ziehen, um nachzuvollziehen, was wir bereits erreicht haben und wo wir noch hin wollen. Basierend auf dem Koeffizientenvektor $\phi^{(n)}$ des Vorhersageproblems $[Z_{[n-1]}|n]$ wollen wir die Vorhersage $\hat{Z}_{n+1}$ bestimmen, also das Problem $[Z_n|n+1]$ lösen. Dazu haben wir den Vorhersageraum $S_n = U_n \oplus V_n$ zerlegt und die Darstellung

$$\hat{Z}_{n+1} = \mathcal{U}_n Z_{n+1} + \alpha_n (Z_1 - \mathcal{U}_n Z_1)$$

erhalten. Beide Projektionsterme ($\mathcal{U}_n Z_{n+1}$ und $\mathcal{U}_n Z_1$) haben wir bereits durch $\phi^{(n)}$ ausgedrückt, sodass zur konkreten Berechnung der Vorhersage $\hat{Z}_{n+1}$ lediglich α_n fehlt. Dies gehen wir nun an; dazu werden wir folgendes Lemma benötigen.

Lemma 3.15 *Seien Z eine (schwach) stationäre Zeitreihe und Projektionen $\mathcal{U}_n : L_2 \rightarrow U_n$ sowie $\mathcal{P}_{n-1} : L_2 \rightarrow S_{n-1}$ wie gehabt gegeben. Dann gilt:*

$$\|Z_1 - \mathcal{U}_n Z_1\|^2 = \|Z_n - \mathcal{P}_{n-1} Z_n\|^2$$

Wir zeigen diese Aussage, indem wir beide Seiten der Gleichung auf denselben Ausdruck bringen. Wir beginnen mit der rechten Seite; es gilt:

$$\begin{aligned} \|Z_n - \mathcal{P}_{n-1} Z_n\|^2 &\stackrel{*1}{=} \left\| Z_n - \sum_{k=1}^{n-1} \phi_k^{(n)} Z_k \right\|^2 \\ &= \left\| \sum_{k=1}^n \lambda_k Z_k \right\|^2 \quad \text{mit} \quad \lambda_k := \begin{cases} -\phi_k^{(n)} & k < n \\ 1 & k = n \end{cases} \\ &= \left\langle \sum_k \lambda_k Z_k, \sum_k \lambda_k Z_k \right\rangle \stackrel{*2}{=} \sum_{i,j} \lambda_i \lambda_j \gamma(|i-j|) \end{aligned}$$

Bei $\star_2$ ging die Stationarität von Z ein (Skalarprodukt $\rightarrow$ Kovarianzfunktion) und bei $\star_1$ haben wir $\mathcal{P}_{n-1}Z_n$ als Linearkombination mit Hilfe des Koeffizientenvektors $\phi^{(n)}$ dargestellt.

Analog können wir nun mit der linken Seite verfahren. Wir wissen nach Satz 3.14, dass sich $\mathcal{U}_n Z_1$ ebenfalls als eine solche Linearkombination darstellen lässt; es gilt:

$$\begin{aligned} \|Z_1 - \mathcal{U}_n Z_1\|^2 &= \|Z_1 - \sum_{k=2}^n \overline{\phi_{k-1}^{(n)}} Z_k\|^2 \\ &= \|\sum_{k=1}^n \eta_k Z_k\|^2 \quad \text{mit} \quad \eta_k := \begin{cases} 1 & k = 1 \\ -\overline{\phi_{k-1}^{(n)}} & k > 1 \end{cases} \\ &= \langle \sum_k \eta_k Z_k, \sum_k \eta_k Z_k \rangle = \sum_{i,j} \eta_i \eta_j \gamma(|i-j|) \end{aligned}$$

Man sieht leicht $\lambda = \begin{pmatrix} -\phi^{(n)} \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -\overline{\phi^{(n)}} \end{pmatrix} = \bar{\eta}$, wodurch die Behauptung folgt. └

Der erhaltene Ausdruck ist gerade die mittlere quadratische Differenz zwischen der 1-Schritt-Vorhersage $\hat{Z}_n = \mathcal{P}_{n-1}Z_n$ des Problems $[Z_{[n-1]}|n]$ und dem tatsächlichen Wert Z_n . In anderen Worten handelt es sich hierbei um den *n-ten mittleren 1SV-Fehler*:

$$e_n := \|Z_n - \hat{Z}_n\|^2 = \|Z_n - \mathcal{P}_{n-1}Z_n\|^2$$

Damit haben wir nun alles zusammen, um eine handhabbare Formel für α_n zu bestimmen. Nach dem Ausdruck auf Seite 15 erhalten wir:

$$\begin{aligned} \alpha_n e_n &= \langle Z_{n+1}, Z_1 - \mathcal{U}_n Z_1 \rangle \\ &\stackrel{3.14}{=} \underbrace{\langle Z_{n+1}, Z_1 \rangle}_{=\gamma(n)} - \sum_{k=2}^n \overline{\phi_{k-1}^{(n)}} \underbrace{\langle Z_{n+1}, Z_k \rangle}_{=\gamma(n-k+1)} \\ &\stackrel{*}{=} \gamma(n) - \sum_{k=1}^{n-1} \overline{\phi_k^{(n)}} \gamma(n-k) \stackrel{n \geq 1}{=} \left\langle \begin{pmatrix} 1 \\ -\overline{\phi^{(n)}} \end{pmatrix}, \gamma^{(n+1)} \right\rangle \end{aligned}$$

Wobei wir bei $\star$ einen Indexshift von k auf $k-1$ gemacht haben. Man rechnet nach, dass für $x, y \in \mathbb{R}^n$ gilt $\langle \bar{x}, y \rangle = \langle x, \bar{y} \rangle$. Damit können wir unsere erste Ergebnisformel in diesem Abschnitt angeben. Mit $\phi^{(n)}$ wie gehabt und $\gamma^{(n+1)} = (\gamma(n) \cdots \gamma(1))^T$ gilt:

$$\alpha_n = \frac{1}{e_n} \left\langle \begin{pmatrix} -\phi^{(n)} \\ 1 \end{pmatrix}, \overline{\gamma^{(n+1)}} \right\rangle, \quad n > 1$$

Da der Wert $e_n = \|Z_n - \hat{Z}_n\|^2$ lediglich dem mittleren und nicht dem konkreten Fehler der vorhergegangenen Prognose $[Z_{[n-1]}|n]$ entspricht, können wir diesen nicht einfach elementar berechnen. Stattdessen ist es möglich, auch für e_n eine rekursive Formel zu bestimmen. Dazu benötigen wir zunächst folgendes Lemma.

Lemma 3.16 Sei $(X, \langle -, - \rangle)$ ein euklidischer Vektorraum und $\|-\| := \langle -, - \rangle^{1/2}$ eine Norm auf diesem. Für $a, b, c \in X$ gilt:

- $\|a - b\|^2 = \|a\|^2 + \|b\|^2 - 2\langle a, b \rangle$
- $\|a - b - c\|^2 = \|a - b\|^2 + \|c\|^2 - 2\langle a - b, c \rangle$

┌ Punkt 1 impliziert Punkt 2, indem man $a \leftarrow a - b$ und $b \leftarrow c$ ersetzt. Punkt 1 folgt aus der Bilinearität und Symmetrie des Skalarproduktes:

$$\begin{aligned} \|a - b\|^2 &= \langle a - b, a - b \rangle \\ &= \underbrace{\langle a, a \rangle}_{=\|a\|^2} + \underbrace{\langle b, b \rangle}_{=\|b\|^2} - 2\langle a, b \rangle \end{aligned}$$

└

Damit können wir nun eine Rekursion für den mittleren Fehler bestimmen. Wobei wir speziell e_{n+1} betrachten, da dabei bereits bekannte Ausdrücke auftauchen.

$$\begin{aligned} e_{n+1} &= \|Z_{n+1} - \hat{Z}_{n+1}\|^2 = \|Z_{n+1} - \mathcal{P}_n Z_{n+1}\|^2 \\ &\stackrel{S_n = U_n \oplus V_n}{=} \|Z_{n+1} - \mathcal{U}_n Z_{n+1} - \mathcal{V}_n Z_{n+1}\|^2 \\ &\stackrel{3.16}{=} \|Z_{n+1} - \mathcal{U}_n Z_{n+1}\|^2 + \|\mathcal{V}_n Z_{n+1}\|^2 - 2\langle Z_{n+1} - \mathcal{U}_n Z_{n+1}, \mathcal{V}_n Z_{n+1} \rangle \end{aligned}$$

Für den ersten dieser Terme zeigt man exakt wie im Beweis von Lemma 3.15, dass

$$\|Z_{n+1} - \mathcal{U}_n Z_{n+1}\|^2 = \|Z_n - \mathcal{P}_{n-1} Z_n\|^2 = e_n$$

gilt. Den zweiten Term werten wir zunächst wie auf Seite 15 aus und wenden dann Lemma 3.15 an, um folgenden Ausdruck zu erhalten:

$$\begin{aligned} \|\mathcal{V}_n Z_{n+1}\|^2 &= \|\alpha_n(Z_1 - \mathcal{U}_1 Z_1)\|^2 \\ &= \alpha_n^2 \|Z_1 - \mathcal{U}_1 Z_1\|^2 \\ &\stackrel{3.15}{=} \alpha_n^2 \|Z_n - \mathcal{P}_{n-1} Z_n\|^2 = \alpha_n^2 e_n \end{aligned}$$

Den letzten Term kann man wie folgt vereinfachen. Insbesondere geht bei $\star$ ein, dass $U_n \perp V_n$ per Konstruktion gilt. Außerdem werten wir $\mathcal{V}_n Z_{n+1}$ wieder wie zuvor aus:

$$\begin{aligned} &\langle Z_{n+1} - \mathcal{U}_n Z_{n+1}, \mathcal{V}_n Z_{n+1} \rangle \\ &= \langle Z_{n+1}, \mathcal{V}_n Z_{n+1} \rangle - \underbrace{\langle \mathcal{U}_n Z_{n+1}, \mathcal{V}_n Z_{n+1} \rangle}_{=0 \text{ wg. } \star} \\ &= \alpha_n \underbrace{\langle Z_{n+1}, Z_1 - \mathcal{U}_1 Z_1 \rangle}_{=\alpha_n e_n \text{ siehe S.17}} = \alpha_n^2 e_n \end{aligned}$$

Insgesamt gilt also $e_{n+1} = e_n + \alpha_n^2 e_n - 2\alpha_n^2 e_n = e_n(1 - \alpha_n^2)$ was wir wie folgt festhalten.

$$e_n = e_{n-1}(1 - \alpha_{n-1}^2), \quad n > 1$$

An dieser Stelle tritt ein weiteres Problem auf: Wollen wir e_n mittels obiger Formel bestimmen, benötigen wir sowohl e_{n-1} als auch α_{n-1} . Diese benötigen dann jedoch wieder e_{n-2} und α_{n-2} und so weiter. Tatsächlich müssen wir einsehen, dass α_n, e_n als auch $\phi^{(n)}$ alle von einander bzw. den vorherigen Größen $(n-1)$ abhängen. Wir werden diese Abhängigkeiten zwar nicht los, können jedoch auch für die Koeffizienten $\phi^{(n)}$ eine Rekursion herleiten und durch Bestimmung geeigneter Startwerte für alle Größen ein effizientes Berechnungsverfahren erhalten.

Zunächst versuchen wir also $\phi^{(n+1)}$ aus $\phi^{(n)}$ zu bestimmen. Sei dazu $\hat{Z}_{n+1} = \mathcal{P}_n Z_{n+1}$ die Lösung von $[Z_{[n]}|n+1]$. Basierend auf $S_n = U_n \oplus V_n$ wählen wir den gleichen Ansatz wie auf Seite 15, welchen wir dank 3.12 und 3.14 (*) wie folgt ausformulieren können.

$$\begin{aligned} \hat{Z}_{n+1} &= \alpha_n Z_1 + \mathcal{U}_n Z_{n+1} - \alpha_n \mathcal{U}_n Z_1 \\ &\stackrel{*}{=} \alpha_n Z_1 + \sum_{k=2}^n \phi_{k-1}^{(n)} Z_k - \alpha_n \sum_{k=2}^n \overline{\phi_{k-1}^{(n)}} Z_k \\ &= \alpha_n Z_1 + \sum_{k=2}^n \left(\phi_{k-1}^{(n)} - \alpha_n \overline{\phi_{k-1}^{(n)}} \right) Z_k \end{aligned}$$

Auf Grund der Eindeutigkeit der Darstellung $\hat{Z}_{n+1} = \mathcal{P}_n Z_{n+1} = \sum_{k=1}^n \phi_k^{(n+1)} Z_k$ liefert Koeffizientenvergleich $\phi_1^{(n+1)} = \alpha_n$ und $\phi_k^{(n+1)} = \phi_{k-1}^{(n)} - \alpha_n \overline{\phi_{k-1}^{(n)}}$ für $k > 1$. Also gilt:

$$\phi^{(n+1)} = \begin{pmatrix} \alpha_n \\ \phi^{(n)} - \alpha_n \overline{\phi^{(n)}} \\ 1 \end{pmatrix}, \quad n >$$

Damit haben wir für alle relevanten Größen bemerkenswert schöne Rekursionsgleichungen gefunden und es bleibt nur noch die Frage nach ihren Startwerten. Dazu betrachten wir den minimalen, noch sinnvollen Fall $n = 1$. Das heißt, wir gehen davon aus Z_1 beobachtet zu haben und wollen die Vorhersage $\hat{Z}_2$ bestimmen. Wir wissen bereits:

$$\hat{Z}_2 = \mathcal{P}_1 Z_2 = \phi_1^{(2)} Z_1 =: \phi Z_1$$

Mit der Darstellung aus Satz 3.10 können wir den Koeffizienten ϕ ebenfalls angeben.

$$\phi \stackrel{3.10}{=} \frac{\langle Z_2, Z_1 \rangle}{\langle Z_1, Z_1 \rangle} = \frac{\gamma(1)}{\gamma(0)}$$

Betrachtet man nun den mittleren Prognosefehler $e_2 = \|Z_2 - \hat{Z}_2\|^2$ ergibt sich:

$$\begin{aligned} e_2 &= \|Z_2 - \phi Z_1\|^2 \\ &\stackrel{3.16}{=} \|Z_2\|^2 + \phi^2 \|Z_1\|^2 - 2\phi \langle Z_2, Z_1 \rangle \end{aligned}$$

All diese Ausdrücke zerfallen, wie wir bereits häufig gesehen haben, zu γ wie folgt:

$$\begin{aligned} e_2 &= \gamma(0) - \phi^2 \gamma(0) + 2\phi \gamma(1) \\ &\stackrel{*}{=} \gamma(0) - \frac{\gamma(1)^2}{\gamma(0)} = \gamma(0)(1 - \phi^2) \end{aligned}$$

Wobei wir bei $\star$ die oben bestimmte Darstellung von ϕ eingesetzt haben. Vergleicht man diesen Ausdruck mit der Rekursionsformel $e_n = e_{n-1}(1 - \alpha_{n-1}^2)$ so erkennt man die gesuchten Startwerte als:

$$e_1 := \gamma(0) \quad \text{und} \quad \alpha_1 := \frac{\gamma(1)}{\gamma(0)}$$

Der Koeffizientenvektor $\phi^{(1)}$ ist nach unserer Definition leer. Der nächste $\phi^{(2)}$ lässt sich dann bereits aus α_1 bestimmen. Insgesamt rechnet man also $\phi^{(1)}, e_1, \alpha_1$, dann $\phi^{(2)}, e_2, \alpha_2$, dann $\phi^{(3)}, e_3, \alpha_3$ usw. in dieser Reihenfolge aus. Dieses Verfahren läuft in linearer Zeit $\mathcal{O}(n)$ und die gesuchte 1-Schritt-Vorhersage ist wie gehabt gegeben als:

$$\hat{Z}_{n+1} = \langle \phi^{(n+1)}, Z_{[n]} \rangle$$

3.4 h-Schritt-Vorhersage

Lit.: [1], [5]. Bevor wir uns dem letzten Teil dieser Arbeit widmen, wollen wir die erhaltenen Ergebnisse des vorherigen Abschnittes etwas verallgemeinern. Anstatt basierend auf einer Beobachtung $Z_{[n]}$ nur den nächsten Wert $\hat{Z}_{n+1}$ vorherzusagen, können wir stattdessen gleich die nächsten $h \in \mathbb{N}$ vielen Werte $\hat{Z}_{n+1}, \dots, \hat{Z}_{n+h}$ vorhersagen.

Das wesentliche Werkzeug dazu wird der folgende Satz. Zur Erinnerung: Wir betrachten orthogonale Projektionen $\mathcal{P}_n : L_2 \rightarrow S_n := \text{span}(Z_{[n]})$ mit $Z_{[n]} := (Z_1, \dots, Z_n)$.

Satz 3.17 (Iterationseigenschaft) Für $m \geq n$ gilt: $\mathcal{P}_n = \mathcal{P}_n \mathcal{P}_m$

Wir wollen zunächst Satz 3.10 benutzen, um die betrachteten Projektionen darzustellen. Dazu benötigen wir eine Orthonormalbasis von S_n , welche wir durch das Gram-Schmidt-Verfahren erhalten können; demnach bilden

$$Y_1 := Z_1, \quad Y_k := Z_k - \mathcal{P}_{k-1} Z_k \quad (k > 1)$$

eine Orthogonalbasis von S_n . Durch Normierung mit $\|Y_k\|_2 = \langle Y_k, Y_k \rangle^{1/2}$ erhalten wir die gesuchte Orthonormalbasis. Für $X \in L_2$ ergibt sich:

$$\begin{aligned} \mathcal{P}_n \mathcal{P}_m X &\stackrel{3.10}{=} \mathcal{P}_n \sum_{k=1}^m \frac{\langle X, Y_k \rangle}{\langle Y_k, Y_k \rangle} Y_k \\ &\stackrel{*}{=} \sum_{k=1}^m \frac{\langle X, Y_k \rangle}{\langle Y_k, Y_k \rangle} \mathcal{P}_n Y_k \end{aligned}$$

Wobei bei $\star$ die Linearität der Projektion einging. Da Y_k Element der betrachteten Orthogonalbasis ist, gilt $\mathcal{P}_n Y_k = Y_k I\{k \leq n\}$. Daher fallen alle Terme obiger Summe für $k > n$ weg und es bleibt wie behauptet:

$$\mathcal{P}_n \mathcal{P}_m X = \sum_{k=1}^n \frac{\langle X, Y_k \rangle}{\langle Y_k, Y_k \rangle} Y_k = \mathcal{P}_n X$$

└

Damit können wir nun für alle $\mathbf{h} = 1, \dots, \mathbf{h}$ die $\mathbf{h}$ -Schritt-Vorhersage $\hat{Z}_{n+\mathbf{h}}$ bestimmen (man missachte den Notationsmissbrauch). Betrachte dazu den folgenden Ausdruck:

$$\begin{aligned} \hat{Z}_{n+h} &= \mathcal{P}_n Z_{n+h} \\ &\stackrel{3.17}{=} \mathcal{P}_n \mathcal{P}_{n+h-1} Z_{n+h} \end{aligned}$$

Die Projektion $\mathcal{P}_{n+h-1} Z_{n+h}$ ist dabei genau die Lösung der 1SV $[Z_{[n+h-1]}|n+h]$, welche nach dem vorherigen Abschnitt gleich $\langle \phi^{(n+h)}, Z_{[n+h-1]} \rangle$ ist. Daher gilt:

$$\begin{aligned} \hat{Z}_{n+h} &= \mathcal{P}_n \sum_{k=1}^{n+h-1} \phi_k^{(n+h)} Z_k \\ &\stackrel{\star}{=} \sum_{k=1}^n \phi_k^{(n+h)} \underbrace{\mathcal{P}_n Z_k}_{= Z_k} + \sum_{k=n+1}^{n+h-1} \phi_k^{(n+h)} \underbrace{\mathcal{P}_n Z_k}_{= \hat{Z}_k} \end{aligned}$$

Wobei bei $\star$ die Linearität der Projektion einging. Dass $\mathcal{P}_n Z_k = Z_k$ für $k \leq n$ gilt, liegt daran, dass dabei bereits $Z_k \in S_n$ ist. Der gleiche Ausdruck $\mathcal{P}_n Z_k = \hat{Z}_k$ für $k > n$ entspricht den $\mathbf{h}$ -Schritt-Vorhersagen $\hat{Z}_{n+h}$. Insgesamt gibt uns obige Gleichung mit $H = n+1, \dots, n+h$ also $\hat{Z}_H = (\hat{Z}_{n+1}, \dots, \hat{Z}_{n+h})$ folgende Formel an die Hand.

$$\hat{Z}_{n+h} = \left\langle \phi^{(n+h)}, \begin{pmatrix} Z_{[n]} \\ \hat{Z}_H \end{pmatrix} \right\rangle$$

4 Schätzung aus Beobachtung

4.1 Autokovarianzfunktion

Lit.: [2], [7]. In diesem Abschnitt wollen wir die letzte fehlende Komponente zur praktischen Vorhersage von Zeitreihen beschreiben: die Autokovarianzfunktion γ . Bislang haben wir zwar ein Verfahren, welches gesuchte $\mathbf{h}$ -Schritt-Vorhersagen effizient berechnen kann, jedoch immer auf Grundlage von γ ; beispielsweise:

$$\alpha_n = \frac{1}{e_n} \left\langle \begin{pmatrix} -\phi^{(n)} \\ 1 \end{pmatrix}, \overline{\gamma^{(n+1)}} \right\rangle, \quad e_1 = \gamma(0), \quad \alpha_1 = \frac{\gamma(1)}{\gamma(0)} \quad \text{etc.}$$

Um unser Verfahren nun endgültig, ausschließlich auf der gemachten Beobachtung $Z_{[n]}$ basieren zu lassen, müssen wir also γ aus jener Beobachtung bestimmen. In anderen Worten suchen wir einen Schätzer $\hat{\gamma}(h)$ für die Autokovarianz $\gamma(h) = \text{Cov}(Z_0, Z_h)$. Wir

gehen ab jetzt von einer *stark* stationären und zentrierten Zeitreihe Z aus. Folgende Überlegung liefert einen sinnvollen Ansatz für einen möglichen Schätzer $\hat{\gamma}(h)$. Es gilt:

$$\gamma(h) = \text{Cov}(Z_0, Z_h) = E(Z_0 Z_h) - \underbrace{E(Z_0)E(Z_h)}_{=0 \text{ da } Z \text{ zentriert}}$$

Wir müssen also den Erwartungswert von $Z_0 Z_h$ schätzen, weshalb es nahe liegt $\hat{\gamma}(h)$ als Mittelwert aller verfügbaren Beobachtungen von $Z_0 Z_h$ zu definieren. Allerdings steht uns nur jeweils eine Beobachtung der $Z_1, \dots, Z_n$ zur Verfügung, sodass dies allein noch nicht sinnvoll ist. Auf Grund der starken Stationarität von Z gilt jedoch weiter

$$\gamma(h) = E(Z_0 Z_h) = E(Z_t Z_{t+h}) \quad \forall t \in \mathbb{Z}$$

denn $E(Z_0 Z_h)$ ist nur abhängig von der gemeinsamen Verteilung von (Z_0, Z_h) , welche jedoch gleich jener von (Z_t, Z_{t+h}) für alle $t \in \mathbb{Z}$ ist. Damit können wir auch die Beobachtung der anderen Z_t benutzen, um $\gamma(h)$ sinnvoll als Mittelwert zu schätzen.

Definition 4.1 (Autokovarianz-Schätzer) Zu einer Beobachtung $Z_{[n]} = (Z_1, \dots, Z_n)$ definieren wir den Autokovarianz-Schätzer $\hat{\gamma}(h)$ mit $h < n$ als:

$$\hat{\gamma}(h) := \frac{1}{n-h} \sum_{t=1}^{n-h} Z_t Z_{t+h}$$

Man beachte, dass die Benutzung dieses Schätzers die maximale Vorhersagetiefe h wegen $h < n$ festlegt. Im Folgenden wollen wir uns davon überzeugen, dass obiger Schätzer $\hat{\gamma}(h)$ typische Güteeigenschaften besitzt.

Satz 4.2 (Erwartungstreue) Für alle $h < n$ ist $\hat{\gamma}(h)$ erwartungstreu, das heißt:

$$E(\hat{\gamma}(h)) = \gamma(h)$$

┌ Diese Eigenschaft lässt sich unter Ausnutzung der Linearität des Erwartungswertes einfach nachrechnen. Zur Erinnerung: Es gilt $E(Z_t Z_{t+h}) = \langle Z_t, Z_{t+h} \rangle = \gamma(h)$.

$$E(\hat{\gamma}(h)) = \frac{1}{n-h} \sum_{t=1}^{n-h} E(Z_t Z_{t+h}) = \frac{n-h}{n-h} \gamma(h) = \gamma(h)$$

└

Als nächstes wollen wir die Konsistenz $\hat{\gamma}(h) \xrightarrow{P} \gamma(h)$ für $n \rightarrow \infty$ dieses Schätzers zeigen. Dafür müssen wir jedoch deutlich mehr arbeiten als für seine Erwartungstreue. Das wesentliche Werkzeug dazu wird der sogenannte Ergodensatz, welcher eine Verallgemeinerung des Gesetzes der großen Zahlen auf stark stationäre stochastische Prozesse (insbesondere also Zeitreihen) darstellt.

Definition 4.3 (ergodische Abbildung) Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $T : \Omega \rightarrow \Omega$ eine Abbildung darauf. Dann heißt T ...

- *maßerhaltend* falls $P^T = P$ gilt, das heißt $P(T^{-1}(A)) = P(A)$ für alle $A \in \mathcal{A}$.
- *ergodisch* falls T maßerhaltend ist und für alle $A \in \mathcal{A}$ mit $T^{-1}(A) = A$ bereits $P(A) \in \{0, 1\}$ gilt.

Satz 4.4 (Ergodensatz) Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $T : \Omega \rightarrow \Omega$ ergodisch. Weiter sei $X : \Omega \rightarrow \mathbb{R}$ eine P -integrierbare Zufallsvariable. Dann gilt:

$$\frac{1}{n} \sum_{k=1}^n X \circ T^{k-1} \xrightarrow{f.s.} E(X) \quad \text{für } n \rightarrow \infty$$

Um den Ergodensatz auf unsere Situation anzuwenden, benötigen wir außerdem noch den Maß-Eindeutigkeitssatz wie folgt. Anschließend formulieren wir den Ergodensatz für Zeitreihen. Beachte: Wir fordern dort erstmalig mehr als Stationarität.

Satz 4.5 (Maß-Eindeutigkeitssatz) Seien μ_1, μ_2 Maße auf $(\Omega, \mathcal{A})$ Messraum mit $\mathcal{A} = \sigma(\mathcal{M})$ wobei $\mathcal{M}$ schnittstabil sei. Dann ist $\mu_1 = \mu_2$ falls gilt:

- $\mu_1(M) = \mu_2(M)$ für alle $M \in \mathcal{M}$
- es gibt $A_n \nearrow \Omega$ aus $\mathcal{M}$ mit $\mu_1(A_n), \mu_2(A_n) < \infty$

Wobei $A_n \nearrow \Omega$ eine aufspannende Folge ist. Das heißt $A_n \subset A_{n+1}$ und $\bigcup_n A_n = \Omega$.

Satz 4.6 (Ergodensatz Zeitreihen) Sei $Y = Y_{\mathbb{N}} \subset \mathbb{R}$ eine stark stationäre und P -integrierbare Zeitreihe über $(\Omega, \mathcal{A}, P)$. Dann gilt für $n \rightarrow \infty$:

$$\frac{1}{n} \sum_{k=1}^n Y_k \xrightarrow{f.s.} E(Y_1)$$

This statement and its proof are actually false. The right-hand side of this asymptotic needs an additional term given by a special RV.

Wie bereits erläutert wollen wir diesen Satz durch Anwendung des Ergodensatzes 3.21 zeigen. Dazu drücken wir Y_k als $Y_k = X \circ T^{k-1} \circ Y$ mit

$$\begin{aligned} T : \mathbb{R}^{\mathbb{N}} &\rightarrow \mathbb{R}^{\mathbb{N}}, (y_k)_{\mathbb{N}} \mapsto (y_{k+1})_{\mathbb{N}} \\ X : \mathbb{R}^{\mathbb{N}} &\rightarrow \mathbb{R}, (y_k)_{\mathbb{N}} \mapsto y_1 \end{aligned}$$

aus. Ferner gilt $X \circ T^{k-1} \circ Y \stackrel{\mathcal{D}}{=} X \circ T^{k-1}$ wobei wir die linke Seite als Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ verstehen und die rechte Seite als Zufallsvariable auf $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\otimes \mathbb{N}}, P^Y)$ für $\mathcal{B} := \{\text{reelle Borel-Mengen}\}$. Man sieht dies mit $x \in \mathbb{R}$ wie folgt:

$$\begin{aligned} P^Y(X \circ T^{k-1} \leq x) &= P(Y^{-1}\{y \in \mathbb{R}^{\mathbb{N}} | X \circ T^{k-1}(y) \leq x\}) \\ &= P(\{\omega \in \Omega | X \circ T^{k-1}(Y(\omega)) \leq x\}) \\ &= P(X \circ T^{k-1} \circ Y \leq x) \end{aligned}$$

Folglich können wir den zu betrachtenden Ausdruck in passender Form ausdrücken:

$$\frac{1}{n} \sum_{k=1}^n Y_k = \sum_{k=1}^n X \circ T^{k-1} \circ Y \stackrel{\mathcal{D}}{=} \sum_{k=1}^n X \circ T^{k-1}$$

Die P^Y -Integrierbarkeit von X folgt direkt aus der P -Integrierbarkeit von Y . Es bleibt zu zeigen, dass T ergodisch ist, das heißt zunächst einmal, dass T messbar

und ma erhaltend ist. Dazu erinnern wir zun achst daran, dass der Produkt-Raum der Borel-Algebren $\mathcal{B}^{\otimes \mathbb{N}}$ durch Zylindermengen erzeugt wird:

$$\begin{aligned} \pi_n : \mathbb{R}^{\mathbb{N}} &\rightarrow \mathbb{R}^n, (y_k)_{\mathbb{N}} \mapsto (y_1, \dots, y_n) \\ \Rightarrow \mathcal{B}^{\otimes \mathbb{N}} &= \sigma\left(\bigcup_{n \in \mathbb{N}} \pi_n^{-1}(\mathcal{B}^{\otimes n})\right) \end{aligned}$$

T messbar

Es reicht, Messbarkeit auf Erzeugern zu pr ufen. Betrachte also f ur $B \in \mathcal{B}^{\otimes n}$:

$$\begin{aligned} T^{-1}(\pi_n^{-1}(B)) &= T^{-1}(B \times \mathbb{R}^{\mathbb{N}}) \\ &= \mathbb{R} \times B \times \mathbb{R}^{\mathbb{N}} = \pi_{n+1}^{-1}(\mathbb{R} \times B) \end{aligned}$$

Da $\mathbb{R} \times B \in \mathcal{B}^{\otimes n+1}$ gilt, liegt obiger Ausdruck in $\mathcal{B}^{\otimes \mathbb{N}}$, was zu zeigen war.

T ma erhaltend

Da Wahrscheinlichkeitsma e immer endlich und Zylindermengen schnittstabil sind sowie die aufspannende Folge $\pi_n^{-1}(\mathbb{R}^n) \nearrow \mathbb{R}^{\mathbb{N}}$ beinhalten, reicht es nach 3.22, die Gleichheit $P^Y(A) = P^Y(T^{-1}(A))$ f ur $A = \pi_n^{-1}(B)$ mit $B \in \mathcal{B}^{\otimes n}$ nachzurechnen:

$$\begin{aligned} P^Y(T^{-1}(A)) &= P(Y \in T^{-1}(A)) = P(T \circ Y \in A) \\ &= P(T \circ Y \in B \times \mathbb{R}^{\mathbb{N}}) = P((Y_2, \dots, Y_{n+1}) \in B) \\ &\stackrel{*}{=} P((Y - 1, \dots, Y_n) \in B) = P(Y \in \underbrace{B \times \mathbb{R}^{\mathbb{N}}}_{=A}) \end{aligned}$$

T ergodisch

Sei $A \in \mathcal{B}^{\otimes \mathbb{N}}$ mit $T^{-1}(A) = A$. Ist $A = \emptyset$ so auch $T^{-1}(A) = \emptyset$ also $P^Y(A) = 0$. Ist $A \neq \emptyset$ so gibt es ein Element $(a_k)_{\mathbb{N}} \in A$. Damit erh alt man aber:

$$\begin{aligned} \Rightarrow \forall x_1 \in \mathbb{R} : (x_1, a_1, a_2, \dots) &\in T^{-1}(A) = A \\ \Rightarrow \forall x_2 \in \mathbb{R} : (x_2, x_1, a_1, \dots) &\in T^{-1}(T^{-1}(A)) = A \\ \vdots & \qquad \qquad \qquad \vdots \end{aligned}$$

Insgesamt ergibt sich also $A = \mathbb{R}^{\mathbb{N}}$ und damit $P^Y(A) = 1$.

Damit k nnen wir nun schlie lich den Ergodensatz 3.21 anwenden und es ergibt sich wie behauptet der folgende Ausdruck.

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^n Y_k &\stackrel{\mathcal{D}}{=} \sum_{k=1}^n X \circ T^{k-1} \xrightarrow{f.s.} E_{P^Y}(X) \\ &= \int_{\mathbb{R}^{\mathbb{N}}} X dP^Y \stackrel{*}{=} \int_{\Omega} \underbrace{X \circ Y}_{=Y_1} dP = E_P(Y_1) \end{aligned}$$

┐

Nun k nnen wir die Konsistenz von $\hat{\gamma}(h)$ nachweisen und damit den Abschnitt beenden.

Satz 4.7 (Konsistenz) Für $n \rightarrow \infty$ gilt $\hat{\gamma}(h) \xrightarrow{\mathcal{P}} \gamma(h)$.

┌ Definiere $Y_t := Z_t Z_{t+h}$, dann ist $Y = Y_{\mathbb{N}}$ wegen der starken Stationarität von Z ebenfalls stark stationär. Unter Annahme von Integrierbarkeit erhält man mit 3.23:

$$\begin{aligned}\hat{\gamma}(h) &= \frac{1}{n-h} \sum_{t=1}^{n-h} Z_t Z_{t+h} = \frac{1}{n-h} \sum_{t=1}^{n-h} Y_t \\ &\xrightarrow{f.s.} E(Y_1) = E(Z_1 Z_{1+h}) = \gamma(h)\end{aligned}$$

Fast sichere Konvergenz impliziert Konvergenz in Wahrscheinlichkeit. ┐

4.2 Verteilung

Lit.: [2], [7]. In diesem Abschnitt wollen wir uns noch zusätzlich zur Vorhersage eines konkreten Punktes $\hat{Z}_h$ die Möglichkeit erarbeiten, ganze Bereiche angeben zu können, in denen der vorherzusagende Wert mit einer vorgegebenen Wahrscheinlichkeit liegt (sogenannte *Konfidenzbereiche*). Wie zuvor gehen wir dabei von einer stark stationären und zentrierten Zeitreihe Z mit Beobachtung Z_B aus. Insbesondere gilt also für ein Wahrscheinlichkeitsmaß Q :

$$Z_b \sim Q \quad \forall b \in B$$

Wir suchen also einen Schätzer $\hat{Q}$ für Q basierend auf der Beobachtung Z_B . Da unsere Beobachtung immer endlich ist und wir nichts weiter über Q wissen, ist es ein sinnvoller Ansatz $\hat{Q}$ als Mittel der den Beobachtungen entsprechenden Dirac-Maßen zu definieren.

Definition 4.8 (Verteilungs-Schätzer) Zu einer Beobachtung Z_B definieren wir den Verteilungs-Schätzer $\hat{Q}$ als:

$$\hat{Q} := \frac{1}{|B|} \sum_{b \in B} \delta_{Z_b}$$

Wobei $\delta_z(A) := I\{z \in A\}$ für $z \in \mathbb{R}$ sowie $A \subset \mathbb{R}$ das Dirac-Maß sei.

Im Folgenden wollen wir untersuchen, ob dieser Schätzer ebenfalls gewissen Güteeigenschaften folgt. Die wohl wichtigste von ihnen ist, dass $\hat{Q}$ tatsächlich auch Wahrscheinlichkeitsmaße liefert.

Satz 4.9 (Wohldefiniertheit) $\hat{Q}$ ist Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$.

┌ Zunächst gilt $0 \leq \hat{Q} \leq 1$ da für alle $A \in \mathcal{B}$ offenbar $0 \leq \delta_{Z_b}(A) = I\{Z_b \in A\} \leq 1$ gilt. Ferner ist $\hat{Q}(\emptyset) = 0$ und $\hat{Q}(\mathbb{R}) = 1$ wie man analog wie folgt sieht:

$$\hat{Q}(\emptyset) = \frac{1}{|B|} \sum_{b \in B} \delta_{Z_b}(\emptyset) = \frac{1}{|B|} \sum_{b \in B} \underbrace{I\{Z_b \in \emptyset\}}_{=0} = 0$$

Es bleibt, sigma-Additivität zu zeigen. Seien $A_n \in \mathcal{B}$ paarweise disjunkt. Es gilt:

$$\begin{aligned}\hat{Q}\left(\bigsqcup_{n \in \mathbb{N}} A_n\right) &= \frac{1}{|B|} \sum_{b \in B} \delta_{Z_b}\left(\bigsqcup_{n \in \mathbb{N}} A_n\right) \\ &= \frac{1}{|B|} \sum_{b \in B} I\left\{Z_b \in \bigsqcup_{n \in \mathbb{N}} A_n\right\}\end{aligned}$$

Da die A_n paarweise disjunkt sind, kann Z_b maximal in einem davon liegen, sodass:

$$I\left\{Z_b \in \bigsqcup_{n \in \mathbb{N}} A_n\right\} = \sum_{n \in \mathbb{N}} I\{Z_b \in A_n\} = \sum_{n \in \mathbb{N}} \delta_{Z_b}(A_n)$$

Wobei diese Summe entsprechend in maximal einem Term ungleich Null ist.

$$\begin{aligned}\hat{Q}\left(\bigsqcup_{n \in \mathbb{N}} A_n\right) &= \frac{1}{|B|} \sum_{b \in B} \sum_{n \in \mathbb{N}} \delta_{Z_b}(A_n) \\ &= \sum_{n \in \mathbb{N}} \frac{1}{|B|} \sum_{b \in B} \delta_{Z_b}(A_n) = \sum_{n \in \mathbb{N}} \hat{Q}(A_n)\end{aligned}$$

┐

Da wir nun wissen, dass der Ausdruck $\hat{Q}$ basierend auf einer Beobachtung Z_B in der Tat ein Wahrscheinlichkeitsmaß liefert, können wir uns im Folgenden darum kümmern, dass dieses auch tatsächlich eine gute Schätzung für das eigentliche Q ist.

Satz 4.10 (Erwartungstreue) $\hat{Q}$ ist erwartungstreu, das heißt $E(\hat{Q}) = Q$.

┐ Diese Eigenschaft lässt sich unter Ausnutzung der Linearität des Erwartungswertes einfach nachrechnen. Für ein beliebiges $A \in \mathcal{B}$ erhält man:

$$\begin{aligned}E(\hat{Q}(A)) &= \frac{1}{|B|} \sum_{b \in B} E(\delta_{Z_b}(A)) = \frac{1}{|B|} \sum_{b \in B} E(I\{Z_b \in A\}) \\ &= \frac{1}{|B|} \sum_{b \in B} \underbrace{P(Z_b \in A)}_{= P^{Z_b}(A) = Q(A)} = \frac{|B|}{|B|} Q(A) = Q(A)\end{aligned}$$

Insgesamt sieht man also wie behauptet $E(\hat{Q}) = Q$.

┐

Satz 4.11 (Konsistenz) $\hat{Q}$ ist konsistent, das heißt $\hat{Q} \xrightarrow{P} Q$ für $|B| \rightarrow \infty$.

┐ Definiere $Y_b := \delta_{Z_b}(A)$ mit $A \in \mathcal{B}$. Dann überträgt sich die Stationarität von $Z_{\mathbb{N}}$ auf $Y_{\mathbb{N}}$. Mit Satz 3.23 erhalten wir für $|B| \rightarrow \infty$:

$$\begin{aligned}\hat{Q}(A) &= \frac{1}{|B|} \sum_{b \in B} Y_b \xrightarrow{f.s.} E(Y_{b_1}) \\ &= E(\delta_{Z_{b_1}}(A)) \stackrel{s.o.}{=} Q(A)\end{aligned}$$

Fast sichere Konvergenz impliziert Konvergenz in Wahrscheinlichkeit.

┐

Literaturverzeichnis

- [1] Brockwell, Peter J.; Davis, Richard A. (1991): Time Series: Theory and Methods. 2. Auflage. New York: Springer-Verlag.
- [2] Depperschmidt, Andrej (2022): Skript zur Vorlesung Stochastische Prozesse. Hamburg: Universität.
- [3] Gilchrist, Warren (2010): Forecasting. Sheffield: City Polytechnic.
- [4] Iske, Armin (2018): Approximation. Berlin: Springer-Spektrum.
- [5] Kreiß, Jens-Peter; Neuhaus, Georg (2006): Einführung in die Zeitreihenanalyse. Berlin, Heidelberg: Springer-Verlag.
- [6] Ledolter, Johannes (2010): Prediction and Forecasting. o.A.
- [7] Rüschemdorf, Ludger (2014): Mathematische Statistik. Berlin, Heidelberg: Springer-Verlag.

Datenverzeichnis

- [8] Börse Frankfurt Historische Kurse und Umsätze DAX (<https://www.boerse-frankfurt.de/index/DAX/kurshistorie/historische-kurse-und-umsaetze>)
- [9] Statistisches Bundesamt Datenbank Code 46241-0004. (<https://www-genesis.destatis.de/genesis/online>)
- [10] Umweltbundesamt Nationale Treibhausgas-Inventare (<https://www.umweltbundesamt.de/daten/klima/treibhausgas-emissionen-in-deutschland>)