

Beiträge zur elementaren Zeitreihenanalyse - Teil 2

von *Günter Nordmeier*, Bad Essen

Kurzfassung: In Teil 1, erschienen im letzten Heft von Stochastik in der Schule, wurde die Methode der gleitenden Mittelwerte vorgestellt. Diese dient dazu, aus einer Zeitreihe den Trend herauszurechnen. In diesem Teil wird der Rest eines Beispiel besprochen, speziell wird aus dem Trend eine Prognose für die Zukunft abgeleitet. Danach wird eine andere Methode zur Berechnung des glatten Teils einer Zeitreihe vorgestellt, nämlich die exponentielle Glättung. Diese Methode wird sodann an einem Beispiel illustriert und mit der Methode der gleitenden Mittelwerte verglichen.

3.2 "Lufttemperaturen in NW-Deutschland von 1961-92 - Fortsetzung

Quellen für das verwendete Datenmaterial

Wir erinnern daran, daß hier als Lufttemperatur NW-Deutschlands das einfache arithmetische Mittel der Lufttemperaturen von Bremen (Flughafen), Essen und Hannover-Langenhagen verstanden wird. Als Quellen für die Monatsmittel der Lufttemperaturen (und bei Bedarf auch der monatlichen Niederschlagsmengen) wurden dabei verwendet:

- für Bremen und Essen von Nov. 1964 weg das Statistische Jahrbuch für die Bundesrepublik Deutschland, Jahrgänge 1966 bis 1992,
- für Hannover-Langenhagen ab Nov. 1964: Stat. Jahrbuch BRD,
- für Essen ab 1961: Statistisches Jahrbuch Nordrhein-Westfalen, ab Jahrgang 1962 (Daten sind bis 1950 zurück verfügbar),
- für Bremen und Hannover 1961 bis Okt. 1964: 'Monatlicher Wetterungsbericht', herausgegeben vom Deutschen Wetterdienst
- Dreißig-Jahres-Mittel 1931-1960: Stat. Jahrbuch BRD.

Die Fortsetzung der Zeitreihen in Tabelle 4

Mit Hilfe dieser Quellen konnten die Daten der Tab. 4 bis einschließlich 1991 berechnet werden. Um die beiden Zeitreihen von Tab. 4 in die Gegenwart zu verlängern, hat der Verfasser die Monatsmittel für 1992 zunächst der Reihe 'Klima-Eilinformation' des Deutschen Wetterdienstes entnommen. Mit dem Jahrgang 1993 des Stat. Jahrbuchs BRD erfolgte eine starke Reduzierung der meteorologischen Angaben. Es werden in Zukunft die Daten der Stationen Bremen und Essen nicht mehr dort nachzulesen sein. Da die Werte für Essen voraussichtlich weiterhin dem Stat. Jahrbuch Nordrhein-Westfalen

entnommen werden können, fehlen uns noch die Daten von Bremen, um die Zeitreihen 'Lufttemperaturen in NW-Deutschland' fortsetzen zu können. Aus praktischen Gründen wollen wir hier jedoch an den Statistischen Jahrbüchern als Quelle festhalten. Unter dieser Bedingung gelingt eine zufriedenstellende Fortsetzung, wenn man mit folgenden Schätzwerten arbeitet:

$$\hat{t}(\text{NW-D}) = \frac{2 \cdot t(\text{Hannover}) + t(\text{Essen})}{3}$$

Mit einer Wahrscheinlichkeit von $\approx 90\%$ schätzt man so die Jahresmittel der Temperaturen für NW-Deutschland (im vorhin festgelegten Sinn) auf $\pm 0,1$ [°C] und die Sommermittel der Lufttemperaturen auf $\pm 0,2$ [°C] genau.

Prognosen

Die beiden Hauptanliegen einer Zeitreihenanalyse sind:

1. Die Entwicklung der untersuchten Querschnittsdaten im fraglichen Zeitraum mit mathematischen Mitteln möglichst genau zu beschreiben.
2. Eine über das Ende der vorliegenden Zeitreihe hinausgehende Prognose zu ermöglichen.

Das zweite Anliegen ist in unserem Beispiel hochinteressant. Wer möchte nicht gern im voraus wissen, welche Temperaturen in den kommenden Jahren herrschen? Führen wir jetzt eine solche Prognose durch. Wir legen unser additives Modell der Zeitreihenanalyse mit den drei Komponenten Trend, Zyklus und Rest zugrunde.

Trend: Abb. 7 (Teil 1) zeigt zwei mögliche Varianten der Trendfortsetzung über 1992 hinaus. Beide Varianten lassen sich aus dem bisherigen Trendkurven-Verlauf schließen. Auch dazwischenliegende Varianten sind denkbar. Variante 1 zeigt die höchstmöglichen Trendwerte, während Variante 2 gewissermaßen die an der unteren Grenze liegenden möglichen zukünftigen Trendwerte einfängt. Man könnte das Verfahren noch stärker mathematisieren. Dazu müßte man den Trend als Funktion von t auffassen und eine Näherungsgleichung für die Trendkurve finden. Wir sehen hier davon ab.

Zyklische Komponente: Wir unterstellen, daß die herausgearbeitete 8-jährige Temperaturschwingung weiter anhält. Dies ist eine zusätzliche Annahme!

Restkomponente: Die mittlere Restkomponente betrug von 1961-92 $\approx 0,4$ [K]. In 78% der erfaßten Jahre galt: $|R_t| \leq 0,4$ [K]. Wir können mit einer Wahrscheinlichkeit von 80% sagen, daß die Restkomponente $|R_t| \leq 0,5$ [K] beträgt, wenn sich die klimatischen Verhältnisse in NW-Deutschland nicht plötzlich grundlegend ändern. Für die Prognose setzen wir der Einfachheit halber die

Restkomponente mit $\pm 0,5$ [K] an. Unsere Modellrechnungen stehen damit unter zusätzlichen Annahmen und die Ergebnisse, also unsere Temperaturvorhersagen für 1993 bis 1996, sind außerdem mit einer Wahrscheinlichkeit behaftet. Dies ist im Unterricht sorgfältig herauszuarbeiten. Den Schülern wird deutlich, was alles dazu gehört, wenn man die Wirklichkeit mit einem mathematischen Modell einfangen will.

Tab. 9: Schätzwerte für die mittlere Jahrestemperatur in NW-Deutschland für 1993-96

(1) Jahr	(2) (3) Trend Variante 1 Variante 2 [°C] [°C]		(4) Zyklische Komponente [K]	(5) Rest mit P = 80% [K]	(6) (7) Schätzintervalle Variante 1 Variante 2 [°C] [°C]	
1993	10,0	9,9	0,0	$\pm 0,5$	9,5-10,5	9,4-10,4
1994	10,1	9,8	-0,6	$\pm 0,5$	9,0-10,0	8,7-9,7
1995	10,2	9,7	-0,7	$\pm 0,5$	9,0-10,0	8,5-9,5
1996	10,3	9,6	-0,1	$\pm 0,5$	9,7-10,7	9,0-10,0

Eine über das Jahr 1996 hinausreichende Schätzung ist nicht möglich, weil die beiden Trendvarianten dann zu sehr auseinandergehen.

4. Exponentielles Glätten einer Zeitreihe

4.1 Einführungsbeispiel

Das folgende Beispiel sollte vom 10. Schuljahrgang an geeignet sein. Zur Trendbestimmung wollen wir nun ein völlig anderes Verfahren heranziehen, nämlich das exponentielle Glätten. Um den Unterschied auch formal hervorzuheben, benutzen wir jetzt zur Bezeichnung der Trendwerte die Variable \bar{y}_t . Wir zeigen das neue Verfahren am Beispiel der mittleren Jahres-Lufttemperaturen von England und Wales (vgl. Tab.3 in Teil 1 bzw. Tab. 10 unten).

Man setzt:

$$\bar{y}_1 = y_1 = y(1981) = 9,74 \text{ [°C]}$$

Den 2. Trendwert berechnet man so:

$$\bar{y}_2 = \alpha \cdot y_2 + (1 - \alpha) \cdot \bar{y}_1$$

Für den sogenannten Glättungsparameter α muß man einen Zahlenwert zwischen 0 und 1 wählen. Welchen Einfluß diese Wahl auf die Glättung hat, werden wir im Abschnitt 4.3 zeigen. Für unser Beispiel wählen wir $\alpha=0,3$.

Es ergibt sich:

$$\bar{y}_2 = 0,3 \cdot 10,36 + 0,7 \cdot 9,74 = 9,93$$

Den 3. Trendwert berechnet man so:

$$\bar{y}_3 = \alpha \cdot y_3 + (1-\alpha) \cdot \bar{y}_2 \quad \text{usw.}$$

Im Beispiel ergibt sich:

$$\bar{y}_3 = 0,3 \cdot 10,46 + 0,7 \cdot 9,93 = 10,09$$

Der Grundgedanke des exponentiellen Glättens ist nun deutlich geworden. Man wichtet den aktuellen Zeitreihenwert y_t stets mit der Glättungskonstanten α und den direkt vorhergehenden Trendwert mit $(1-\alpha)$:

$$\bar{y}_t = \alpha \cdot y_t + (1-\alpha) \cdot \bar{y}_{t-1}$$

Tab. 10: Lufttemperaturen in England und Wales

t	y_t	\bar{y}_t
1981	9,74	9,74
1982	10,36	9,93
1983	10,46	10,09
1984	9,99	10,06
1985	9,23	9,81
1986	9,13	9,61
1987	9,48	9,57
1988	10,19	9,75
1989	10,93	10,11

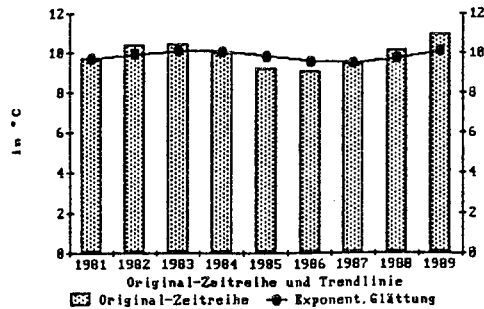


Abb. 8: Lufttemperaturen - exponentieller Trend mit $\alpha=0,3$

4.2 Zum Namen des Verfahrens

Wie wir soeben dargestellt haben, gilt beim exponentiellen Glätten:

$$\bar{y}_1 = y_1 \quad \bar{y}_2 = \alpha \cdot y_2 + (1-\alpha) \cdot \bar{y}_1 \quad \bar{y}_3 = \alpha \cdot y_3 + (1-\alpha) \cdot \bar{y}_2 \quad \text{usw.}$$

Durch Einsetzen erhält man:

$$\bar{y}_2 = \alpha \cdot y_2 + (1-\alpha) \cdot y_1$$

$$\begin{aligned} \bar{y}_3 &= \alpha \cdot y_3 + (1-\alpha) \cdot [\alpha \cdot y_2 + (1-\alpha) \cdot y_1] \\ &= \alpha \cdot y_3 + \alpha(1-\alpha) \cdot y_2 + (1-\alpha)^2 \cdot y_1 \end{aligned}$$

$$\begin{aligned} \bar{y}_4 &= \alpha \cdot y_4 + (1-\alpha) \cdot [\alpha \cdot y_3 + \alpha(1-\alpha) \cdot y_2 + (1-\alpha)^2 \cdot y_1] \\ &= \alpha \cdot y_4 + \alpha(1-\alpha) \cdot y_3 + \alpha(1-\alpha)^2 \cdot y_2 + (1-\alpha)^3 \cdot y_1 \\ &\quad \text{usw.} \end{aligned}$$

Allgemein:

$$\bar{y}_t = \alpha \cdot y_t + \alpha(1-\alpha) \cdot y_{t-1} + \alpha(1-\alpha)^2 \cdot y_{t-2} + \dots + (1-\alpha)^{t-1} \cdot y_1$$

Jetzt wird deutlich, warum das Verfahren "exponentielles Glätten heißt: Die Werte der Zeitreihe werden rückwärtsschreitend mit $(1-\alpha)$ exponentiell abnehmend gewichtet. Es ist also möglich, jeden Trendwert direkt aus den Werten der Zeitreihe zu berechnen. In unserem Beispiel ergibt sich:

$$\begin{aligned} \bar{y}(1989) &= \left\{ \begin{array}{l} 0,3 \cdot 10,93 \\ + 0,3 \cdot 0,7 \cdot 10,19 \\ + 0,3 \cdot 0,7^2 \cdot 9,48 \\ + 0,3 \cdot 0,7^3 \cdot 9,13 \\ + 0,3 \cdot 0,7^4 \cdot 9,23 \\ + 0,3 \cdot 0,7^5 \cdot 9,99 \\ + 0,3 \cdot 0,7^6 \cdot 10,46 \\ + 0,3 \cdot 0,7^7 \cdot 10,36 \\ + 0,7^8 \cdot 9,74 \end{array} \right. = \left\{ \begin{array}{l} 0,3 \cdot 10,93 \\ + 0,2100 \cdot 10,19 \\ + 0,1470 \cdot 9,48 \\ + 0,1029 \cdot 9,13 \\ + 0,0720 \cdot 9,23 \\ + 0,0504 \cdot 9,99 \\ + 0,0353 \cdot 10,46 \\ + 0,0247 \cdot 10,36 \\ + 0,0576 \cdot 9,74 \end{array} \right. \approx 10,11 \end{aligned}$$

Bei der Betrachtung der letzten allgemeinen Gleichung für die Trendwerte und dieses Rechenbeispiels sehen die Schüler:

1. Beim exponentiellen Glätten gehen alle bisherigen Zeitreihenwerte in die Berechnung eines Trendwerts mit ein.
2. Die Gewichtung der Daten nimmt rückwärtsschreitend exponentiell ab bzw. vorwärtsschreitend exponentiell zu.
3. Beim exponentiellen Glätten fehlen keine unteren und oberen Randwerte.

4.3 Der Einfluß des Glättungsparameters

Um etwas über den Einfluß des Glättungsparameters α zu erfahren, gehen wir zunächst experimentell vor. Dazu glätten wir unsere ausgewählte Zeitreihe der Lufttemperaturen von England und Wales mit $\alpha=0,3$ bzw. $0,5$ und $0,7$. Das Ergebnis ist in Abb. 9 zu sehen. Dies führt uns zu der Vermutung, daß größere Parameter die in einer Zeitreihe enthaltenen

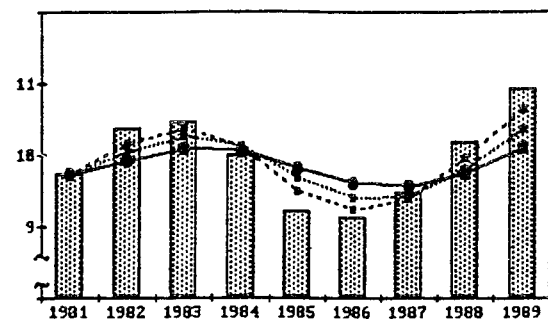


Abb. 9: Lufttemperaturen in England und Wales - Exponentielles Glätten:
— $\alpha = 0,3$ - - - $\alpha = 0,5$ - - - $\alpha = 0,7$

Unterschiede auch mehr hervorheben. Die Trendlinie für $\alpha=0,7$ schwingt am stärksten, die Trendlinie für $\alpha=0,3$ ist am "ruhigsten". Daß dies so sein muß, können Schüler aus den Formeln für die Glättung schließen. Das Unterrichtsgespräch über diese Phänomene sollte etwa zu folgenden Ergebnissen führen:

1. Je stärker die jüngeren Zeitreihenwerte berücksichtigt werden sollen, desto größer muß α gewählt werden.
2. Je stärker die zurückliegenden Werte der Zeitreihe berücksichtigt werden sollen, desto kleiner muß α gewählt werden.
3. Je größer α gewählt wird, desto stärker wird die Anpassung an neue Entwicklungen (an Tendenzänderungen) vollzogen.
4. Je kleiner α gewählt wird, desto stärker wird die Reihe geglättet.
5. Beim Glätten mit $\alpha=1$ wird die Zeitreihe selbst reproduziert. Beim Glätten mit $\alpha=0$ entsteht eine Parallele zur Zeitachse durch den Anfangswert der Reihe. Diese beiden Extreme dienen nicht zur Trendbestimmung, weshalb man α zwischen 0 und 1 wählt.

Es hängt vom jeweiligen Untersuchungsziel ab, welchen Wert man α gibt. In unserem Beispiel wird man α klein wählen, weil es einen langfristigen Trend aufzuspüren gilt. Abb. 9 zeigt übrigens recht deutlich, daß die Zeitreihe von Kimber zu kurz ist, um einen langfristigen Trend - etwa im Sinne einer Klimaänderung - auszumachen. Wir sehen wellenförmige Linien vor uns, die nichts darüber aussagen, wie es nach 1989 weitergeht.

Anregung: Untersuchen Sie die in Tab. 4 enthaltenen Zeitreihe der Lufttemperaturen in Nordwestdeutschland mit Hilfe der exponentiellen Glättung.

4.4 Vergleich verschiedener Glättungsmethoden

Alle bisher dargestellten Glättungsmethoden gehören zu den sogenannten Tiefpaßfilterungen: Hohe Frequenzen bzw. Zyklen/Schwingungen von kurzer Länge werden unterdrückt. Niedrige Frequenzen bzw. lange Zyklen/Schwingungen werden hervorgehoben. Dadurch wird der Trend sichtbar. Eine gute Einführung in die Praxis der Tiefpaßfilterungen gibt Schönwiese (1992). Die Methode der gleitenden Mittelwerte läßt sich zur Methode der gewichteten gleitenden Mittelwerte verallgemeinern (vgl. Abschnitt 3 in Teil 1). Auch die Methode der exponentiellen Glättung läßt sich erweitern, z.B. auf verschiedene Komponenten ausdehnen (Holt-Winters-Verfahren).

Das Verfahren des exponentiellen Glättens ist wegen seiner vielfältigen Erweiterungsmöglichkeiten im Grundsatz der gleitenden Mittelwertbildung überlegen. In Abb. 10 sieht man u.a. zwei Trendlinien. Die durchgehende dunkle Linie entstand durch exponentielles Glätten mit $\alpha=0,3$, während die gestrichelte Linie ein dreigliedriges gleitendes Mittel darstellt, das in den Vordaten verankert wurde. Als punktierte Linie wurde noch die mittlere Lufttemperatur der Jahre 1951-80 eingezeichnet. Damit findet man drei Linien in einer Abbildung, die zu vergleichen und interpretieren sind.

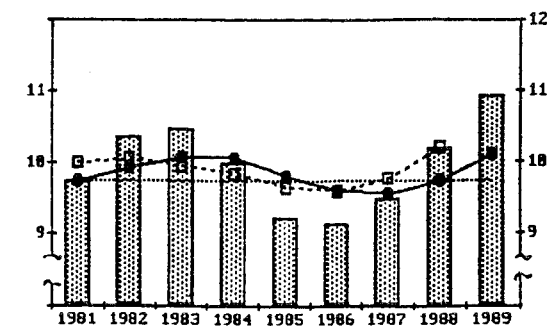


Abb. 10: Lufttemperaturen in England und Wales - Vergleich von Glättungsmethoden:
— exponentiell - - - Gleit. Mittel

Abschließend sei darauf hingewiesen, daß die Zeitreihe der Lufttemperaturen von England und Wales und die von Nordwestdeutschland im Zeitraum 1981-89 einen auffallend ähnlichen Verlauf aufweisen. Dieses Phänomen zu untersuchen und Methoden zu entwickeln, wie man zwei Zeitreihen "verbinden" kann, könnte ein interessantes Vorhaben sein.

Literatur

Siehe Teil 1 in *Stochastik in der Schule* 13(1993), Heft 3.