

Wavelets und ihre Anwendung in der Statistik

Michael Schomaker

3. Mai 2005

Inhaltsverzeichnis

1	Einführung in Wavelets	2
1.1	Fourier-Reihen	2
1.2	Wavelets	4
1.2.1	Das Haar System	4
1.2.2	Die Wavelet-Darstellung	5
1.2.3	Glattere Wavelets	8
2	Statistische Grundlagen	10
2.1	Kerndichteschätzer	10
2.2	Nonparametrische Regression	10
2.2.1	Kernel Regression	11
2.2.2	Schätzung über orthonomale Reihen	12
3	Wichtige statistische Anwendungen von Wavelets	14
3.1	Nonparametrische Regression	14

Kapitel 1

Einführung in Wavelets

1.1 Fourier-Reihen

Der französische Mathematiker und Physiker Jean-Baptiste Fourier fand zu Beginn des neunzehnten Jahrhunderts heraus, daß jede quadratisch integrierbare Funktion beliebig genau durch eine Summe von Sinus- und Cosinus-Funktionen approximiert werden kann:

$$f(x) = \frac{1}{2}a_0 + \sum_{j=1}^J (a_j \cos(jx) + b_j \sin(jx))$$

Je mehr Funktionsterme hinzugefügt werden (also je größer J ist), desto genauer ist die Approximation. Je größer j , desto größer die Frequenz der

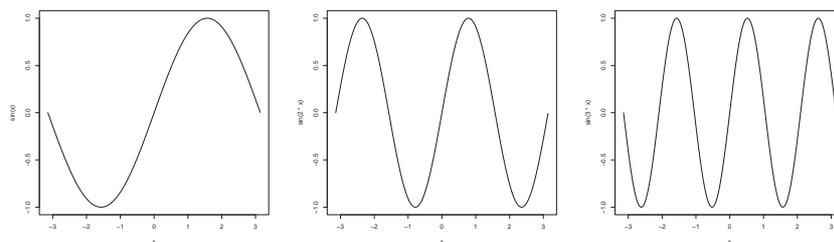


Abbildung 1.1: Die Funktionen $\sin x$, $\sin(2x)$ und $\sin(3x)$

einzelnen Sinus- bzw. Cosinusschwingungen. In Abb.(1.1) sind Sinusschwingungen verschiedener Frequenz dargestellt.

Man sagt, daß eine Funktion dem quadratintegralen Raum $L^2[a, b]$ angehört, genau dann, wenn $\int_a^b f^2(x)dx < \infty$. Fouriers Aussage gilt für alle Funktionen

$f \in L^2[-\pi, \pi]$.

Eine wichtige Eigenschaft im Hinblick auf die Fourier-Reihe ist, daß die Menge der ganzzahligen Sinus- und Cosinusschwingungen eine orthogonale Basis bezüglich des $L^2[-\pi, \pi]$ bildet.

Orthogonal bedeutet, daß das innere Produkt $\langle f, g \rangle = \int f(x)g(x) dx$ zweier Funktionen $f, g \in L^2[a, b]$ Null ergibt.

Die folgende Abb.(1.2) soll zeigen, daß eine Erhöhung der Summationsgren-

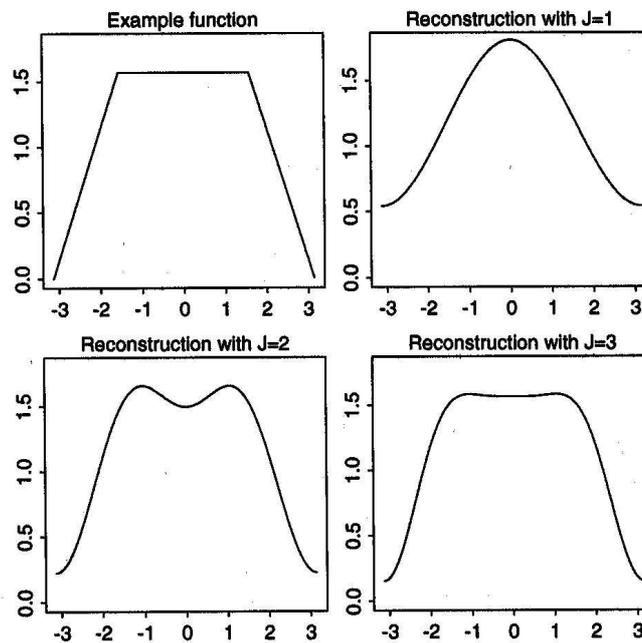


Abbildung 1.2: Beispielfunktion(o.l.), und Approximationen mit verschiedenem Summationsindex

ze J mehr Funktionsterme hinzufügt, und damit die Genauigkeit verbessert.

Des weiteren sei noch erwähnt, daß sich die Fourier-Koeffizienten wie folgt berechnen:

$$a_j = \frac{1}{\pi} \langle f, \cos(j \cdot) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(jx) dx, \quad j = 0, 1, \dots$$

$$b_j = \frac{1}{\pi} \langle f, \sin(j \cdot) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(jx) dx, \quad j = 0, 1, \dots$$

1.2 Wavelets

Wie man im vorherigen Kapitel gesehen hat, können bestimmte Funktionen sehr einfach mit Hilfe der Fourier-Reihe approximiert werden. Im folgenden Kapitel werden nun andere Methoden aufgezeigt, mit denen Funktionen teilweise besser und schneller approximiert werden können. Wesentlicher Unterschied ist, daß die globale Betrachtung nun einer lokalen weicht. Als Einstieg hierzu dient vor allem die Haar-Funktion, an deren Beispiel die Ideen von Wavelets erläutert werden, bevor dann im weiteren noch andere relevante Funktionen präsentiert werden sollen.

1.2.1 Das Haar System

Die *Haar-Funktion*

$$\psi(x) = \begin{cases} 1, & 0 \leq x < \frac{1}{2} \\ -1, & \frac{1}{2} \leq x < 1 \\ 0, & \text{sonst} \end{cases}$$

die am Anfang nicht sonderlich spektakulär erscheinen mag (siehe Abb.(1.3)), wird (wie auch andere Funktionen, die noch später erwähnt werden) als ein *mother wavelet* bezeichnet. Aus ihr entsteht eine "Familie" von Funktionen, die aus gestreckten und verschobenen Haar-Funktionen besteht:

$$\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k)$$

Eine Erhöhung des Streckungsindex j bewirkt hierbei ein "Zusammenpres-

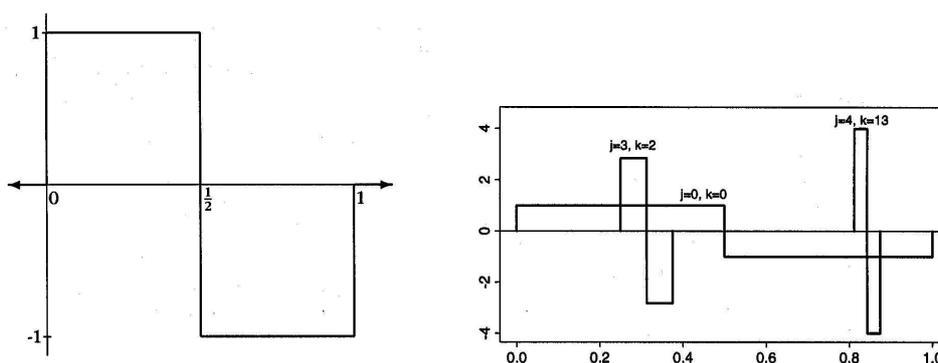


Abbildung 1.3: Haarfunktion (links), gestreckte und verschobene Haarfunktionen (rechts)

sen“ der Funktion, eine Erhöhung des Verschiebungsindex k eine Verschiebung entlang der x -Achse. Abb.(1.3) soll dies auch noch einmal verdeutlichen.

Wichtige Eigenschaften und Folgerungen:

- Die Menge $\{\psi_{j,k}, j, k \in \mathbb{Z}\}$ bildet ein vollständiges Orthonomalsystem im $L^2(\mathbb{R})$

Eine Menge von Funktionen $\{f_j\}$ wird dann als vollständiges Orthonomalsystem bezeichnet, wenn die einzelnen Funktionen zueinander paarweise orthogonal sind, $\|f_j\| = 1$ für jedes j und die einzige Funktion die zu allen Funktionen f_j orthogonal ist, die “Nullfunktion“ ist.

- $\int_{-\infty}^{\infty} \psi_{j,k}(x) dx = 0$
- Sei f^{J_0} eine stückweise konstante Funktion auf Intervallen der Länge $\frac{1}{2^{J_0}}$. Es ist möglich f^{J_0} als die Summe einer *Approximationsfunktion* (die konstant auf Intervallen der doppelten Größe ist) und einer *Detailfunktion* (die die Ungenauigkeit zwischen Approximation und wahrer Funktion widerspiegelt) darzustellen:

$$f^{J_0} = f^{J_0-1} + g^{J_0-1}$$

Man kann dies natürlich sukzessiv weiterführen, indem man dann f^{J_0-1} aufsplittet in $f^{J_0-2} + g^{J_0-2}$, dann f^{J_0-2} aufsplittet, usw. Führt man dies bis zur größten Approximationsfunktion durch, so läßt sich f_{J_0} durch eine Summe von Detailfunktionen und einer sehr groben Approximationsfunktion darstellen.

Da die Abweichung zwischen wahrer Funktion und (nur) der Summe aus den Detailfunktionen beliebig minimiert werden kann, und man die Detailfunktionen als eine Summe von verschobenen und gestreckten Haar-Funktionen darstellen kann, läßt sich folgern, daß jede Funktion $f \in L^2(\mathbb{R})$ beliebig genau durch eine Linearkombination von verschobenen und gestreckten Haarfunktionen $\psi_{j,k}$ dargestellt werden kann.

1.2.2 Die Wavelet-Darstellung

Im letzten Kapitel wurde erörtert, daß sich eine Funktion $f \in L^2(\mathbb{R})$ durch eine Summe von verschobenen und gestreckten Haar-Funktionen darstellen läßt. Dies soll in diesem Kapitel verallgemeinert werden. Ziel ist es, eine Grundlage zu finden, die ermöglicht, daß Funktionen durch *beliebige* Wavelets (also

nicht zwingend durch die Haar-Funktion) approximiert werden können.

Es sei

$$V_j = \{f \in L^2(\mathbb{R}) : f \text{ ist stückweise konstant auf } [k2^{-j}, (k+1)2^{-j}), k \in \mathbb{Z}\}$$

Dann stellt die Menge der Räume $(V_j)_{j \in \mathbb{Z}}$ eine "Leiter" von Unterräumen dar, bei der mit Erhöhung des Indizes j um 1 die Intervalllänge halbiert wird. Die Räume $(V_j)_{j \in \mathbb{Z}}$ besitzen folgende Eigenschaften:

1. $\dots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \dots$;
2. $\bigcap_{j \in \mathbb{Z}} V_j = 0, \quad \overline{\bigcup_{j \in \mathbb{Z}} V_j} = L^2(\mathbb{R})$;
3. $f \in V_j$ nur wenn $f(2 \cdot) \in V_{j+1}$
4. $f \in V_j$ impliziert $f(\cdot - k) \in V_0$ für alle $k \in \mathbb{Z}$

Wenn Sequenzen von Räumen den obigen vier Bedingungen genügen und eine Funktion $\phi \in V_0$ existiert, deren Menge $\{\phi_{0,k} = \phi(\cdot - k), k \in \mathbb{Z}\}$ eine orthonomale Basis für V_0 bildet, dann existiert auch eine Funktion ψ , so daß gilt:

$$P^j f = P^{j-1} f + \sum_{k \in \mathbb{Z}} \langle f, \psi_{j-1,k} \rangle \psi_{j-1,k}$$

$P^j f$ bezeichnet dabei die Abbildung der Funktion f auf den Raum V_j (die "beste" Approximation von f im V_j). ϕ wird dabei auch als *Skalierungsfunktion* oder *father wavelet* bezeichnet. Man erkennt, daß obige Formel nichts anderes als eine Verallgemeinerung des letzten Kapitels ist. Eine Funktion wird durch eine grobe Approximation und eine Summe von speziellen Funktionen (Wavelets) beschrieben.

Entscheidend ist nun, daß die Approximationsfunktion durch eine Linearkombination von *father Wavelets* ϕ dargestellt werden kann, während die Detailfunktion (wie schon gesehen) durch eine Summe von verschobenen und gestreckten *mother Wavelets* ψ repräsentiert wird.

Es ergibt sich also im wesentlichen:

$$P^j f = \sum_{k \in \mathbb{Z}} c_{j,k} \phi_{j-1,k} + \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j-1,k}$$

Mother Wavelets und *father Wavelets* sind zueinander orthogonal. Man beachte hierbei die Analogie zu den Fourier Reihen: Die Basen Cosinus und Sinus sind dort auch zueinander orthogonal. Doch dies ist nicht der einzige Punkt indem sich die Ähnlichkeit zu den Fourier-Reihen sehen läßt. Die Berechnung der Skalierungsfunktionskoeffizienten ($c_{j,k}$) und der Waveletkoeffizienten ($d_{j,k}$) (also die Gewichtung der einzelnen mother- und father-Wavelets) ergibt sich auf ähnliche Weise wie bei den Fourierkoeffizienten:

$$c_{j,k} = \langle f, \phi_{j,k} \rangle = \int f(x) \phi_{j,k}(x) dx$$

$$d_{j,k} = \langle f, \psi_{j,k} \rangle = \int f(x) \psi_{j,k}(x) dx$$

Abb.(1.4) soll noch einmal verdeutlichen, wie eine Funktion mit Hilfe von

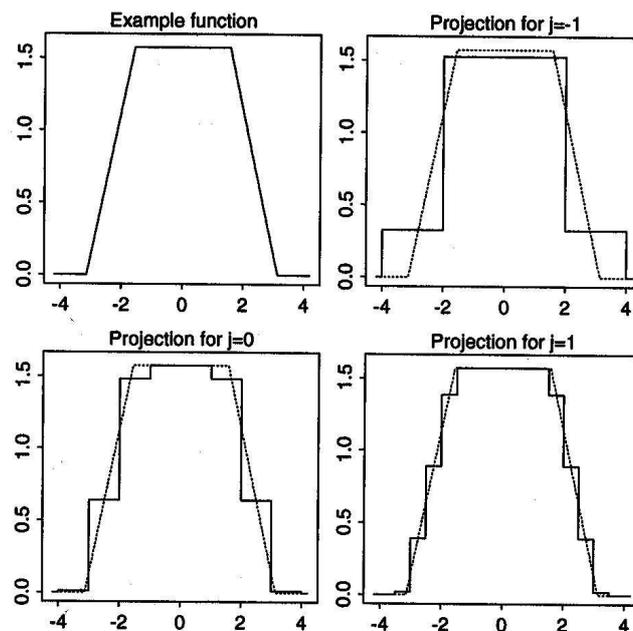


Abbildung 1.4: Beispielfunktion und Approximationen mit Hilfe der Haar-funktion für verschiedene Intervalllängen

Wavelets approximiert werden kann.

1.2.3 Glattere Wavelets

Alle Beispiele und Graphiken wurden bisher von der Haarfunktion unterstützt. Dies ist zwar ein schöner Einstieg in solch ein Thema, aber in der Praxis empfiehlt es sich andere, glattere Wavelets zu benutzen. Natürlich gibt es sehr viele davon. Als Beispiel für weitere Wavelets sollen hier Funktionen der "Daubechies Familie" dienen. Abb.(1.5) stellt verschiedene Skalierungsfunktionen ϕ und verschiedene Wavelets ψ dar. In Abb.(1.6) wird noch einmal aufgezeigt, wie eine Approximation mit der "Deubechies Familie" aussehen kann.

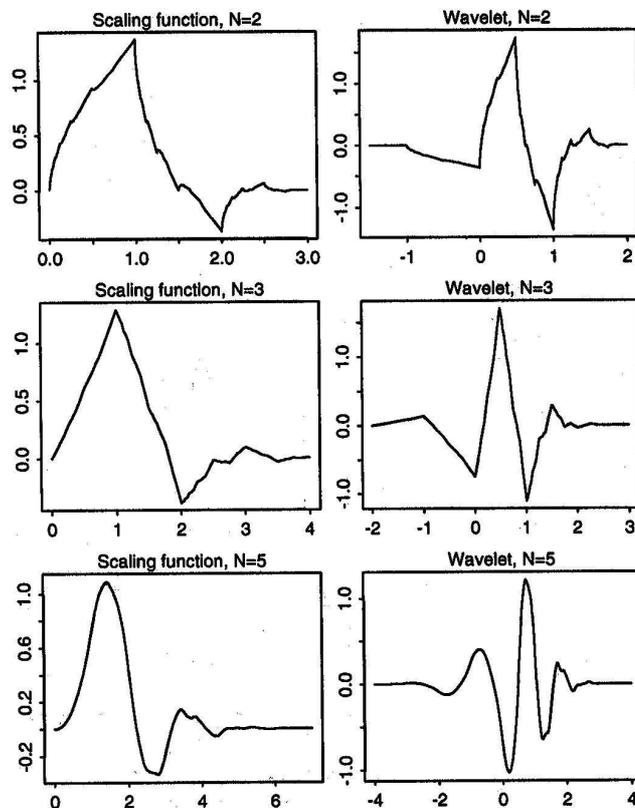


Abbildung 1.5: Wavelets und Skalierungsfunktionen in der "Daubechies Familie"

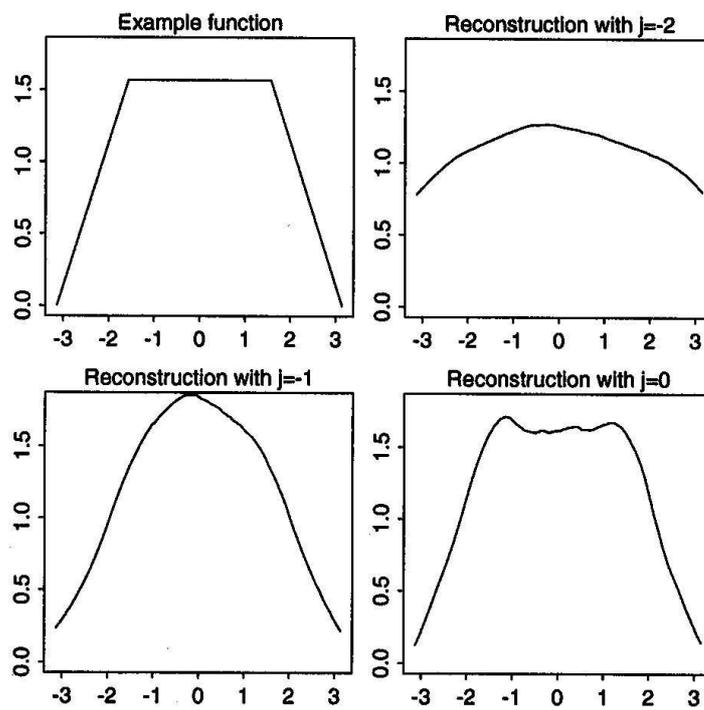


Abbildung 1.6: Approximation einer Funktion mit Hilfe von Wavelets aus der "Daubechies Familie"

Kapitel 2

Statistische Grundlagen

2.1 Kerndichteschätzer

In der explorativen Datenanalyse interessiert man sich oft für die Gestalt einer Verteilung. Hat man nun Daten, und möchte etwas über die Form der Verteilung aussagen, empfiehlt sich natürlich zunächst ein Histogramm zu zeichnen. Nachteil bei einer solchen Darstellung ist jedoch, daß die Form der Verteilung durchaus stark von der Wahl der Intervallbreite abhängen *kann*. Liegt eine Beobachtung x_i beispielsweise am linken Rand eines Intervalls $[b, c]$, und eine weitere Beobachtung x_{i-1} am rechten Rand des vorhergehenden Intervalls $[a, b]$, so wird diese Beobachtung überhaupt nicht in die "Wertung" des Intervalls $[b, c]$ aufgenommen, obwohl sie eventuell viel näher am Datenpunkt x_i liegt als eine weitere Beobachtung x_{i+1} am rechten Ende des Intervalls $[b, c]$. Des weiteren möchte man natürlich oft mit glatten Dichten rechnen. Um diesem Problem gerecht zu werden, kann man *Kerndichteschätzer* benutzen, die von folgender Gestalt sind:

$$\hat{f}(x) = \frac{1}{n\lambda} \sum_{i=1}^n K\left(\frac{x - X_i}{\lambda}\right)$$

Das heißt, abhängig von einer Kernfunktion K , werden Datenpunkte die näher an der Beobachtung liegen stärker gewichtet. In Abb.(2.1) sind verschiedene solcher Kernfunktionen dargestellt.

2.2 Nonparametrische Regression

Die nonparametrische Regression unterscheidet sich von der parametrischen im wesentlichen dadurch, daß der Regressionskurve keine bestimmte Struktur

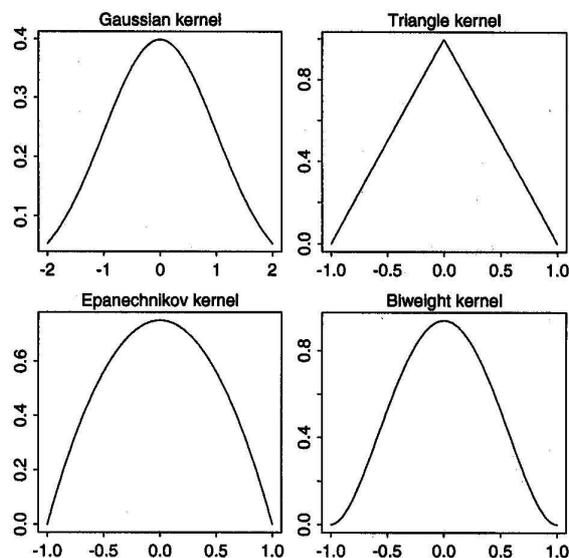


Abbildung 2.1: Verschiedene Kernfunktionen

(z.B. eine lineare Struktur) unterstellt wird. Jede ‘‘Technik‘‘ im Bereich der nonparametrischen Regression unterliegt anderen Annahmen, die grundsatzlich aber schwacher sind als bei der parametrischen Regression.

In den folgenden Abschnitten wird, ohne dadurch die Allgemeinheit zu beschranken, davon ausgegangen, da die Datenpunkte x_1, \dots, x_n aquidistant sind.

2.2.1 Kernel Regression

Als erstes soll in diesem Abschnitt kurz der ‘‘naive‘‘ Schatzer fur die Regressionskurve betrachtet werden:

$$\tilde{f}(u) = \begin{cases} Y_i, & \frac{i-1}{n} \leq u < \frac{i}{n}, \quad i = 1, \dots, n \\ 0, & u = 1 \end{cases}$$

Dieser macht nichts anderes, als die Datenpunkte per Treppenfunktion zu verbinden (siehe Abb.(2.2)). Um aber eine endgultige, glatte Kurve zu erhalten reicht dieser Schatzer naturlich nicht aus. Man kann -in Analogie zum Kerndichteschatzer- dieses Problem so angehen, da man Datenpunkte, die naher an einem Punkt liegen, starker gewichtet als diejenigen, die weiter ent-

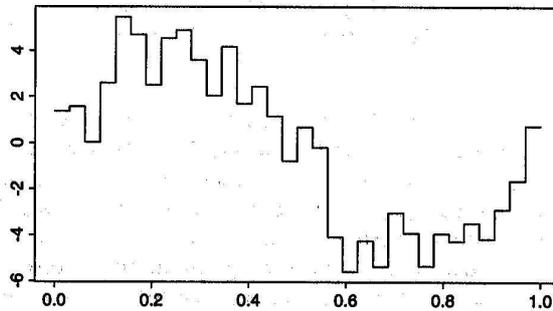


Abbildung 2.2: Naiver Schätzer der Regressionsfunktion

fernt davon liegen. Ein neuer, besserer Schätzer kann also wie folgt aussehen:

$$\hat{f}(u) = \int_0^1 \frac{1}{\lambda} K\left(\frac{u-v}{\lambda}\right) \tilde{f}(v) dv$$

Erneut hängt die geschätzte Funktion dann natürlich vor allem von der Wahl der Kernfunktion und der Wahl der Bandbreite λ ab. Abb.(2.3) soll noch einmal verdeutlichen, wie sehr die Bandbreite die Schätzungsfunktion beeinflussen kann.

2.2.2 Schätzung über orthonomale Reihen

Eine weitere Möglichkeit um die Regressionsfunktion zu schätzen, besteht darin, die Funktion über die Fourier-Reihe oder Polynome, die ein vollständiges Orthonomalsystem darstellen (siehe Kap(1.2.1.)) anzunähern. Dies würde also heißen, daß sich die Regressionsfunktion unter anderem auch wie folgt schätzen ließe:

$$\hat{f}_J(x) = \frac{1}{2} \hat{a}_0 + \sum_{j=1}^J (\hat{a}_j \cos(jx) + \hat{b}_j \sin(jx))$$

In diesem Fall berechnen sich die Koeffizienten \hat{a}_j und \hat{b}_j wie folgt:

$$\hat{a}_j = \langle \tilde{f}, \cos(j \cdot) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \tilde{f}(x) \cos(jx) dx, \quad j = 0, 1, \dots$$

$$\hat{b}_j = \langle \tilde{f}, \sin(j \cdot) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \tilde{f}(x) \sin(jx) dx, \quad j = 1, 2, \dots$$

Abb.(2.4) zeigt, wie die Wahl des Summationsindex J das Ergebnis beeinflussen kann.

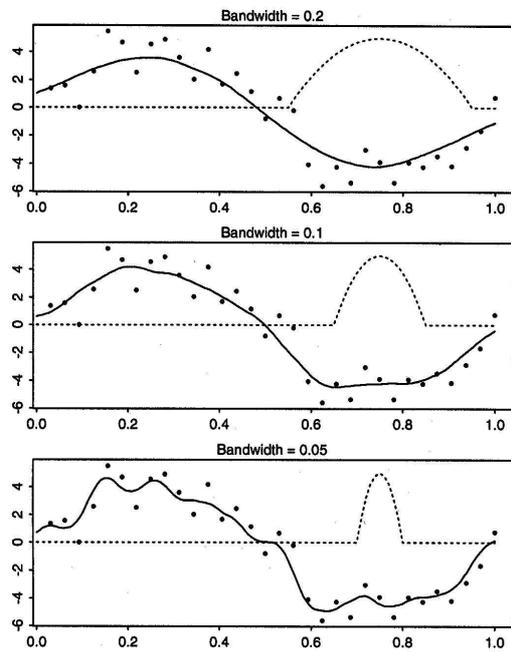


Abbildung 2.3: Regressionsfunktion bei verschiedenen Bandbreiten

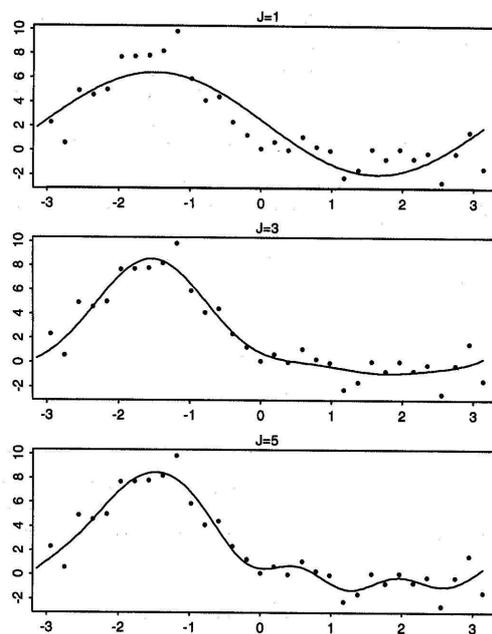


Abbildung 2.4: Regressionsfunktion bei verschiedenem Summationsindex J

Kapitel 3

Wichtige statistische Anwendungen von Wavelets

In diesem Kapitel soll kurz umrissen werden, wie Wavelets mit statistischen Anwendungen verknüpft werden können. Es ist also nichts weiter als eine Verknüpfung der ersten beiden Kapitel.

3.1 Nonparametrische Regression

Erneut soll, ohne Beschränkung der Allgemeinheit, angenommen werden, daß die Datenpunkte x_1, \dots, x_n äquidistant sind.

Die Abbildung von f auf den Approximationsraum kann wie folgt ausgedrückt werden:

$$(P^J f)(u) = \sum_k c_{J,k} \phi_{J,k}(u)$$

Die Funktion f wird also über eine gewichtete Summe von Skalierungsfunktionen approximiert. Hat man kein komplettes Wissen über die Funktion, so muß der Koeffizient $c_{J,k}$ natürlich geschätzt werden:

$$\hat{f}(u) = \sum_k \hat{c}_{J,k} \phi_{J,k}(u)$$

$$\hat{c}_{J,k} = \langle f^\sim, \phi_{J,k} \rangle = \int_0^1 f^\sim \phi_{J,k}(u) du$$

Eine Kombination der letzten beiden Gleichungen ergibt dann ineinander

eingesetzt und vereinfacht folgenden *wavelet-basierten Schätzer der Regressionsfunktion*:

$$\hat{f}(u) = \sum_{i=1}^n Y_i \int_{(i-1)/n}^{i/n} E_J(u, v) dv$$

wobei $E_J(u, v)$ wie folgt definiert ist:

$$E_J(u, v) = \sum_k \phi_{J,k}(u) \phi_{J,k}(v)$$

Die Regressionskurve läßt sich also über eine durch Skalierungsfunktionen gewichtete Funktion ausdrücken.

Wie genau die approximierte Funktion dann aussieht hängt natürlich von der Wahl des Summationsindex J und der Wahl der Skalierungsfunktion (und damit der Wavelet-Familie) ab. Abb.(3.1) gibt hierzu noch einen kleinen Einblick:

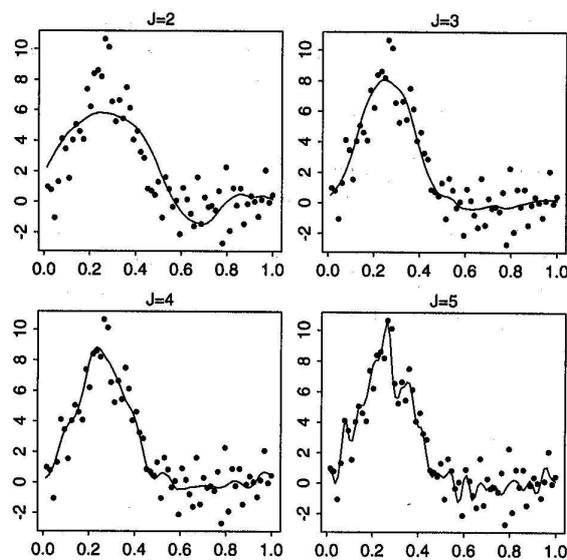


Abbildung 3.1: Wavelet-basierte Schätzungen der Regressionskurve für verschiedene J

Abschließende Bemerkung

Wie man gesehen hat, ermöglichen Wavelets generell eine schnelle Approximation von quadratintegrierbaren (nicht unbedingt stetigen) Funktionen.

Die Anwendungen von Wavelets sind daher reichlich. In der Statistik beschränken sie sich vornehmlich auf den Bereich der Dichteschätzung (der hier nicht näher betrachtet wurde) und auf den Bereich der Regression.

Weitere Anwendungen außerhalb der Statistik gibt es häufig. Hier eine kleine Auswahl: In der Mathematik zur numerischen Lösung partieller Differentialgleichungen, im Bereich der Signal- und Bildbearbeitung und in der Biometrik.

Literatur

R. Todd Ogden, *Essential Wavelets for Statistical Applications and Data Analysis*, Birkhäuser Verlag 1997