Glättungsverfahren in der Biometrie

Ein historischer Abriss

Smoothing methods in biometry

A historic review

• Michael G. Schimek[1]

Vor rund 25 Jahren haben in Deutschland nichtparametrische Glättungsverfahren in die Statistik und etwas zeitverzögert auch in die Biometrie Eingang gefunden. In den frühen 1980er Jahren setzte ein regelrechter Boom in der theoretischen und bald auch rechenintensiven Statistik (engl. 'computational statistics') ein. Im Vordergrund standen univariate nichtparametrische Verfahren für die Dichte- und Kurvenschätzung. Wirklich interessant wurden Glättungsmethoden für die Biometrie jedoch erst in ihrer multivariaten Ausformung. Dieser 'Dimensionswechsel' wirft bis heute offene methodische Fragen auf. Es darf daher nicht wundern, dass das vereinfachende Paradigma der additiven Regression, realisiert in den generalisierten additiven Modellen (GAM), den Siegeszug der Glättungsverfahren Anfang der 1990er Jahre eingeleitet hat. Parallel dazu hat es neue Algorithmen und bedeutende Softwareentwicklungen, vor allem in den statistischen Programmiersprachen S und R, gegeben. Neuere Entwicklungen von Glättungsverfahren finden sich in der Überlebenszeitanalyse, der Longitudinalanalyse, den gemischten Modellen und in der Funktionaldatenanalyse, teilweise unter Einbeziehung Bayesianischer Konzepte. Ganz neu sind statistische Methoden mit Glättungsbezug in der Bioinformatik.
In diesem Artikel wird nicht nur ein allgemeiner historischer Überblick gegeben, sondern auch versucht speziell die Aktivitäten im deutschsprachigen Raum zu skizzieren. Auch die derzeitige Situation wird einer kritischen Betrachtung unterzogen. Schlussendlich wird eine große Anzahl relevanter Literaturhinweise gegeben.

Schlüsselwörter: Biostatistik, Dichteschätzung, Glättungsverfahren, Geschichte, Kernschätzer, Kurvenschätzung, lokale Polynome, Regression, Splines

In Germany around 25 years ago nonparametric smoothing methods have found their way into statistics and with some delay also into biometry. In the early 1980's there has been what one might call a boom in theoretical and soon after also in computational statistics. The focus was on univariate nonparametric methods for density and curve estimation. For biometry however smoothing methods became really interesting in their multivariate version. This 'change of dimensionality' is still raising open methodological questions. No wonder that the simplifying paradigm of additive regression, realized in the generalized additive models (GAM), has initiated the success story of smoothing techniques starting in the early 1990's. In parallel there have been new algorithms and important software developments, primarily in the statistical program-

[1] *Medizinische Universität Graz, Institut für Medizinische Informatik, Statistik und Dokumentation, Graz, Austria*

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000009.shtml
Bitte zitieren als: Schimek MG. Glättungsverfahren in der Biometrie: ein historischer Abriss. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc09.
gbv(M20050329-1)

ming languages S and R. Recent developments of smoothing techniques can be found in survival analysis, longitudinal analysis, mixed models and functional data analysis, partly integrating Bayesian concepts. All new are smoothing related statistical methods in bioinformatics.

In this article we aim not only at a general historical overview but also try to sketch activities in the German-speaking world. Moreover, the current situation is critically examined. Finally a large number of relevant references is given.

Keywords: biostatistics, curve estimation, density estimation, history, kernel estimator, local polynomials, regression, smoothing methods, splines

Einleitung

In diesem Artikel wird versucht, die Bedeutung von Glättungsverfahren für die Biometrie kritisch zu beleuchten. Diese international weit verbreitete, rechenintensive Methodik ist aus der modernen Statistik nicht mehr wegzudenken. Ziel ist es, ihre Positionierung in der Biostatistik bzw. Biometrie aus historischer Sicht aufzuzeigen. Eine solche Betrachtung ist niemals nur die Beschreibung einer Entwicklung in einem Forschungsfeld, sondern enthält immer auch eine persönliche Sicht, die von den eigenen Erfahrungen und Präferenzen in Theorie und Anwendung geprägt wird. Dies möge der werte Leser/die werte Leserin dem Verfasser zugute halten. Die notwendige Kürze einer solchen Abhandlung erzwingt zusätzlich eine inhaltliche Fokussierung, so erstrebenswert thematische Vollständigkeit auch sein mag.

Was ist nun die grundlegende Idee des Glättens? Sie besteht darin, eine funktionelle Beziehung zwischen unterschiedlichen Messgrößen zu finden. Eine wichtige Anwendung ist die Regressionsanalyse. In der klassischen, parametrischen Regression ist diese Beziehung festgelegt (z.B. eine Gerade). In der nichtparametrischen glättenden Regression hingegen bestimmen die Datenpunkte selbst die funktionelle Form, die abgesehen von gewissen Glattheitsanforderungen (z.B. an die Ableitungen der Funktion) beliebig ist.

Was ist ein Glätter (engl. 'smoother')? Ein statistisches Werkzeug für die Zusammenfassung der Beobachtungen an einer abhängigen Messgröße Y als Funktion einer oder mehrerer Prädiktormessgrößen $X_1,\ldots, X_p$. Der resultierende Schätzer ist weniger variabel als die Beobachtungen von Y selbst, was den Namen 'Glätter' erklärt.

Prinzipiell gibt es zwei Anwendungen in der Biometrie: Als deskriptives (exploratorisches) Hilfsmittel zur Visualisierung von Punktwolken (engl. 'scatterplots'; siehe z.B. [44]) und als Methode der Schätzung der Abhängigkeit des Mittels von Y von einem oder mehreren Prädiktoren (siehe z.B. [46]).

Wozu benötigen wir Glätter? Abgesehen von der eingangs erwähnten Regressionsanalyse vor allem für die Dichteschätzung.

Wie funktioniert ein Glätter? Er kann als gewichteter Mittelungsprozess charakterisiert werden (auch andere Erklärungsansätze sind gebräuchlich und nützlich). Die Mittelung erfolgt in der Nachbarschaft um den jeweiligen Zielpunkt über einen geordneten metrischen Prädiktor. Die Nachbarschaft (Bandweite) muss geeignet festgelegt werden. Es handelt sich somit um eine nichtparametrische Strategie. Im Gegensatz zu starren parametrischen Ansätzen ist die Abhängigkeit Ys von den $X_1,\ldots, X_p$ flexibel. Ein Gewichtskonzept kontrolliert die funktionelle Abhängigkeit. Für die Funktionsschätzung kann ein gewichtetes Mittel unter Einsatz einer Kernfunktion K zur Anwendung gebracht werden. Ein typisches Beispiel (für den Fall eines Prädiktors, p=1) ist der Nadaraya-Watson-Kernschätzer [36], [73], definiert durch

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} y_i K((x - x_i)/h)}{\sum_{i=1}^{n} K((x - x_i)/h)},$$

wobei n für den Stichprobenumfang und h für die Bandweite stehen (h>0). Die Bandweite kontrolliert die Glattheit der Funktionsschätzung. Mittels Nadaraya-Watson-Kernschätzer wird eine kontinuierliche Funktion erzeugt (unter der Annahme einer kontinuierlichen Kernfunktion K). Der Kern K ist eine beschränkte, integrierbare, reellwertige Funktion der Form

$$\lim_{x\to\infty} |x| K(x) = 0.$$

Üblicherweise wird für K eine symmetrische Wahrscheinlichkeitsdichtefunktion gewählt. Hier sind einige wichtige Beispiele:

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000009.shtml
Bitte zitieren als: Schimek MG. Glättungsverfahren in der Biometrie: ein historischer Abriss. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc09.

$$K_1(x) = I(x \in [-0.5,+0.5]),$$
$$K_2(x) = \frac{3}{4}(1-x^2)I(x \in [-1,+0.1]),$$
$$K_3(x) = \frac{15}{32}(3-10x^2+7x^4)I(x \in [-1,+0.1]),$$
$$K_4(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2).$$

In Abbildung 1 sind sie dargestellt.

Es existieren auch andere Kernschätzer, wie beispielsweise der Gasser-Müller-Kernschätzer (siehe später) oder der Parzen-Rosenblatt-Kernschätzer [38], [42].

Darüber hinaus gibt es noch weitere wichtige Glättungskonzepte. Hier sind die glättenden Splines und die lokalen Polynome zu nennen (für eine Einführung siehe Kapitel 1 und 9 in [49]).

Die Optimierung eines Kleinstquadratekriteriums mit Straffunktion führt zu glättenden Splines und stellt einen völlig anderen Zugang dar. Dieser Glätter kann daher nicht als Ergebnis eines gewichteten Mittelungsprozesses verstanden werden. Als Beispiel betrachten wir die kubische Splineregression mit dem Kriterium

$$\sum_{i=1}^{n}(y_i - f(x_i))^2 + \lambda \int f''(x)^2 dx,$$

wobei f für die glatte Funktion und λ (λ>0) für den Glättungsparameter (hat die selbe Aufgabe wie die Bandweite bei den Kernschätzern) stehen. Über dieses Kriterium wird ein Ausgleich zwischen der Datentreue und der Glattheit von f in Abhängigkeit von λ angestrebt. Die Lösung dieses Optimierungsproblems ist eine glättende Splinefunktion. James R. Thompson und Richard A. Tapia haben sich 1978 erstmals systematisch mit Optimierungsaufgaben unter Straftermen auseinandergesetzt (siehe [62]). Der wahrscheinlich erste moderne Beitrag zu glättenden Splines ist Wahba & Wold (1975) [67].

Lokale Polynome sind wiederum anders motiviert. Erste Arbeiten im statistischen Regressionskontext gehen auf die 70er Jahre des vorigen Jahrhunderts zurück (insbesondere seien Stone (1977) [61] und Cleveland (1979) [3] erwähnt). Die Grundidee ist, dass eine glatte Funktion durch ein Polynom niedrigen Grades in der Nachbarschaft eines beliebigen Punktes x gut approximiert werden kann. Als Beispiel soll die lokale lineare Approximation

$$\mu(x_i) \approx a_0 + a_1(x_i - x)$$

für $x-h \le x_i \le x+h$ (h ist die Bandweite) dienen. Diese Anpassung kann durch lokal gewichtete Kleinstquadrate erfolgen. Die Gewichte werden durch einen Kern K wie bei den Kernschätzern festgelegt. Die Koeffizienten a_0 und a_1 ergeben sich aus der Minimierung von

$$\sum_{i=1}^{n} K((x_i - x)/h)(y_i - (a_0 + a_1(x_i - x)))^2.$$

Eine aktuelle Diskussion der lokalen polynomialen Regression findet sich in Cleveland & Loader (1996) [2].

Unterschiedlichen Glättern, wie Kernschätzern oder die noch nicht erwähnten Nearest-Neighbour-Schätzern, kommt auch Bedeutung für die Dichteschätzung zu, die folgende Ziele verfolgt: Die Aufdeckung von Formmerkmalen in Daten mittels einer Dichtefunktion $f(\boldsymbol{X})$ sowie die Vorhersage von Y aus der gemeinsamen Dichtefunktion $f(\boldsymbol{X},Y)$. Analog zur Regression, wenn die Daten keiner bekannten parametrischen Form (z.B. der Normalverteilung) folgen, ist nichtparametrische Dichteschätzung angezeigt. Allgemein kann gesagt werden, dass parametrische Schätzer im Vergleich zu nichtparametrischen dazu tendieren, niedrigere Varianz zu haben. Ist jedoch die unterstellte Form falsch, wächst der Bias beträchtlich. Nichtparametrische Methoden haben natürlich auch einen Bias. Er verschwindet jedoch asymptotisch für eine kontinuierliche Zielfunktion. Hervorragende Einführungen in die nichtparametrische Dichteschätzung finden sich in Silverman (1986) [54], Simonoff (1996) [57] und Scott (2004) [51].

Die hier vorgestellten Glättungsverfahren beruhen auf unterschiedlichen mathematisch-statistischen Konzeptionen. Annahmen wie auch Interpretationsmöglichkeiten sind daher nicht einheitlich. Glättende Splines lassen sich auch Bayesianisch motivieren und zeigen ebenso Zusammenhänge mit stochastischen Prozessen [29], [31]. Zum Teil können jedoch die hier genannten formalen Glättungskonzepte ineinander übergeführt werden (so genannte Kernäquivalenz; siehe die grundlegende Arbeit von Silverman [56]). Oft stimmen die resultierenden Schätzer überein, aber die Schätzfehler können abhängig vom Modell sein. Im Bayesschen Kontext bei apriori-Verteilungen auf Glättungsparameter ergeben sich größere Fehler [65].

Die Ansätze ergänzen einander in dem Sinne, dass sie für spezifische Anwendungen (z.B. in der Biometrie) unterschiedlich geeignet sind und differenzierte numerische Anforderungen stellen (bedeutend für Algorithmen- und Softwareentwicklung). Wie oben erwähnt, die Behandlung von Schätzfehlern kann in den

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000009.shtml
Bitte zitieren als: Schimek MG. Glättungsverfahren in der Biometrie: ein historischer Abriss. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc09.

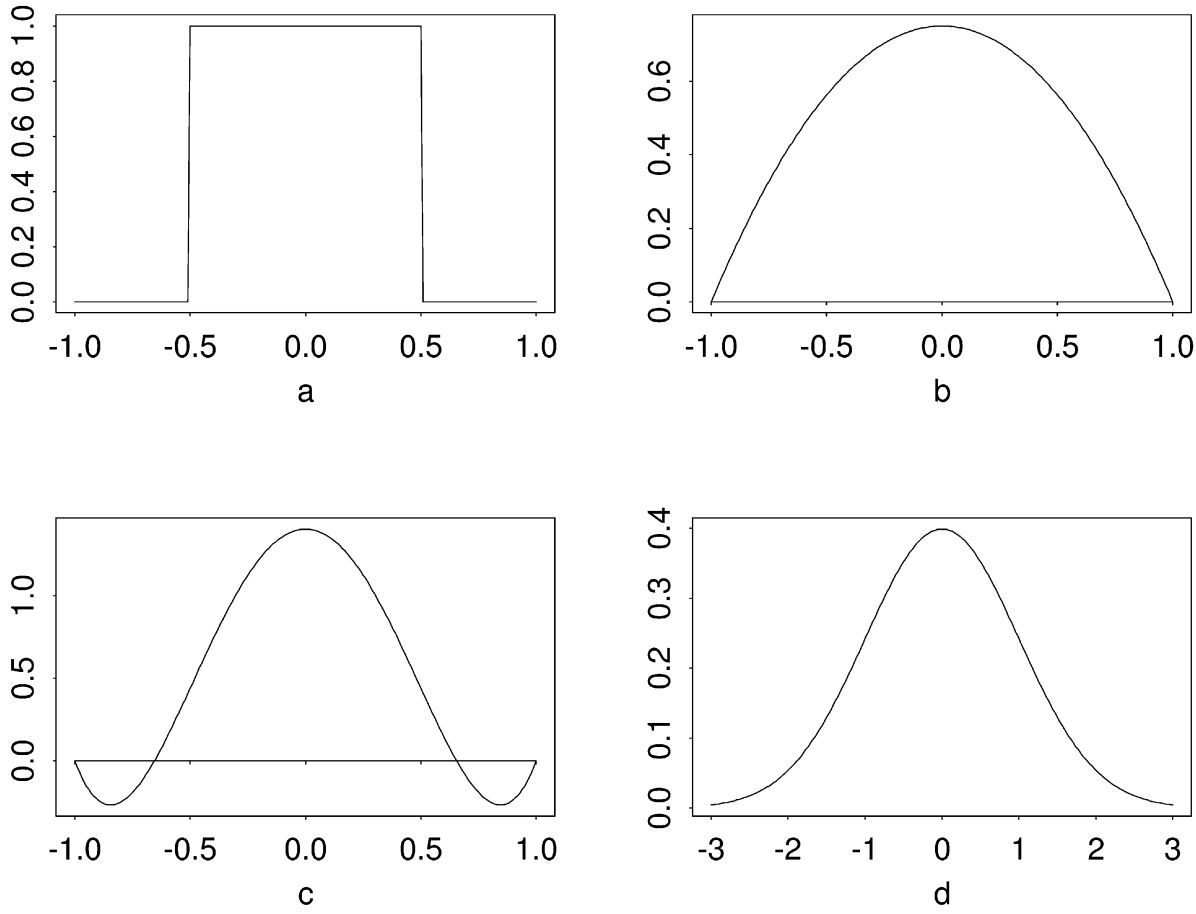

Abbildung 1: a: Uniformer Kern K_1; b: Epanechnikov Kern K_2 ; c: Kern 4. Ordnung K_3; d: Normaler Kern K_4

einzelnen Glättungsansätzen unterschiedlich sein, was auch die Interpretation von Ergebnissen beeinflusst.

Hier sei noch eine abschließende Bemerkung erlaubt: Unabhängig vom angewandten Glättungskonzept stellt sich die Frage wie man die Bandweite *h* bzw. den Glättungsparameter λ optimal wählt. Beiden Parametern kommt die zentrale Rolle zu, den Ausgleich zwischen der Varianz und dem Bias der zu schätzenden glatten Funktion bezüglich der Extreme 'Interpolation' einerseits und 'maximaler Glattheit' (i.e. Linearität) andererseits zu kontrollieren. Die Literatur zur Theorie und Praxis der Wahl von *h* und λ nimmt daher breiten Raum ein. Eine Diskussion würde den Rahmen dieses Artikels sprengen. Einen guten Überblick bieten die Kapitel 2 und 4 in Schimek (2000) [49]. Sehr zu empfehlen ist auch der kritische Artikel von Marron (1996) [34].

Die historischen Anfänge

Erste Glättungsideen finden sich bereits im 19. Jahrhundert. Nur was von Hand berechenbar war, konnte sich damals durchsetzen. Ein frühes Anwendungsgebiet war die Versicherungsmathematik (Berechnung von Mortalitäts- und Morbiditätsraten). Von lokaler polynomialer Anpassung wird bereits in de Forest (1873) [5] und Gram (1883) [18] berichtet. In diesem Zusammenhang hat sich J. Spencer 1904 in 'On the graduation of rates of sickness and mortality' [60] erstmals mit der Bandweitenwahl auseinandergesetzt. Der erste Beitrag zur Splineregression geht auf E. Whittaker zurück: 'On a new method of graduation' [74]. Er wurde im Jahre 1923 veröffentlicht.

Der älteste und am weitesten verbreitete Dichteschätzer ist das Histogramm. Bereits für das Jahr 1661 wird berichtet, dass John Graunt diese Methode in einem Vortrag bei der Royal Society in London mit dem Titel 'Bills of Mortality' verwendet habe (siehe [62], S. 2ff). Auch in diesem Fall war die Motivation eine demographische Aufgabenstellung.

Als weiteres historisch bedeutendes Anwendungsgebiet stellt sich die mathematische Lösung von Ausgleichs- und Interpolationsproblemen in den Naturwissenschaften dar. Ein verwandtes Aufgabenfeld war einiges später die Approximation von Funktionen in der damals neuen digitalen Computertechnik. Viele

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000009.shtml
Bitte zitieren als: Schimek MG. Glättungsverfahren in der Biometrie: ein historischer Abriss. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc09.

Probleme sind erst durch den Einsatz von Digitalcomputern wissenschaftlich interessant und numerisch handhabbar geworden. Es muss jedoch angemerkt werden, dass es sich damals üblicherweise nicht um stochastische Betrachtungen (also keine Berücksichtigung von Schätzfehlern) der Probleme gehandelt hat. Die zufallsbezogene Sichtweise ist der modernen Statistik ab den 60er Jahren des vorigen Jahrhunderts vorbehalten geblieben. Was die damit verbundenen Algorithmen angeht, konnte die rechenintensive Statistik auf einigen Vorarbeiten der numerischen Mathematik aufbauen (z.B. auf den bahnbrechenden Arbeiten für Splines von C. de Boor; siehe Monographie de Boor [4]).

Statistisches Glätten und Biometrie

• Frühe Entwicklungen

Am Anfang standen theoretische Fragestellungen im Kontext von Kernschätzern, univariate Probleme betreffend (i.e. ein Prädiktor *X*), im Vordergrund. Abgesehen von den schon oben erwähnten Arbeiten von Nadaraya (1964) [36] und Watson (1964) [73] sind hier folgende Beiträge hervorzuheben: Rosenblatts (1956) 'Remarks on some nonparametric estimates of a density function' [42], der weitere Arbeiten einer ganzen Generation von Forschern/Forscherinnen folgen sollten (siehe später); Parzens (1962) 'On estimation of probability density functions and mode' [38] und Priestleys & Chaos (1972) 'Nonparametric curve fitting' [39]. Erst im Laufe der 1970er Jahre wird dieses Forschungsgebiet zunehmend perzipiert.

Sieht man von Whittaker (1923) [74] ab, eine Arbeit, die heute im Lichte der Splinemethodik gesehen wird, ist um 1970 die Geburtsstunde der glättenden Splines zu veranschlagen. Als Pioniere sind George Kimeldorf und Grace Wahba zu nennen. Ihre Titel lauteten 'Spline functions and stochastic processes' [30] sowie 'A correspondence between Bayesian estimation of stochastic processes and smoothing by splines' [29]. Weitere wichtige Arbeiten hat Grace Wahba als Alleinautorin 1978 [71], 1979 [70], 1983 [69] und 1985 [68] veröffentlicht. Wahba, die Physikerin und spätere Ordinaria für Statistik an der University of Wisconsin-Madison, ist zweifelsohne eine Wegbereiterin der Glättungsverfahren in der modernen uni- und multivariaten Statistik. Wie schon die Arbeiten mit Kimeldorf verraten, liegt ihr Hauptinteresse auf Splineglättung mit Bezug zur Vorhersage stochastischer Prozesse und zum Bayes-Ansatz. Wahbas Monographie von 1990 'Spline Models for Observational Data' [72] verschafft einen guten Überblick über den Beitrag ihrer 'Schule' (nicht wenige ihrer einstigen Doktoranden sind heute selbst als Professoren Vertreter dieser Spezialisierung). Ihren Arbeiten ist es zu verdanken, dass diese Methodik zunehmend auch in Europa akademische Verbreitung gefunden hat. Viele ihrer Ideen fanden Ende der 1970er Jahre bereitwillige Aufnahme bei theoretischen Statistikern wie Bernard W. Silverman (damals University of Bath, UK). Mit seiner Arbeit 'Some aspects of the spline smoothing approach to non-parametric regression curve fitting' aus 1985 [55] hat er das Feld für praktische Anwendungen bestellt. Auf ihn geht auch die erste Implementation glättender Splines (FORTRAN Programm 'BATHSPLINE') zurück [47]. Weiters sind seine Monographie 'Density Estimation for Statistics and Data Analysis' aus 1986 [54] und gemeinsam mit Peter Green (University of Bristol, UK) sein Buch aus 1994 'Nonparametric Regression and Generalized Linear Models' [19] zu nennen. Beide Werke haben bis heute großen Einfluss auf die Biometrie, vor allem auch wegen der enthaltenen Beispiele. Neben nordamerikanischen und britischen Statistikern/Statistikerinnen ist hier vor allem ein lange Zeit in Deutschland wirkender Schweizer, Theo Gasser, zu nennen (siehe später).

Warum hat es seit den ersten Anfängen so lange gedauert, bis Forschungsergebnisse zu Glättungsmethoden in der Statistik aufgegriffen wurden?

Die Antwort ist einfach und besteht aus zwei Teilen: (1) Die mathematische und ebenso die angewandte Statistik waren bis weit in die 80er Jahre des vorigen Jahrhunderts, in Deutschland noch länger, von klassischen parametrischen Verfahren dominiert. (2) Die für die nichtparametrische Statistik erforderliche Rechnerleistung war die längste Zeit, mangels computertechnischer Grundlagen nicht ausreichend. Weiters ist noch das geringe Interesse an exploratorischen und graphischen Verfahren außerhalb der USA zu nennen. Seit den 80er Jahren - dank des enormen Fortschritts der Hardware- und Softwareentwicklung und dem Wechsel von Mainframes zu Arbeitsplatzrechnern - ist es regelrecht zu einem Boom, vorerst in ausgewählten Zentren in den USA, in Kanada, im UK und in Australien gekommen. Die Ausrichtung war sowohl mathematische Statistik als auch rechenintensive Statistik. Noch gab es nur wenige Anwendungen weil die Literatur sehr formal und geeignete Software noch nicht kommerziell erhältlich war.

Mit der weiteren, auch geographischen Verbreitung dieser Forschungsbereiche, vor allem auf dem europäischen Festland, wurden neue Themen aufgegriffen, unter ihnen Vorschläge für alternative (vor allem lokale) univariate Glätter, für die Wahl der Bandweite bzw. des Glättungsparameters und zunehmend auch für multivariate Probleme. Viele dieser Entwicklungen basierten immer noch auf der Annahme unabhängig identisch verteilter Fehler, was biometrischen Anwendungen häufig widerspricht.

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000009.shtml
Bitte zitieren als: Schimek MG. Glättungsverfahren in der Biometrie: ein historischer Abriss. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc09.

Der Beitrag Deutschlands

Ende der 1970er Jahre entstand an der Ruprecht-Karls-Universität Heidelberg eine erste Initiative unter der Leitung des schon erwähnten Theo Gasser. Es hätte aber auch anders kommen können, würde Gasser in den 1980er Jahren einen an ihn erteilten Ruf der Karl-Franzens-Universität Graz angenommen haben, was einen großen Gewinn für die österreichische biostatistische Forschung bedeutet hätte.

Für den deutschen Sprachraum selten wurde in Heidelberg anwendungsorientiert theoretische Statistik, und das war völlig neu, ausgerichtet auf biometrische Fragestellungen betrieben! Diese Aktivitäten sollten bald Aufwind bekommen durch den Sonderforschungsbereich 123 'Stochastische Mathematische Modelle' (1978-1992; Sprecher Willi Jäger) an der Ruprecht-Karls-Universität Heidelberg. In Gassers Teilprojekt 'Methoden der Zeitreihenanalyse' war der Fokus auf Kernschätzverfahren, Residualanalyse und Methoden der Bandweitenwahl, jeweils unter der Annahme zeitabhängiger Fehler. Wichtige Anwendungen waren die Modellierung von biologischen Wachstumsprozessen und die Analyse von EEG-Daten. Besonders hervorzuheben ist eine frühe Arbeit mit dem Titel 'Kernel estimation of regression functions' [14]. Dieser Beitrag ist in einem von Gasser und von dem schon erwähnten Pionier Murray Rosenblatt herausgegebenen Sammelband, dem historisch ersten Buch über Glättungsverfahren in der Statistik, 1979 erschienen. In dem Artikel mit Hans Georg Müller, einem der Doktoranden Gassers der ersten Stunde und heute Professor für Biostatistik an der University of California - Davis, wird ein Kernschätzer vorgeschlagen, der in der Folge in der Statistik unter der Bezeichnung Gasser-Müller-Schätzer bekannt geworden ist.

Andere inzwischen selbst eine Professur bekleidende Doktoranden und Mitarbeiter von damals sind Wolfgang Härdle (Humboldt Universität zu Berlin) und Alois Kneip (Universität Bonn). Beide sind heute außerhalb der Biometrie beschäftigt, letzterer jedoch weiterhin an der Funktionaldatenanalyse interessiert. Weitere, auch für biometrische Anwendungen, bedeutende Arbeiten sind von Gasser et al. (1984) 'Nonparametric regression analysis of growth curves' [12], von Gasser, Müller & Mammitzsch (1985) 'Kernels for nonparametric curve estimation' [13], von Gasser, Sroka & Jennen-Steinmetz (1986) 'Residual variance and residual pattern in nonlinear regression' [16] und von Gasser, Kneip & Köhler (1991) 'A flexible and fast method for automatic smoothing' [11]. Gassers Koautor Volker Mammitzsch (Phillips-Universität Marburg) zählt ebenfalls zu den ersten Professoren Deutschlands, die sich eingehend mit Glättungsmethoden auseinandergesetzt haben.

Gasser, nunmehr an der Universität Zürich tätig, ist der unangefochtene Doyen der deutschsprachigen Glätterszene. Vor allem in der Biometrie ist sein internationaler Einfluss unübersehbar. Ein weiterer facheinschlägig forschender Biostatistiker an der Universität Zürich ist Burkhardt Seifert (z.B. [52], [53]).

Was die Splinemethodik angeht, ist an der Schwelle zu den jüngeren Entwicklungen eine deutsche Mathematikerin zu erwähnen, Angelika van der Linde (Universität Bremen). Sie hat diverse Ideen Grace Wahbas und Bernard W. Silvermans aufgenommen, aus Bayesianischer Sicht weiterentwickelt und für biometrische Anwendungen fruchtbar gemacht [64], [63].

• Jüngere Entwicklungen

Für die 1990er Jahre kann man zweifelsohne von einem internationalen Innovationsschub sprechen. Neben Nordamerika, Australien und Europa etablierte sich auch der asiatische Raum. Computationalen Problemen wird zunehmend Aufmerksamkeit gezollt. Aus mathematisch-statistischer Sicht steht diese Periode vor allem für rechenintensive multivariate Ansätze. Zu zentralen Herausforderungen werden der so genannte 'Fluch der Dimensionalität', die mehrdimensionale Bandweitenwahl und effiziente statistische Algorithmen (inklusive deren Umsetzung in professioneller Software).

Seit den von Wahba unter strikten Annahmen eingeführten Thin-Plate-Splines (siehe z.B. [72]) sind die technischen Schwierigkeiten, die durch mehrdimensionales Glätten aufgeworfen werden, bekannt. Bislang ist unter keiner der bekannten Glättungsmethoden eine allgemein befriedigende praktische Lösung in Sicht. Es liegt daher nahe, das Problem durch geeignete Annahmen (vor allem Additivität) zu vereinfachen. Diese Idee wurde von den nordamerikanischen Biostatistikern Trevor Hastie und Robert Tibshirani mit ihren generalisierten additiven Modellen (GAM) erfolgreich umgesetzt. Die dazu notwendigen Vorarbeiten fanden eine monographische Zusammenfassung in 'Generalized Additive Models' [26]. Aus numerischer Sicht hat der wenig aufwendige Backfitting-Algorithmus den GAM zum Durchbruch verholfen (erstmals vorgeschlagen in Breiman & Friedman [1]). Überdies hat Hastie mit GAIM das erste Programm zur Auswertung solcher Regressionsmodelle zur Verfügung gestellt, das defacto in unveränderter Form bis heute die Basis für die Modellschätzung in S-Plus bildet. Auch in XploRe [21] wurden geeignete Algorithmen realisiert.

Diese Kombination aus Theorie, Algorithmus und praktischer Umsetzung in Verbindung mit einigen interessanten biometrischen Anwendungen hat dieser Modellklasse zu einem enormen Aufschwung verhol-

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000009.shtml
Bitte zitieren als: Schimek MG. Glättungsverfahren in der Biometrie: ein historischer Abriss. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc09.

fen. Es gibt zwei Anwendungsfelder, in denen nichtparametrische Glättungsverfahren, insbesondere GAM, zur Standardmethodik gehören: (1) Die epidemiologische Modellierung von Luftschadstoffen, inbesondere von ultrafeinen Partikeln (engl. 'Particulate Matter'; [50], [58]) und Modellierungen in der Pflanzenökologie [76]. Die gehäufte praktische Anwendung führt stets zu statistischen Erweiterungen. So ist man beispielsweise in der Epidemiologie an die Grenzen der etablierten GAM-Methodik gestoßen, was zu neuen Ansätzen, so z.B. zu den penalisierten GAM geführt hat [35], [7]. In der Pflanzenökologie wiederum hat es eine Erweiterung von GAM zu vektoriellen GAM gegeben [75].

Weitere Entwicklungen der letzten Jahre sind statistische Tests und Methoden der Modellkritik bei glättender Regression. Laufend wird an Algorithmen für neue Aufgaben und an Effizienzverbesserungen bekannter numerischer Methoden gearbeitet. Auch Glättungsverfahren unter Nicht-Standard-Annahmen (z.B. für zeitliche und räumliche Daten) nehmen einen breiten Raum ein.

Bei vielen medizinischen und biowissenschaftlichen Anwendungen spielt die Schätzung von Regressionskoeffizienten und ihren Schätzfehlern (z.B. zur Risikoabschätzung) eine große Rolle. Es darf daher nicht wundern, dass es seit geraumer Zeit Bestrebungen gibt, die Flexibilität der nichtparametrischen Ansätze mit den Vorteilen parametrischer Modelle (generalisierte lineare Modelle und gemischte Modelle) zu verbinden. Solche Konzepte führen die Bezeichnung 'semiparametrisch'. Die wichtigste Person auf diesem Gebiet war anfänglich Paul Speckman (University of Missouri - Columbia). Seine Arbeit 'Kernel smoothing in partial linear models' von 1988 [59] ist nicht nur ein Meilenstein in der theoretischen Entwicklung, sondern auch für die biometrische Forschungsgemeinde von Nutzen. Bei Schimek (2000) [45] finden sich numerisch effiziente Algorithmen. Um Verallgemeinerungen dieser Modelle hat sich die Biostatistikerin Joan Staniswalis (University of Texas at El Paso) verdient gemacht (für einen Überblick siehe Schimek [43], S. 293-312). Derzeit erleben die semiparametrischen Modelle einen Boom in der Umweltepidemiologie. Der Autor sieht für diese Modellklasse in den gesamten Biowissenschaften eine große Zukunft.

Weiters ist zu beobachten, dass die grundlegenden Glättungsverfahren für Dichte- und Regressionsschätzung Standardwerkzeuge der Statistik geworden sind (in vielen Softwarepaketen realisiert) und dass heute Glätter zunehmend als Bausteine von komplexen non- oder semiparametrischen oder gemischten Modellen, auch für die Analyse von Longitudinaldaten, Verwendung finden. Darüber hinaus können wir sehen, dass Glätter mit anderen formalen Konzepten (Bayes, Kalman-Filter, neuronale Netze, statistisches Lernen) kombiniert werden [9], [49], [25], [32], [6]. Für die Biometrie wichtig sind auch glättungsbasierte nonparametrische Ansätze für die Hazardregression [37], ([9], Kapitel 9), ([10], Kapitel 5) und die variierenden Koeffizientenmodelle [27]. Immer mehr biowissenschaftliche Anwendungsfelder werden von derartigen Methoden durchdrungen. Das gilt auch für die neuen Herausforderungen durch hochdimensionale Schätz- und Vorhersageprobleme, wie sie in der Bioinformatik vorkommen (für einen Überblick siehe [48]).

Eine neuere Entwicklung mit noch wenig genutztem Potential ist die Funktionaldatenanalyse. Sie geht im Wesentlichen auf James O. Ramsay (McGill University Montreal) und Bernard W. Silverman zurück. Eine gute Zusammenfassung der Methoden findet sich in ihrer Monographie von 1997 [41]. Zahlreiche Anwendungen, vor allem in den Biowissenschaften (betreffend Pathologie, Wachstum, Physiologie, 'Biometrie' im Sinne von Personenmerkmalserkennung) beinhaltet ihr Buch von 2002 [40]. Auch zur Funktionaldatenanalyse gibt es von Theo Gasser Beiträge bis in die jüngste Zeit (z.B. [17]).

Abschließend sei hier noch eine Bemerkung zur Berechenbarkeit von Glättungsverfahren erlaubt. Es besteht kein Zweifel, dass derartige Methoden numerisch anspruchsvoll und äußerst rechenintensiv sind. Ihre Berechenbarkeit verdanken wir jedoch nicht nur der Leistungsfähigkeit heutiger Digitalcomputer, sondern auch dem Fortschritt bei den statistischen Algorithmen. Es würde den Rahmen dieser Betrachtung sprengen, hier auf die historischen Details einzugehen (eine aktuelle Aufarbeitung der Thematik findet sich in [20]). Wichtige deutsche Beiträge zur effizienten Berechnung von Kernschätzern gehen auf den schon erwähnten Statistiker Wolfgang Härdle zurück. Sein Buch 'Smoothing Techniques. With Implementations in S' aus 1990 [24] hat erstmals Algorithmen für Glättungsverfahren in das Zentrum der Betrachtung gestellt. Die statistische Programmiersprache S, damals noch wenig verbreitet im deutschen Sprachraum, ist heute vor allem durch das kommerzielle S-Plus und seine nichtkommerzielle Variante R zum de-facto Standard in der modernen statistischen Analyse, aber auch für das Prototyping von Algorithmen, geworden. Die Verfügbarkeit der Sprachen S und R hat einen enormen Einfluss auf die Verbreitung der Glättungsverfahren ausgeübt. Wichtige Bücher für den Biostatistiker/die Biostatistikerin sind Venables & Ripley (2002) [66] sowie Everitt & Rabe-Hesketh (2001) [8].

Der Beitrag Deutschlands

Die 1990er Jahre in Deutschland sind von der Gründung dreier statistischer Sonderforschungsbereiche

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000009.shtml
Bitte zitieren als: Schimek MG. Glättungsverfahren in der Biometrie: ein historischer Abriss. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc09.

gekennzeichnet, die interessanterweise alle zu einem gewissen Teil auch Glättungsmethoden zum Forschungsgegenstand erkoren haben. Der erste wurde an der Humboldt Universität zu Berlin (SFB 373 'Quantifikation und Simulation ökonomischer Prozesse'; 1994-2003, Sprecher Wolfgang Härdle) eingerichtet. Die beiden anderen sind an der Ludwig-Maximilians-Universität München (SFB 386 'Statistische Analyse diskreter Strukturen'; Sprecher Ludwig Fahrmeir) und an der Universität Dortmund (SFB 475 'Komplexitätsreduktion in multivariaten Datenstrukturen'; Sprecherin Ursula Gather) beheimatet. Die beiden noch aktiven Sonderforschungsbereiche integrieren biometrische Teilprojekte. Auch nach dem Ende des Sonderforschungsbereichs 123 und Theo Gassers Ruf an die Universität Zürich sind einschlägige Forschungsaktivitäten an der Ruprecht-Karls-Universität Heidelberg noch bis 2003 zu beobachten (z.B. theoretische Beiträge zu GAM [33]). Der Wechsel Enno Mammens an die Universität Mannheim hinterlässt eine Lücke in Heidelberg, was Glättungsverfahren anbelangt.

Leider ist ein stetiger Rückgang einschlägiger Forschungsaktivitäten an allen Universitätsstandorten, vielleicht abgesehen von München, zu verzeichnen.

Schlussfolgerungen

In Deutschland kann man im internationalen Vergleich immer noch Vorbehalte gegenüber Glättungsmethoden in der akademischen Statistik, die ja traditionell mathematisch orientiert ist, orten. Die vielfältigen alternativen formalen Konzepte bzw. Modellierungen scheinen vielen Mathematikern/Mathematikerinnen nicht logisch zwingend zu sein. Noch größer ist der Widerstand gegen Baysianische Ansätze wegen derer vermeintlicher Subjektivität. Völlig anders ist die Situation an angloamerikanischen Universitäten wo den Glättungsverfahren in Forschung, Lehre und Anwendung breiter Raum gegeben wird.

Glättungsmethoden haben zweifelsohne ein großes Anwendungspotential. Anwendungsgeleitete statistische Forschung ist im deutschen Sprachraum noch immer nicht der Regelfall. Akademische Statistiker/Statistikerinnen sind überwiegend ausgebildete Mathematiker/Mathematikerinnen, die sich dem Fach Statistik erst später genähert haben. Als solche zeigen sie natürlich spezifische fachliche Präferenzen und eher geringen Bezug zu Datenproblemen aus der Praxis.

Nicht nur in der angloamerikanischen Welt ist die Tendenz zu erkennen, dass Absolventen/Absolventinnen statistischer Studiengänge auch als akademische Lehrer/Lehrerinnen bestellt werden. Die Berufung einer solchen Person ist in Deutschland die Ausnahme. Glättungsmethoden sind auch in der akademischen Lehre wenig verankert. Daher muss mittelfristig mit einem Rückgang des wissenschaftlichen Nachwuchses auf diesem Gebiet und mit einer Abnahme des Einsatzes dieser Methoden gerechnet werden. Es ist zu befürchten, dass derartige Kompetenz gerade dort wo sie am meisten gebraucht werden wird, nämlich in den modernen Biowissenschaften einschließlich der Genforschung, nicht ausreichend vorhanden sein wird.

Gründe für einen Rückgang der einschlägigen Forschungsaktivitäten in Deutschland in den letzten zehn Jahren sieht der Verfasser in einer Verlagerung des akademischen Interesses auf neue Forschungsgebiete (z.B. Bioinformatik: die neu eingerichteten Studiengänge boomen, bieten jedoch nur wenig Ausbildung in Biostatistik), in der forcierten Sparpolitik im Hochschulwesen und in der stärkeren Konzentration der statistischen Forschung an wenigen Standorten, was viele Mittel der Forschungsförderung bindet.

Eine erwartbare Beobachtung ist die im internationalen Vergleich geringe Zahl publizierter Anwendungen deutscher Autoren/Autorinnen. Die Biometrie als anwendungsorientiertes Fach würde viele Perspektiven für den vorteilhaften Einsatz von Glättungsmethoden bieten. Beispielsweise in den Programmheften der Biometrischen Kolloquien der Deutschen Region der Internationalen Biometrischen Gesellschaft ist ein gehäuftes Auftreten von einschlägigen Beiträgen nur Mitte der 1980er bis Mitte der 1990er Jahre zu registrieren. Betrachtet man die International Biometric Conference in Freiburg 2002, die das Biometrische Kolloquium inkludiert hat [28], stehen einige einschlägige (auch eingeladene) Vorträge im internationalen Teil keinem einzigen im deutschen Teil gegenüber. Auch bei den Arbeitstagungen über Methoden und Werkzeuge der Informatik für die Statistik auf der Reisensburg ist derselbe Trend zu erkennen.

Ähnliches gilt auch für die europäischen COMPSTAT Symposien der International Association for Statistical Computing, die auch Entwicklungen im deutschen Sprachraum widerspiegeln. In den Proceedings des Berliner Symposiums im Jahr 2002 [22] findet sich kein einziger deutscher Beitrag zum Glätten im weitesten Sinn, wohl aber einige (einschließlich eingeladener) von Forschern/Forscherinnen anderer Nationalitäten. Eine COMPSTAT Spezialveranstaltung zum Thema 'Smoothing' fand das letzte Mal 1994 auf dem österreichischen Semmering statt [23]. An dieser Stelle sollte nicht verschwiegen werden, dass seither international eine Vielzahl theoretischer Ergebnisse vorgelegt und zahlreiche neue Anwendungen aufgezeigt worden sind. Die Schweiz und Österreich folgen leider dem für Deutschland aufgezeigten Muster.

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000009.shtml
Bitte zitieren als: Schimek MG. Glättungsverfahren in der Biometrie: ein historischer Abriss. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc09.

Danksagung

Der Verfasser möchte zwei anonymen Gutachtern für wertvolle Literaturhinweise und kritische Anmerkungen danken. Ebenso gilt sein Dank Dr. Gerhard Bachmaier (Medizinische Universität Graz) für die Konvertierung des unter LATEX erzeugten Manuskripts in ein Word-dokument.

Korrespondenzadresse:

• Prof. Dr. Dr. Michael G. Schimek, Medizinische Universität Graz, Institut für Medizinische Informatik, Statistik und Dokumentation, Auenbruggerplatz 2, A-8036 Graz, Austria
michael.schimek@meduni-graz.at

Literatur:

[1] Breiman L, Friedman JH. Estimating optimal transformations for multiple regression and correlation (with discussion). J Amer Statist Assoc. 1985;80:580-619.

[2] Cleveland WS, Loader C. Smoothing by Local Regression: Principles and Methods. In: Härdle W, Schimek MG, eds. Statistical Theory and Computational Aspects of Smoothing. Heidelberg: Physica-Verlag; 1996. p. 10-49.

[3] Cleveland WS. Robust locally weighted regression and smoothing scatterplots. J Amer Statist Assoc. 1979;74:829-36.

[4] de Boor C. A Practical Guide to Splines. New York: Springer; 1978.

[5] de Forest. On some methods of interpolation applicable to the graduation of irregular series. In: Annual Report of the Board of Regents of the Smithsonian Institution for 1871. 1873. p. 275-339.

[6] Denison DGT, Holmes CC, Mallick BK, Smith AFM. Bayesian Methods for Nonlinear Classification and Regression. New York: Wiley; 2002.

[7] Eilers PHC, Schimek MG. Generalized additive models in particulate matter studies: statistical and computational perspectives. Bulletin de l'Institut International de Statistique. 2003;54ème Session, Livraison 1:332-5.

[8] Everitt B, Rabe-Hesketh S. Analyzing Medical Data Using S-Plus. New York: Springer; 2001.

[9] Fahrmeir L, Tutz G. Multivariate Statistical Modelling based on Generalized Linear Models. New York: Springer; 1996.

[10] Fan J, Gijbels I. Local Polynomial Modelling and Its Applications. London: Chapman & Hall; 1996.

[11] Gasser T, Kneip A, Köhler W. A flexible and fast method for automatic smoothing. J Amer Statist Assoc. 1991;86:643-52.

[12] Gasser T, Müller HG, Köhler W, Molinari L, Prader A. Nonparametric regression analysis of growth curves. Ann Statist. 1984;12:210-9.

[13] Gasser T, Müller HG, Mammitzsch V. Kernels for nonparametric curve estimation. J Roy Statist Soc. 1985;B 47:238-52.

[14] Gasser T, Müller HG. Kernel estimation of regression functions. In: Gasser T, Rosenblatt M, eds. Smoothing Techniques for Curve Estimation. New York: Springer; 1979. p. 23-68.

[15] Gasser T, Rosenblatt M, eds. Smoothing Techniques for Curve Estimation. New York: Springer; 1979.

[16] Gasser T, Sroka L, Jennen-Steinmetz C. Residual variance and residual pattern in nonlinear regression. Biometrika. 1986;73:625-33.

[17] Gasser T. Functional Data Analysis and Human Growth. In: International Biometric Society. Proceedings of the XXth International Biometric Conference, Berkeley, California. Vol. II. 2000. p. 251-3.

[18] Gram JP. Über Entwickelung reeller Functionen in Reihen mittels der Methode der kleinsten Quadrate. J Math. 1883;94:41-73.

[19] Green P, Silverman BW. Nonparametric Regression and Generalized Linear Models. A Roughness Penalty Approach. London: Chapman & Hall; 1994.

[20] Grossmann W, Schimek MG, Sint PP. The history of COMPSTAT and key-steps of statistical computing during the last 30 years. In: Antoch J, ed. COMPSTAT 2004. Proceedings in Computational Statistics. Heidelberg: Physica-Verlag; 2004. p. 1-35.

[21] Härdle W, Klinke S, Müller M. XploRe - Academic Edition. The Interactive Statistical Computing Environment (CD-ROM with Handbook). Berlin: Springer-Verlag; 1999.

[22] Härdle W, Rönz B, eds. COMPSTAT 2002. Proceedings in Computational Statistics. Heidelberg: Physica-Verlag; 2002.

[23] Härdle W, Schimek MG, eds. Statistical Theory and Computational Aspects of Smoothing. Heidelberg: Physica-Verlag; 1996.

[24] Härdle W. Smoothing Techniques. With Implementations in S. New York: Springer; 1991.

[25] Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning : data mining, inference, and prediction. New York: Springer; 2001.

[26] Hastie TJ, Tibshirani RJ. Generalized Additive Models. London: Chapman & Hall; 1990.

[27] Hastie TJ, Tibshirani RJ. Varying-coefficients models. J Roy Statist Soc. 1993;B 55:757-96.

[28] International Biometric Society. Proceedings of the XXIst International Biometric Conference, Freiburg, Germany; 2002.

[29] Kimeldorf G, Wahba G. A correspondence between Bayesian estimation of stochastic processes and smoothing by splines. Ann Math Statist. 1970; 41:495-502.

[30] Kimeldorf G, Wahba G. Spline functions and stochastic processes. Sankhya. 1970;A 132:173-80.

[31] Kohn R, Ansley CF. A new algorithm for spline smoothing based on smoothing a stochastic process. SIAM Journal of Scientific Computing. 1987;8:33-48.

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000009.shtml
Bitte zitieren als: Schimek MG. Glättungsverfahren in der Biometrie: ein historischer Abriss. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc09.

[32] Kohn R, Smith M, Chan D. Nonparametric regression using linear combinations of basis functions. Statistics and Computing. 2001;11:313-22.

[33] Mammen E, Linton O, Nielsen J. The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. Ann Statist. 1999;27:1443-90.

[34] Marron JS. A personal view of smoothing and statistics. In: Härdle W, Schimek MG, eds. Statistical Theory and Computational Aspects of Smoothing. Heidelberg: Physica-Verlag; 1996. p. 1-9.

[35] Marx BD, Eilers PHC. Direct generalized additive modeling with penalized likelihood. Computational Statistics and Data Analysis. 1998;28:193-209.

[36] Nadaraya EA. On estimating regression. Theory of Probability and its Applications. 1964;10:186-96.

[37] O'Sullivan F. Fast computation of fully automated log-density and log-hazard estimators. SIAM J Sci Statist Comput. 1987;9:363-79.

[38] Parzen E. On estimation of probability density functions and mode. Ann Math Statist. 1962;33:1065-76.

[39] Priestley MB, Chao MT. Nonparametric curve fitting. J Roy Statist Soc. 1972;B 34:385-92.

[40] Ramsay JO, Silverman BW. Applied Functional Data Analysis. Methods and Case Studies. New York: Springer; 2002.

[41] Ramsay JO, Silverman BW. Functional Data Analysis. New York: Springer; 1997.

[42] Rosenblatt M. Remarks on some nonparametric estimates of a density function. Ann Math Statist. 1956;27:832-37.

[43] Schimek MG, ed. Semiparametric function estimation and testing. Statistics and Computing. 2001;11:291-335.

[44] Schimek MG, Kubik W. Möglichkeiten und Grenzen des Werkzeuges XploRe aus der Sicht des Biometrikers. In: Enke H, Gölles J, Haux R, Wernecke KD, eds. Methoden und Werkzeuge für die exploratorische Datenanalyse in den Biowissenschaften. Stuttgart: G. Fischer; 1992. p. 129-39.

[45] Schimek MG. Estimation and inference in partially linear models with smoothing splines. J Statist Plan Infer. 2000;91:525-40.

[46] Schimek MG. Non-parametric regression techniques for biometric problems: Concepts and software. In: Adlassnig KP, Grabner G, Bengtsson S, Hansen R, eds. Medical Informatics Europe 1991. Berlin: Springer-Verlag; 1991. p. 562-6.

[47] Schimek MG. Non-parametric spline regression by Bathspline: Foundations and application. In: Faulbaum F, Haux R, Jöckel KH, eds. SOFTSTAT '89. Fortschritte der Statistik-Software 2. Stuttgart: G. Fischer; 1990. p. 224-34.

[48] Schimek MG. Penalized logistic regression in gene expression analysis. Methods Inform Med. 2004;43:439-44.

[49] Schimek MG. Smoothing and Regression. Approaches, Computation and Application. New York: Wiley; 2000.

[50] Schwartz J. The use of generalized additive models in epidemiology. In: XVIIth International Biometric Conference, Hamilton, Canada. Proceedings. Vol.1. 1994. p. 55-80.

[51] Scott DW. Multivariate density estimation and visualization. In: Gentle JE, Härdle W, Mori Y, eds. Handbook of Computational Statistics. New York: Springer; 2004. p. 517-38.

[52] Seifert B, Brockmann M, Engel J, Gasser T. Fast algorithms for nonparametric curve estimation. J Comp Graph Statist. 1994;4:192-213.

[53] Seifert B, Gasser T. Variance properties of local polynomials and ensuing modifications. In: Härdle W, Schimek MG, eds. Statistical Theory and Computational Aspects of Smoothing. Heidelberg: Physica-Verlag; 1996. p. 50-79.

[54] Silverman BW. Density Estimation for Statistics and Data Analysis. London: Chapman & Hall; 1986.

[55] Silverman BW. Some aspects of the spline smoothing approach to non-parametric regression curve fitting (with discussion). J R Statist Soc. 1985;B 47:1-52.

[56] Silverman BW. Spline smoothing: The equivalent variable kernel method. Ann Statist. 1984;12:896-916.

[57] Simonoff JS. Smoothing Methods in Statistics. New York: Springer; 1996.

[58] Smith RL, Davis JM, Speckman P. Airborne particles and mortality. In: Nychka D, Piegorsch WW, Cox LH, eds. Case Studies in Environmental Statistics. New York: Springer; 1998. p. 91-120.

[59] Speckman P. Kernel smoothing in partial linear models. J Royal Statist Soc. 1988;B 50:413-36.

[60] Spencer J. On the graduation of rates of sickness and mortality. J Inst Act. 1904;38:334-47.

[61] Stone CJ. Consistent nonparametric regression (with discussion). Ann Statist. 1977;5:595-645.

[62] Thompson JR, Tapia RA. Nonparametric Function Estimation, Modeling, and Simulation. Philadelphia: SIAM; 1990.

[63] van der Linde A, Osius G. Estimation of nonparametric risk functions in matched case-control studies. Statistics in Medicine. 2001;20:1639-62.

[64] van der Linde A, Witzko KH, Jöckel KH. Spatio-temporal analysis of mortality using splines. Biometrics. 1995;51:1352-60.

[65] van der Linde A. Smoothing errors. Statistics. 1998;31:91-114.

[66] Venables WN, Ripley BD. Modern Applied Statistics with S-Plus. 4. ed. New York: Springer; 2002.

[67] Wahba G, Wold S. A completely automatic French curve. Commun Statist. 1975;4:1-17.

[68] Wahba G. A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. Ann Statist. 1985;13:1378-1402.

[69] Wahba G. Bayesian 'confidence intervals' for the cross-validated smoothing spline. J Roy Statist Soc. 1983; B 45:133-50.

[70] Wahba G. Convergence rates of thin plate smoothing splines when the data are noisy. In: Gasser T, Rosenblatt

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000009.shtml
Bitte zitieren als: Schimek MG. Glättungsverfahren in der Biometrie: ein historischer Abriss. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc09.

M, eds. Smoothing Techniques for Curve Estimation. Berlin: Springer-Verlag; 1979. p. 233-45.

[71] Wahba G. Improper priors, spline smoothing and the problem of guarding against model errors in regression. J Roy Statist Soc. 1978; B 40:364-72.

[72] Wahba G. Spline models for observational data. Philadelphia, Pa.: Society for Industrial and Applied Mathematics; 1990. CBMS-NSF regional conference series in applied mathematics; 59.

[73] Watson GS. Smooth regression analysis. Sankhya. 1964;A 26:359-72.

[74] Whittaker E. On a new method of graduation. Proc Edinburgh Math Soc. 1923;41:63-75.

[75] Yee TW, Mackenzie M. Vector generalized additive models in plant ecology. Ecological Modelling. 2002;157:141-56.

[76] Yee TW, Mitchell ND. Generalized additive models in plant ecology. J Veg Sci. 1991;2:587-602.

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000009.shtml
Bitte zitieren als: Schimek MG. Glättungsverfahren in der Biometrie: ein historischer Abriss. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc09.
gbp(M2005002)29.1)