Abstract

This Ph.D. thesis treats consistency, robustness and discontinuity-preserving issues of M-kernel estimators in one-and two-dimensional regression.
The statistical model in the one-dimensional case looks as follows: consider the task of estimating the regression function $m : [0,1] \longrightarrow I\!R$ from a dataset of random variables $Y_1, \ldots, Y_n$ measured at the design points $0 \leq x_1 \leq \ldots \leq x_n \leq 1$. This is done by the M-kernel smoother $m_n(x)$ defined by

$$m_n(x) \in \arg\min_{y \in I\!R} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_n} K\left(\frac{x - x_i}{h_n}\right) \frac{1}{g_n} L\left(\frac{y - Y_i}{g_n}\right),$$

where K is a kernel function (i.e. $\int K(u)du = 1$) and L is some score function. Notice that this estimator is a generalization of the ordinary kernel estimator which is obtained with $L(\cdot) = (\cdot)^2$. The M-kernel estimator was first introduced by Härdle and Gasser(1984) with a monotone score function L (and fixed scale parameter g_n). Härdle and Gassergive a restricted valid proof for consistency and show the minimax property. Chu et al. (1998) take a density as (redescending) score function L and let $g_n \to 0$. They observe a good jump-preserving property. However, as shown in this thesis, consistency cannot be achieved under the assumptions given in their paper.
In this thesis, complete proofs of robustness for both monotone and redescending M-kernel smoothers are provided. It is further shown that the monotone M-kernel smoother is asymptotically robust while the redescending M-kernel smoother as introduced by Chu et al. does not have this property. On the other hand, as shown in this thesis, the latter estimator is consistent close to a discontinuity ("jump-preserving"), if the noise level is not very high. The monotone M-kernel smoother, however, does not have this property.
But if we leave the scale parameter g_n constant, even robustness can be obtained for the redescending M-kernel estimator without loosing the jump-preserving property.
Since discontinuities in the two-dimensional regression model can have many different shapes, the model of a two-dimensional regression function with a one-dimensional subset of discontinuities is formalized based on a differential-geometric approach. It is shown that the redescending M-kernel smoother even preserves sharp corners which is, in combination with qualitative asymptotic robustness, a unique property among nonparametric smoothers.
In simulations it is observed that the redescending M-kernel smoother is not robust against outliers. Therefore, the Trimmed M-Kernel Estimator is introduced combining the corner-preserving property of the redescending M-kernel smoother and the robustness against outliers of the LTS estimator.
Finally, the estimators are compared in two-dimensional simulation examples.

Zusammenfassung

Die vorliegende Dissertation behandelt Konsistenz, Robustheit und Erhalten von Unstetigkeitsstellen durch M-Kernschätzer in ein- und zweidimensionaler Regression.
Das statistische Modell im eindimensionalen Fall stellt sich folgendermaßen dar: Es sei eine Regressionsfunktion $m : [0\,,1] \longrightarrow I\!R$ durch eine Stichprobe $Y_1, \ldots, Y_n$ zu schätzen, die an den Stützstellen $0 \leq x_1 \leq \ldots \leq x_n \leq 1$ gemessen wird. Dies geschieht durch den M-Kernschätzer $m_n(x)$, der durch

$$m_n(x) \in \arg\min_{y \in I\!R} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_n} K\left(\frac{x - x_i}{h_n}\right) \frac{1}{g_n} L\left(\frac{y - Y_i}{g_n}\right),$$

definiert wird, wobei K ein Kern ist (d.h. $\int K(u)du = 1$) und L eine Scorefunktion. Man beachte, dass dieser Schätzer eine Verallgemeinerung des gewöhnlichen Kernschätzers ist, der durch $L(\cdot) = (\cdot)^2$ erhalten wird. Der M-Kernschätzer wurde von Härdle und Gasser (1984) mit einer monotonen Scorefunktion und konstantem Skalenparameter g_n eingeführt. Sie liefern einen – allerdings nur eingeschränkt gültigen – Konsistenzbeweis und zeigen die Minimax-Eigenschaft. Chu et al. (1998) wählen eine Dichte als (zurückfallende) Scorefunktion und lassen $g_n \to 0$. Dabei stellen sie eine gute sprungerhaltende Eigenschaft des Schätzers fest. Allerdings ist, wie in dieser Arbeit gezeigt wird, der Schätzer unter den Annahmen von Chu et al. (1998) nicht konsistent.
In dieser Arbeit werden vollständige Beweise sowohl für monotone als auch zurückfallende Schätzer geführt. Außerdem wird gezeigt, dass der monotone M-Kernschätzer asymptotisch robust ist, während der zurückfallende M-Kernschätzer, so wie er von Chu et al. eingeführt wurde, nicht diese Eigenschaft besitzt. Auf der anderen Seite ist, wie in dieser Arbeit bewiesen wird, ist der zurückfallende M-Kernschätzer konsistent dicht an Diskontinuitäten ("sprungerhaltend"), wenn das Rauschen nicht zu stark ist. Diese Eigenschaft hat wiederum der monotone M-Kernschätzer nicht.
Wenn wir allerdings den Skalenparameter g_n konstant lassen, ist sogar der zurückfallende M-Kernschätzer robust, ohne seine sprungerhaltende Eigenschaft zu verlieren.
Da Unstetigkeiten im zweidimensionalen Regressionsmodell viele verschiedene Formen annehmen können, wird das Modell einer zweidimensionalen Regressionsfunktion mit einer eindimensionalen Menge von Unstetigkeitsstellen mit Hilfe eines differentialgeometrischen Ansatzes formalisiert. Es wird gezeigt, dass der zurückfallende M-Kernschätzer sogar spitze Ecken erhält, was in Kombination mit der qualitativen asymptotischen Robustheit eine einzigartige Eigenschaft ist.
In Simulationen stellte sich heraus, dass der zurückfallende M-Kernschätzer nicht robust gegen Ausreißer ist. Daher wird der Getrimmte M-Kernschätzer eingeführt, der die eckenerhaltende Eigenschaft des zurückfallenden M-Kernschätzers mit der Robustheit des LTS-Schätzers gegen Ausreißer vereinigt.
In zweidimensionalen Simulationen werden die Schätzer abschließend verglichen.