

Nächste-Nachbarn-Verfahren zur Reservierung für Einzelschäden

Jens M. Dittmer

Universität Hamburg, Department Mathematik

Hamburg, 16. November 2006
ASTIN-Herbsttagung



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

Problemstellung

Situation:

- Heterogenes Risikokollektiv
- Gesamtschadenssumme dominiert von Einzelschäden

Ziel:

- Prognose der Spätschäden (zu bereits gemeldeten Ereignissen)

Voraussetzung:

- Daten der großen Einzelschäden der vorhergehenden Jahre vorliegend
- Schadenverläufe jeweils unterteilt nach Abwicklungsjahren

Wahl des Reservierungsverfahrens

- Gebräuchlichste Methoden verwenden aggregierte Kollektivdaten
 - ▶ Verlust der zusätzlichen Information über individuelle Schadenverläufe

Vorschlag von *T. Mack (2002)*:

- Vergleiche gegebenen Einzelschadenverlauf mit früheren Schäden
 - ▶ Fortsetzung gemäß *ähnlicher* Vorbilder, den *nächsten Nachbarn*



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

Das Vorgehen

- Spezifizierung der betrachteten Methoden
- Empirischer Vergleich
- k -nächste-Nachbarn-Verfahren: Verteilung des Schätzfehlers
- Fazit zur praktischen Anwendung



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

Betrachtete Reservierungsverfahren

- Als Vergleichsmaßstab: Chain-Ladder-Verfahren
- Nächste-Nachbarn-Methoden
 - Ein Vergleichsschaden herangezogen
 - Mehrere Vergleichsschaden herangezogen



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

Chain-Ladder-Verfahren

Bezeichnungen:

- C_{ik} : Kumulierte Portfolio-Gesamtschäden der Abwicklungsjahre 1 bis k aus Anfalljahr i
- Beobachtet: C_{ik} für $i + k \leq I + 1$
(I : aktuelles Kalenderjahr)

Chain-Ladder-Verfahren

- Schätzung multiplikativer Zuwachsfaktoren

$$\hat{f}_k = \frac{\sum_{i=1}^{l-k} C_{i,k+1}}{\sum_{i=1}^{l-k} C_{ik}}$$

- ▶ Prognose zukünftiger Schadenstände

$$\widehat{C}_{ik} := C_{i,l+1-i} \cdot \hat{f}_{l+1-i} \cdot \dots \cdot \hat{f}_{k-1}$$



Chain-Ladder-Verfahren

Bei Einzelschäden:

- ▶ Anwendung der geschätzten Zuwachsfaktoren \hat{f}_k auf einzelne Schadenverläufe



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

Nächste-Nachbarn-Verfahren

Ein Vergleichsschaden

Bezeichnungen:

- $x \in \mathbb{R}^p$: Betrachteter Schaden (Schadenstände der bislang beobachteten p Abwicklungsjahre)
- $Y \in \mathbb{R}$: Zugehörige zu prognostizierende Spätschäden des Folgejahres
- (X_i, Y_i) , $1 \leq i \leq n$: Bekannte Schadenverläufe potenzieller Vorbilder
 - $X_i \in \mathbb{R}^p$: Schadenstände der ersten p Abwicklungsjahre
 - $Y_i \in \mathbb{R}$: Schadenzuwachs im $(p + 1)$ -ten Abwicklungsjahr

Nächste-Nachbarn-Verfahren

Ein Vergleichsschaden

Wahlmöglichkeiten betreffen...

- Bestimmung der *Ähnlichkeit* der \mathbb{R}^P -wertigen Schadenverläufe
- Festlegung einer geeigneten Fortsetzung des Schadenverlaufs (mittels des gefundenen Vorbilds)

Abstand zwischen Schadenverläufen

- Festlegung einer Norm auf \mathbb{R}^p
- Euklidische Norm:

$$\|u\| := \sqrt{\sum_{i=1}^p u_i^2}, \quad u = (u_1, \dots, u_p) \in \mathbb{R}^p$$

- ▶ ...gewichtet alle Abwicklungsjahre gleich
- Absolutbetrag der p -ten Komponente:

$$\|u\| := |u_p|, \quad u = (u_1, \dots, u_p) \in \mathbb{R}^p$$

- ▶ ...berücksichtigt nur den letzten bekannten Schadenstand

Festlegung der Fortsetzung

Additiver oder multiplikativer Ansatz

Ähnlichster Schaden zu $x \in \mathbb{R}^p$ sei $(\tilde{X}, \tilde{Y}) \in \mathbb{R}^p \times \mathbb{R}$

- Verwendung des *additiven Zuwachses*

$$\hat{Y}_{(\text{add})} := \tilde{Y}$$

- Verwendung des *multiplikativen Zuwachses*

$$\hat{Y}_{(\text{mult})} := x_p \cdot \frac{\tilde{Y}}{\tilde{X}_p}$$



Nächste-Nachbarn-Verfahren

Mehrere Vergleichsschäden

Verfeinerung des Verfahrens (*T. Mack, 2002*):

- Mehrere ($k \in \{1, \dots, n\}$) Vorbilder zur Schätzung heranziehen
- Gewichtete Mittelung nach *Ähnlichkeit* zum betrachteten Schaden



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

Nächste-Nachbarn-Verfahren

Mehrere Vergleichsschäden

Bezeichnungen wie eben:

- $x \in \mathbb{R}^p$: Betrachteter Schaden (Schadenstände der bislang beobachteten p Abwicklungsjahre)
- $Y \in \mathbb{R}$: Zugehörige zu prognostizierende Spätschäden
- (X_i, Y_i) , $1 \leq i \leq n$: Bekannte Schadenverläufe potenzieller Vorbilder

Nächste-Nachbarn-Schätzer

$$\hat{Y}_{\text{knn}} := \frac{\sum_{i=1}^n K\left(\frac{X_i - x}{R(k)}\right) \cdot Y_i}{\sum_{i=1}^n K\left(\frac{X_i - x}{R(k)}\right)}$$

- $R(k)$: Abstand von x zum k -ten Nachbarn unter den X_i
- $K(u) > 0$ nur für $\|u\| < 1$
- ▶ Mittelung der $k - 1$ nächsten Nachbarn gemäß relativem Abstand zu x



Festlegung der Gewichtsfunktion

Modifizierter *Epanechnikov*-Kern

$$K(u) := \kappa \cdot ((1 - \|u\|^2) + \delta) \cdot \mathbf{1}_{[0,1]}(\|u\|), \quad \delta > 0$$

Dabei:

- Norm $\|\cdot\|$ auf \mathbb{R}^p beliebig
- κ : Normierungskonstante, so dass $\int K(u) du = 1$



Festlegung der Gewichtsfunktion

Modifizierter *Epanechnikov*-Kern

$$K(u) := \kappa \cdot ((1 - \|u\|^2) + \delta) \cdot 1_{[0,1)}(\|u\|), \quad \delta > 0$$

Dann:

- ▶ $K(u) > 0 \Leftrightarrow \|u\| < 1$
 - ▶ Mittelung über $k - 1$ Vorbilder
- ▶ Gewichte $K\left(\frac{X_i - x}{R(k)}\right)$ umso größer, je geringer $\|X_i - x\|$



Empirischer Vergleich der Verfahren

Datenbeispiel aus Haftpflichtversicherung:

- Abwicklungsverläufe von 1346 Großschadenereignissen (1973-2004) eines deutschen Versicherungsunternehmens
- Betrachtung *bezahlter* Schäden (inflationsbereinigt)



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

Backtesting

- Für jeden beobachteten Einzelschaden
 - ab zweitem Abwicklungsjahr
 - ab zweitem Anfalljahr
- Prognose auf Basis der bis zum vorherigen Jahr bekannten Information
- Vergleich von Prognosen und tatsächlich eingetretenen Schäden



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

Backtesting

Reserveschätzung mit Backtesting für

- Schäden $m \in \{1, \dots, n_{ik}\}$
- aus Schadenanfalljahr $i \in \{2, \dots, 31\}$
- und Abwicklungsjahr $k \in \{2, \dots, 33 - i\}$

(n_{ik} : Anzahl Schäden aus Anfalljahr i und Abwicklungsjahr k)



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

Backtesting

- Y_{ikm} : Beobachteter Schaden m aus Anfalljahr i und Abwicklungsjahr k
- \hat{Y}_{ikm} : Zugehöriger prognostizierter Schaden
- Analyse der Residuen $\hat{Y}_{ikm} - Y_{ikm}$



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

Backtesting

Beurteilungskriterien:

- Aggregation der quadrierten Residuen
 - auf Einzelschadenebene
 - gruppiert nach Abwicklungs- und Anfalljahr
- Quantile der empirischen Verteilung aller $|\hat{Y}_{ikm} - Y_{ikm}|$



Beurteilungskriterien beim Backtesting

- Aggregierte quadrierte Einzelschadenresiduen:

$$SSR_{\text{Einzel}} := \sum_{i=2}^{31} \sum_{k=2}^{33-i} \sum_{m=1}^{n_{ik}} (\hat{Y}_{ikm} - Y_{ikm})^2$$

- ▶ Beurteilung der Vorhersage einzelner Schäden



Beurteilungskriterien beim Backtesting

- Aggregierte quadrierte Jahresresiduen:

$$SSR_{\text{Jahr}} := \sum_{i=2}^{31} \sum_{k=2}^{33-i} \left(\sum_{m=1}^{n_{ik}} \hat{Y}_{ikm} - \sum_{m=1}^{n_{ik}} Y_{ikm} \right)^2$$

- ▶ Beurteilung der Vorhersage der kumulierten Schäden bei festgehaltenem Anfall- und Abwicklungsjahr

Vorhersagegenauigkeit: Ein empirischer Vergleich

Betrachtete Verfahren (Übersicht):

- Chain-Ladder:

- $\hat{Y}_{CL} := x_p \cdot \hat{f}_p$

- Nächste-Nachbarn (ein Vergleichsschaden):

- $\hat{Y}_{(add)} = \tilde{Y}$

- $\hat{Y}_{(mult)} = x_p \cdot \frac{\tilde{Y}}{\tilde{X}_p}$

- k-Nächste-Nachbarn:

- $\hat{Y}_{knn} := \frac{\sum_{i=1}^n K\left(\frac{X_i - x}{R(k)}\right) \cdot Y_i}{\sum_{i=1}^n K\left(\frac{X_i - x}{R(k)}\right)}$



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

Vorhersagegenauigkeit: Ein empirischer Vergleich

Zur Abschwächung von Kalenderjahreffekten:

- Jeweils nur Berücksichtigung von (maximal) 10 Jahre zurückliegenden Schadenverläufen
 - ▶ Verbesserung aller betrachteten Verfahren

Methoden mit einem Vergleichsschaden

Varianten:

- Additive oder multiplikative Fortsetzung
- Wahl der Norm $||\cdot||$

- *Euklidisch*: $||u|| = \sqrt{\sum_{i=1}^p u_i^2}$

- *Letzte Komponente*:
 $||u|| = |u_p|$



Methoden mit einem Vergleichsschaden: Varianten

Vergleich mit dem Chain-Ladder-Verfahren

Verfahren	Norm	SSR_{Einzel}	SSR_{Jahr}
Chain-Ladder		$3,09 \cdot 10^{11}$	$4,06 \cdot 10^{11}$
\hat{Y}_{add}	Euklidisch	$4,85 \cdot 10^{11}$	$5,48 \cdot 10^{11}$
\hat{Y}_{add}	Letzte Komp.	$5,09 \cdot 10^{11}$	$5,41 \cdot 10^{11}$
\hat{Y}_{mult}	Euklidisch	$5,57 \cdot 10^{11}$	$6,44 \cdot 10^{11}$
\hat{Y}_{mult}	Letzte Komp.	$5,19 \cdot 10^{11}$	$5,61 \cdot 10^{11}$



Methoden mit einem Vergleichsschaden

Ergebnisse:

- Chain-Ladder-Verfahren noch deutlich überlegen
- Vergleich der Nächste-Nachbarn-Methoden:
 - ▶ Additive Fortsetzung vorzuziehen
- Unter diesen: Wahl von $\|\cdot\|$ nachrangig, euklidische Norm leicht präziser

k -Nächste-Nachbarn-Verfahren

Vergleich \hat{Y}_{knn} mit \hat{Y}_{CL} :

- Berechnung SSR_{Einzel} und SSR_{Jahr} für unterschiedliche Werte von k
- Verwendung euklidischer Norm
- Beachte: $\hat{Y}_{knn} = \hat{Y}_{(add)}$ im Fall $k = 1$



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

k -Nächste-Nachbarn-Verfahren

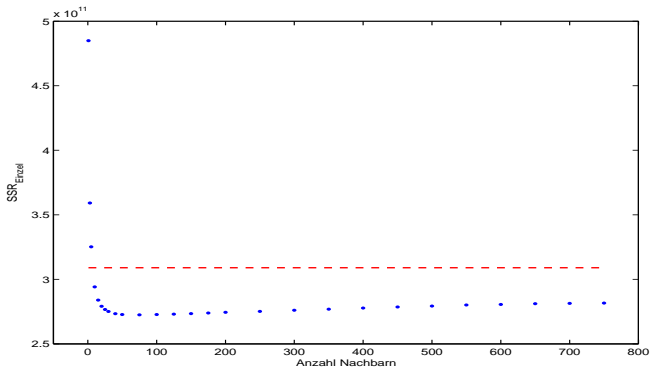


Abbildung: SSR_{Einzel} bei Chain-Ladder- (rot) und k -Nächste-Nachbarn-Verfahren (blau)

k -Nächste-Nachbarn-Verfahren

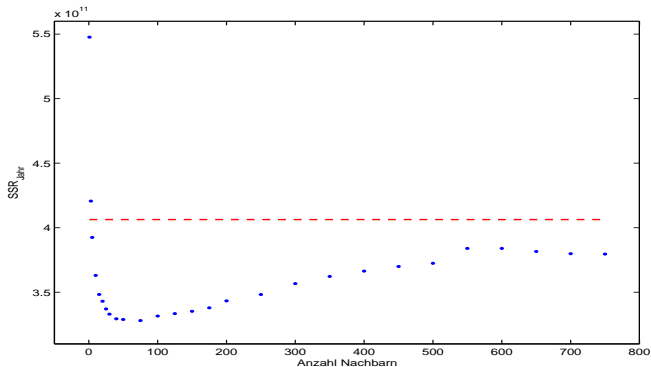


Abbildung: SSR_{Jahr} bei Chain-Ladder- (rot) und k -Nächste-Nachbarn-Verfahren (blau)

k -Nächste-Nachbarn-Verfahren

Aggregierte quadrierte Residuen

Erkenntnisse:

- Für breite Spanne an Werten für k :
Mit k -Nächste-Nachbarn-Verfahren präzisere Prognose der Einzelschäden als mit Chain-Ladder
- Genauere Vorhersage auch der aggregierten Schäden bei festgehaltenem Anfall- und Abwicklungsjahr

k -Nächste-Nachbarn-Verfahren

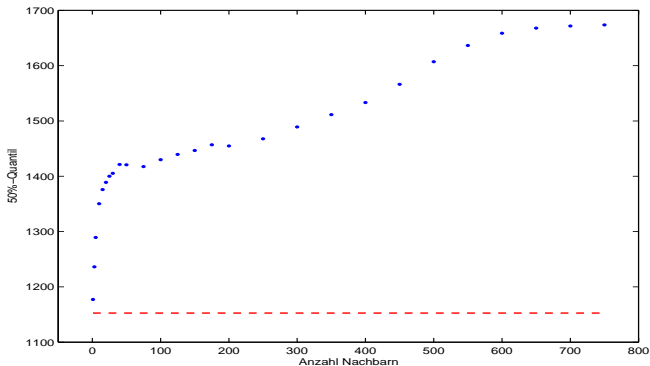


Abbildung: 50%-Quantil der absoluten Residuen bei Chain-Ladder- (rot) und k -Nächste-Nachbarn-Verfahren (blau)

k -Nächste-Nachbarn-Verfahren

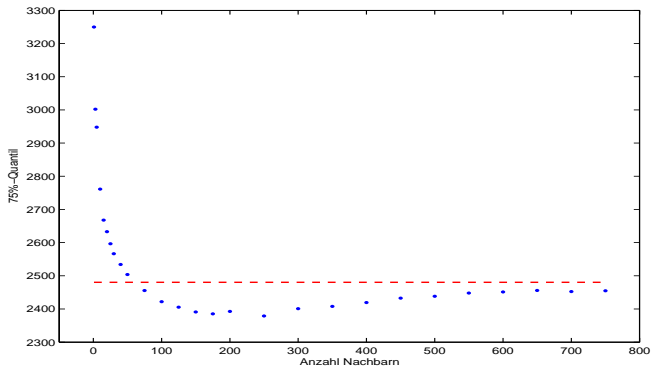


Abbildung: 75%-Quantil der absoluten Residuen bei Chain-Ladder- (rot) und k -Nächste-Nachbarn-Verfahren (blau)

k -Nächste-Nachbarn-Verfahren

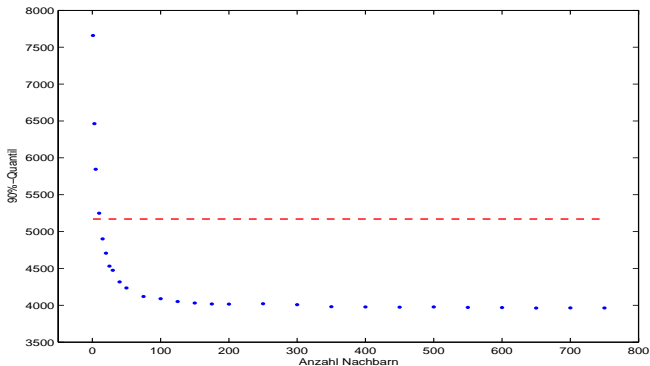


Abbildung: 90%-Quantil der absoluten Residuen bei Chain-Ladder- (rot) und k -Nächste-Nachbarn-Verfahren (blau)

k-Nächste-Nachbarn-Verfahren

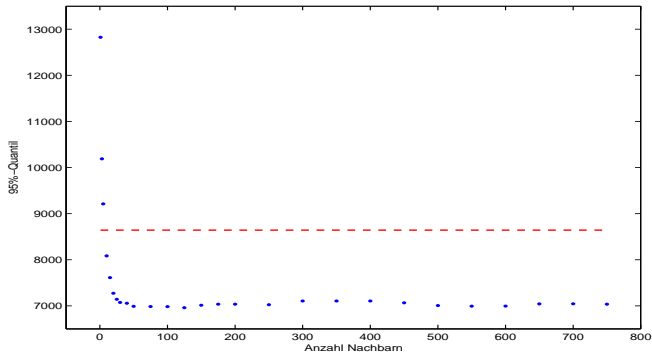


Abbildung: 95%-Quantil der absoluten Residuen bei Chain-Ladder- (rot) und k-Nächste-Nachbarn-Verfahren (blau)

k -Nächste-Nachbarn-Verfahren

Quantile der absoluten Residuen

Erkenntnisse:

- Ein Großteil (50%) der Schäden wird durch das Chain-Ladder-Verfahren etwas genauer vorhergesagt
- Bei Chain-Ladder jedoch wesentlich häufiger starke Abweichungen (insbesondere durch Unterschätzung von Großschäden)



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

k -Nächste-Nachbarn-Verfahren

Weitere Ergebnisse:

- Prognosegüte weitestgehend unempfindlich gegenüber Variation der Norm der Art

$$\|u\| := \sqrt{\sum_{i=1}^p a_i u_i^2}, \quad a_i \geq 0$$

- Vorteile des k -Nächste-Nachbarn-Verfahrens speziell in ersten zehn Abwicklungsjahren

Theoretische Untersuchung der k -Nächste-Nachbarn-Methode

Vorgehen:

- Interpretation Reserveschätzung als nichtparametrischer Regressionsansatz
- Wahl der Anzahl an Nachbarn
- Erwartete Spätschäden: Schätzfehler und Konfidenzintervalle
- Prognosefehler



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

Stochastische Modellierung der Reserveschätzung

Nichtparametrischer Regressionsansatz:

- Modell: $Y = m(X) + \sigma(X) \cdot \varepsilon$
 - $X \in \mathbb{R}^p$: Schadenstände der ersten p Abwicklungsjahre
 - $Y \in \mathbb{R}$: Schadenzuwachs im $(p + 1)$ -ten Abwicklungsjahr
 - $\varepsilon \in \mathbb{R}$: Fehler mit $E(\varepsilon) = 0$ und $\text{Var}(\varepsilon) = 1$ (unabhängig von X)
-
- ▶ $m(x) = E(Y|X = x)$
 - ▶ $\sigma^2(x) = \text{Var}(Y|X = x)$



Nichtparametrischer Regressionsansatz

Modell: $Y = m(X) + \sigma(X) \cdot \varepsilon$

- Annahme: $m, \sigma : \mathbb{R}^p \rightarrow \mathbb{R}$ glatte Funktionen
- Idee: Interpretation von \hat{Y}_{knn} als Schätzer für $m(x)$
($x \in \mathbb{R}^p$ fest)



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

Schätzung von $m(x) = E(Y|X = x)$

► Nächste-Nachbarn-Schätzer

$$m_n(x) := \frac{\sum_{i=1}^n K\left(\frac{X_i - x}{R_n(k_n)}\right) \cdot Y_i}{\sum_{i=1}^n K\left(\frac{X_i - x}{R_n(k_n)}\right)}$$

Dabei:

- $(X_1, Y_1), \dots, (X_n, Y_n), (X, Y)$ iid
- $k_n \in \{1, \dots, n\}$: Anzahl an Nachbarn
- $R_n(k_n) := \inf\{t \geq 0 \mid \sum_{j=1}^n \mathbf{1}_{\{\|X_j - x\| \leq t\}} \geq k_n\}$
(Abstand von x zum k_n -ten Nachbarn unter den X_j)

Festlegung der Anzahl an Nachbarn (k_n)

Falls k_n zu klein:

- ▶ Keine ausreichende Mittelung der Vergleichsschäden
- ▶ Varianz der Schätzung sehr groß

Festlegung der Anzahl an Nachbarn (k_n)

Falls k_n zu groß:

- ▶ Viele Vergleichsschäden X_i mit großem Abstand zu x
 - ▶ Häufig starke Abweichungen zwischen Funktionswerten $m(X_i)$ und $m(x)$
 - ▶ Systematischer Fehler bei Schätzung von $m(x)$
- ▶ Bandweite $R_n(k_n)$ sehr groß
 - ▶ Relative Abstände $\frac{\|X_i - x\|}{R_n(k_n)}$ überwiegend klein
 - ▶ Unter „ähnlichsten“ Vorbildern kaum noch differenzierte Gewichtung



Asymptotische Verteilung des Schätzfehlers

$$m_n(x) - m(x)$$

Wichtigste Voraussetzungen:

- X besitzt Lebesgue-Dichte
- Y darf weitestgehend beliebig sein (z.B. Punktmasse in 0)
- Keine Einschränkung der Wahl von $\|\cdot\|$
- $\lim_{n \rightarrow \infty} k_n = \infty, \lim_{n \rightarrow \infty} \frac{k_n}{n} = 0$
- $\eta := \lim_{n \rightarrow \infty} \sqrt{k_n} n^{-4/(4+p)} < \infty$
(z.B. $\eta = 0$)



Asymptotische Verteilung des Schätzfehlers

$$m_n(x) - m(x)$$

Wichtigste Voraussetzungen:

- $0 < N_1 \leq K(u) \leq N_2$ für alle $\|u\| < 1$
- $K(u) = 0$ für alle $\|u\| \geq 1$
- $\int K(u) du = 1$
- $\int K(u) u_\alpha du = 0$ für $\alpha \in \{1, \dots, p\}$



Asymptotische Verteilung des Schätzfehlers

$$m_n(x) - m(x)$$

Dann gilt:

$$\sqrt{k_n}(m_n(x) - m(x)) \xrightarrow{n \rightarrow \infty} \mathcal{N}\left(B(x), c \cdot \sigma^2(x) \int K^2(u) du\right)$$

Dabei:

- c : Volumen der Einheitskugel im \mathbb{R}^p bezüglich $\|\cdot\|$
- $B(x)$: Asymptotischer Bias
(verschwindet im Fall $\eta = 0$)

Asymptotische Verteilung des Schätzfehlers

$$m_n(x) - m(x)$$

- Beweis durch Bedingen nach $R_n(k_n)$, mit zentralem Grenzwertsatz von Berry-Esséen
- Techniken in Anlehnung an *Y.-P. Mack (1981)*



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

Konfidenzintervalle für $m(x) = E(Y|X = x)$

Herleitung asymptotischer Konfidenzintervalle aus
Normalapproximation

- Annahme: k_n klein genug gewählt, dass $\eta = B(x) = 0$ (systematischer Fehler vernachlässigbar)
- Konsistente Schätzung $\sigma^2(x) = \text{Var}(Y|X = x)$ durch

$$\sigma_n^2(x) := \left| \frac{\sum_{i=1}^n K\left(\frac{X_i - x}{R_n(k_n)}\right) \cdot Y_i^2}{\sum_{i=1}^n K\left(\frac{X_i - x}{R_n(k_n)}\right)} - m_n^2(x) \right|$$



Konfidenzintervalle für $m(x) = E(Y|X = x)$

Asymptotisches Konfidenzintervall zum Niveau $1 - \alpha$:

$$m_n(x) \pm z_{1-\alpha/2} \sqrt{\frac{c}{k_n} \sigma_n^2(x) \int K^2(u) du}$$

$z_{1-\alpha/2}$: $(1 - \frac{\alpha}{2})$ -Quantil der Standardnormalverteilung



Prognosefehler $m_n(x) - Y$

- Statt Schätzung von $E(Y|X = x)$ nun Prognose von Y im Blickpunkt
- Zu Schätzfehler $m_n(x) - m(x)$ kommt hinzu:
Zufällige Schwankung von Y um $E(Y|X = x)$
 - ▶ $m_n(x) - Y = m_n(x) - m(x) + \underbrace{m(x) - Y}_{=-\sigma(x) \cdot \varepsilon}$
 - ▶ Faltung von Schätzfehler und Prognoseanteil



Prognosefehler $m_n(x) - Y$

$$m_n(x) - Y = (m_n(x) - m(x)) - \sigma(x) \cdot \varepsilon$$

► Varianzzerlegung

$$\begin{aligned}\text{Var}(m_n(x) - Y) &= \text{Var}(m_n(x) - m(x)) + \sigma^2(x) \\ &\approx \sigma^2(x) \left(1 + \frac{c}{k_n} \int K^2(u) du \right)\end{aligned}$$

Dabei:

- $c \int K^2(u) du \leq 2$ bei Euklidischer Norm und modifiziertem Epanechnikov-Kern
 - Einfluss des Schätzfehlers für wachsende k_n schnell gering

Schätz- und Prognosefehler

Bei Zusammenfassung mehrerer *gleichartiger* Schäden:

- Prognoseanteile $Y - E(Y|X = x) = \sigma(x) \cdot \varepsilon$
diversifizierbar, da voneinander unabhängig
- Verwendete Vorbilder oft dieselben
 - ▶ Durchschnittlicher Schätzfehler nicht systematisch verringert



Schätz- und Prognosefehler

Konsequenz für Beurteilung Prognosefehler:

- ▶ Dominanz des Prognoseanteils bei Einzelschadenvorhersage
- ▶ Dominanz des Schätzfehlers bei Zusammenfassung gleichartiger Schäden



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

Praktische Anwendung des k -Nächste-Nachbarn-Verfahrens

- k -Nächste-Nachbarn-Verfahren vorteilhaft in Situationen heterogener Risikokollektive (dominiert von Großschäden)
- Präzisere Vorhersage als mit Chain-Ladder-Verfahren insbesondere in früheren Abwicklungsjahren

Praktische Anwendung des k -Nächste-Nachbarn-Verfahrens

- Ziel: Vorhersage eines zukünftigen, *mehrjährigen* Abwicklungsverlaufes (Jahre $p + 1$ bis L)

Mehrfache Durchführung des Verfahrens:

- Prognose der Schäden getrennt für jedes einzelne Abwicklungsjahr
- Zusätzlich: Vorhersage der Summe zukünftiger Schäden in einem einzigen Schritt
- ▶ Konfidenzintervalle erhältlich für erwartete Schäden sowohl in den einzelnen Perioden als auch in der gesamten Betrachtungszeit

Wahl der Anzahl an Nachbarn k_n

- Für breite Spanne an Werten für k_n vergleichsweise genaue Prognose
- Im betrachteten Beispiel beste Ergebnisse, wenn $k_n \geq 10$ und maximal 10-20% der verfügbaren Daten verwendet
- Genauerer Aufschluss durch Datenanalyse mit Backtesting oder Cross Validation



DAV




DEUTSCHE
AKTUARVEREINIGUNG e.V.

Schwierigkeiten

Offene Probleme:

- Prognosefehler für (im Portfolio) aggregierte Schäden ergibt sich nicht aus Modell
 - ▶ Empirische Untersuchung: Auch hier verhältnismäßig genaue Prognosen des k -Nächste-Nachbarn-Verfahrens
- Normalapproximation des Schätzfehlers:
Voraussetzung Lebesgue-Dichte an Schadenverlauf $X \in \mathbb{R}^p$ der ersten p Abwicklungsjahre zu einschränkend
 - ▶ Allgemeineres Resultat scheint möglich



-  Thomas Mack (2002)
Schadenversicherungsmathematik.
2. Auflage; Verlag Versicherungswirtschaft, Karlsruhe.
-  Yue-Pok Mack (1981)
Local properties of k -nn regression estimates.
SIAM Journal on Algebraic and Discrete Methods 2,
311-323.
-  Jens M. Dittmer (2006)
*Nearest-Neighbour Methods for Reserving with respect to
Individual Losses.*
Blätter der DGVFM XXVII, 647-664.

