Beiträge zur Stichproben-Theorie
und
Anmerkungen zu einem Bootstrap-Grenzwertsatz

Dissertation zur Erlangung des akademischen Grades
„Doktor der Philosophie“

vorgelegt von Dipl.-Math. Hans-Ulrich Hingst
am Fachbereich Politik- und Sozialwissenschaften
der Freien Universität Berlin

3. Mai 2003

1. Gutachter: Prof. Dr. H. Skarabis

2. Gutachter: Prof. Dr. J. Gordesch

Datum der Disputation: 17.07.2003

Einleitung und Zusammenfassung

Die vorliegende Arbeit behandelt drei Problembereiche aus der mathematischen Statistik.
Die beiden ersten Kapitel liefern Beiträge zur „traditionellen“ Stichproben-Theorie, während das letzte Kapitel mit F.Strobl's Bootstrap-Grenzwertsatz für Statistische Funktionale befaßt ist.

Im I.Kapitel wird eine Ungleichung entwickelt, die eine Beziehung herstellt zwischen der Abhängigkeit der Stichprobenvariablen beim n-maligen Ziehen ohne Zurücklegen aus Ω_N und der Größe von Stichprobe und Grundgesamtheit, wenn die Grundgesamtheit Ω_N aus N Objekten gleicher Auswahlchance besteht. Mit $\mathcal{F}(N)$ wird die Verteilung des Merkmals X in Ω_N bezeichnet.
Es wird eine Stichprobe vom Umfang n aus Ω_N gezogen, und an jedem Objekt die zugehörige Ausprägung des Merkmals X gemessen.
Ausgangspunkt für die Entwicklung der o.g. Ungleichung ist ein unveröffentlichtes Skriptum von einem Vortrag auf der gemeinsamen Jahrestagung der Polnischen und der Deutschen Statistischen Gesellschaft aus dem Jahre 2000. Hierin hat der Autor eine Abschätzung vorgestellt, die eine Vor-Version der Ungleichung darstellt; es war aber noch keine Angabe über deren Präzision möglich. Nach der Tagung entstand das Vorhaben, die Arbeit daran fortzuführen und wesentlich zu vertiefen. Ein neuer Ansatz liegt hier vor, und es ist nun eine Aussage über die Präzision der mit der Ungleichung zum Ausdruck kommenden Abschätzung möglich: Sind die N gemessenen Werte an den Objekten von Ω_N paarweise verschieden, dann gilt in der Ungleichung das Gleichheitszeichen.
Vom betrachteten Merkmal X wird zwar die Reellwertigkeit, aber keine spezielle Meßbarkeit vorausgesetzt. Daher wurde auch nicht mit Verteilungsfunktionen, sondern mit Wahrscheinlichkeitsfunktionen argumentiert.

Das n-malige Ziehen ohne Zurücklegen enthält in folgendem Sinne ein größeres Potential für die Informationsgewinnung über $\mathcal{F}(N)$ gegenüber dem n-maligen Ziehen mit Zurücklegen: Beim Ziehen ohne Zurücklegen sind neben den paarweise verschiedenen Stichprobendaten auch deren absolute Häufigkeiten zur Informationsgewinnung über $\mathcal{F}(N)$ nutzbar; tritt ein Stichprobenwert k-fach auf, weiß man, dass seine relative Häufigkeit in $\mathcal{F}(N)$ mindestens k / N ist; ein $k > 1$ liefert daher Informationen über $\mathcal{F}(N)$. Beim Ziehen ohne Zurücklegen liefert jeder Stichprobenwert Informationen über $\mathcal{F}(N)$.
Beim Ziehen mit Zurücklegen liefern nur die paarweise verschiedenen Stichprobendaten Informa-

tionen über $\mathcal{F}(N)$. Tritt hier ein Stichprobenwert k-fach auf mit $k > 1$, liefert das die selbe Information über $\mathcal{F}(N)$ wie sein 1-faches Auftreten: Man weiß nur, dass die relative Häufigkeit dieses Wertes mindestens $1 / N$ ist. Beim Ziehen mit Zurücklegen liefern also Daten, die schon einmal aufgetreten sind, keine neuen Informationen über $\mathcal{F}(N)$.

Dieses größere Potential wird z.B. von der Stichprobenfunktion $\overline{X}_n$ für die Schätzung von $\mu = E(X)$ genutzt. Direkt nachrechenbar ist die bekannte Tatsache, dass der Mittlere Quadratische Fehler von $\overline{X}_n$ beim Ziehen ohne Zurücklegen aus Ω_N für alle $n > 1$ kleiner ist als beim Ziehen mit Zurücklegen aus Ω_N.

Betreibt man für den Fall des Ziehens mit Zurücklegen einen zusätzlichen „Selektions-Aufwand“, läßt in einer Schätzfunktion die „nicht benötigten“ (weil schon aufgetretenen) Daten weg und berücksichtigt nur die paarweise verschiedenen Daten, kann das die Effektivität der Schätzfunktion erhöhen - quasi als „Lohn“ für den Zusatz-Aufwand. D.Basu (1958) hat die μ-Schätzung durch das Arithmetische Mittel untersucht. Er konnte zeigen, dass die Varianz dieser Schätzfunktion beim Ziehen mit Zurücklegen kleiner / gleich jener beim Ziehen ohne Zurücklegen ist, wenn man im ersten Fall den Mittelwert nur aus den paarweise verschiedenen Daten und im zweiten Fall den Mittelwert aus allen Daten bildet.

Beim n-maligen Ziehen mit Zurücklegen aus Ω_N mögen m verschiedene Objekte getroffen werden. Es kann folgender Fall eintreten: Zwischen den Einzelzügen vergeht notwendig Zeit, so dass man bei hinreichend genauer Messung für das selbe Objekt verschiedene Meßwerte erhält. Dann hat man es also in Wahrheit nicht mit einer Stichprobe vom Umfang n zu tun, sondern mit m Zeitreihen. Diese Tatsache bleibt verborgen, wenn nicht zusätzlich zu den Meßwerten auch die gezogenen Objekte notiert wurden (z.B. mangels vorhandener Kennzeichnung).

Das Ziehen mit Zurücklegen aus Ω_N kann also spezielle Fehleinschätzungen zur Folge haben, die beim Ziehen ohne Zurücklegen aus Ω_N nicht auftreten.

Den beschriebenen Vorteilen des Ziehens ohne Zurücklegen aus Ω_N steht ein gewichtiger Tatbestand gegenüber: Die zugehörigen Stichproben-Variablen sind stochastisch abhängig.

Dass dies ein Nachteil sein kann, offenbart sich z.B. dann, wenn für kompliziertere Schätzfunktionen Verteilungsparameter berechnet werden sollen. „Traditionelle“ Methoden verlangen in der Regel die Unabhängigkeit der Stichproben-Variablen. Delta-Methode oder Varianten des Bootstrap-Verfahrens stoßen ebenfalls an gewisse Grenzen, denn hier wird im Falle abhängiger Stichproben-Variablen vorausgesetzt, dass die zu analysierende Stichprobenfunktion ein Arithmetisches Mittel oder zumindest eine hinreichend glatte Funktion eines Arithmetischen Mittels darstellt.

In den sonstigen Fällen werden in der statistischen Praxis für die Analyse von Stichproben-

funktionen beim Ziehen ohne Zurücklegen aus Ω_N trotzdem Methoden angewendet, welche an sich die Unabhängigkeit der Stichproben-Variablen voraussetzen. Dies wird dann als gerechtfertigt angesehen, wenn N relativ groß ist, und der Stichprobenumfang n dazu relativ klein ist. Hintergrund dafür ist der von W.Feller (1950) formulierte Vergleich der Auswahl-Wahrscheinlichkeiten von n Objekten aus Ω_N, wenn einerseits mit, und andererseits ohne Zurücklegen gezogen wird. Um den Abhängigkeitsgrad der Stichproben-Variablen $X_1, \ldots, X_n$ zu beurteilen, ist eine i.a. relativ umfangreiche Untersuchung vonnöten. Sei m die Anzahl der paarweise verschiedenen Merkmalswerte, die an den Objekten von Ω_N gemessen werden. Mit P_{ZoZ} wird die gemeinsame Wahrscheinlichkeitsfunktion der Stichproben-Variablen beim n-maligen Ziehen ohne Zurücklegen aus Ω_N bezeichnet ($n \leq N$). Sei $D := P_{ZoZ}\left(X_1 = x_1, \ldots, X_n = x_n \right) - \prod_{k=1}^{n} P\left(X_k = x_k \right)$.

Erst der Nachweis, dass der Abstand $|D|$ klein ausfällt für alle möglichen Stichprobenrealisationen $(x_1, \ldots, x_n)$, läßt den Schluß zu, dass die Abhängigkeit der $X_1, \ldots, X_n$ nur gering ist. Es wird gezeigt, dass man lediglich im Falle $m = N$ davon ausgehen kann, dass der o.g. Vergleich der Auswahl-Wahrscheinlichkeiten mit dieser Untersuchung gleichwertig ist.

Aus der hier entwickelten Ungleichung erhält man eine nur von n und N abhängige Abschätzung von $|D|$. Daraus ergibt sich für jede Anzahl $m \in \{1, \ldots N\}$:

Zu gegebenem Stichprobenumfang n wird die Abhängigkeit der Stichproben-Variablen beim Ziehen ohne Zurücklegen aus Ω_N beliebig klein, wenn N hinreichend groß ist.

Es wird ein Grenzwertsatz angegeben, der besagt, dass der Abstand $|D|$ zu gegebenem n mit $N \to +\infty$ gleichmäßig gegen Null konvergiert.

Hieraus ergibt sich unmittelbar ein weiterer Grenzwertsatz über den Abstand zwischen den Wahrscheinlichkeitsfunktionen der m-dimensionalen Hypergeometrischen Verteilung und der m-fachen Multinomialverteilung. Es werden Bedingungen angegeben, unter denen dieser Abstand mit $N \to +\infty$ gleichmäßig gegen Null konvergiert. Dieser Grenzwertsatz ist eine Verschärfung und Erweiterung eines von W.Feller (1953) speziell für $m = 2$ angegebenen Satzes.

Außerdem wird der o.g. Vergleich der Auswahl-Wahrscheinlichkeiten W.Feller's hier mathematisch präzisiert. Dieser Vergleich ergibt sich aus dem Identitätsfall der Ungleichung, wenn die Zuordnung zwischen Ω_N und der Menge der Merkmalswerte bijektiv und also $m = N$ ist.

Zu gegebenem $N > 4$ hat man für $n > 2$ die folgende Interpretation:

Die Abhängigkeit von n Entnahmen ohne Zurücklegen aus Ω_N nimmt ab mit wachsendem n; sie ist am kleinsten, wenn $n = N-1$ ist. Die Abhängigkeit der Entnahmen bei Vollerhebung ist also kurioserweise größer als jene bei Zurücklassen eines einzigen Objektes.

In der allgemeinen Fassung der Ungleichung aus dem I.Kapitel bleibt zunächst noch (λ_N / N), die maximale relative Häufigkeit des Modalwertes, als nach oben abzuschätzende Unbestimmte. Die Abschätzung von (λ_N / N) wird im II.Kapitel vorgenommen.

Die Literatur-Recherchen des Autors ergaben, dass offenbar keine Veröffentlichungen zu dieser Thematik vorliegen. Für infragekommende Schätzfunktionen fand sich auch kein Zugang zur vorhandenen Theorie der „extreme statistics“ ; die nötigen Voraussetzungen sind im vorliegenden Fall nicht erfüllt.

Es wurde der folgende Weg beschritten.

Die N Merkmalswerte, auf die sich die maximale relative Häufigkeit (λ_N / N) bezieht, werden aufgefaßt als Realisationen von N Zufallsgrößen $Z_1, \ldots, Z_N$, von denen vorausgesetzt wird, dass sie zu einer Folge $\mathcal{X} = (Z_k)_{k \in \mathbb{N}}$ von unabhängigen, identisch verteilten, kardinal meßbaren Zufallsgrößen gehören. Die Verteilung der Z_k $(k \in \mathbb{N})$ wird mit $\mathcal{F}$ bezeichnet, $\mathcal{F}$ besitzt die Verteilungsfunktion F.

Sei F_N die zu den Daten gehörige Realisation der Empirischen Verteilungsfunktion IF_N.

Mit einer geeigneten Schrittweite $2 \cdot h_N$ gilt die folgende Identität:

$$\lambda_N / N \;=\; \sup_{t \in \mathbb{R}} \left\{ F_N(t + h_N) \;-\; F_N(t - h_N) \right\} \;.$$

Es werden mehrere Schätz-Versionen für (λ_N / N) angegeben, die sich einerseits aus zwei Hauptversionen der Abschätzung der Kolmogorov / Smirnov -Distance $\mathbb{D}_N$ und andererseits aus den Annahmen über die Verteilungsfunktion F und die Schrittweite $2 \cdot h_N$ ergeben. Während Annahmen über die Schrittweite in erster Linie die (λ_N / N) - Abschätzung nach unten betreffen, sind die Voraussetzungen für F maßgeblich für die Erreichbarkeit der Abschätzung nach oben.

Für den wichtigen Fall einer stückweise stetigen Verteilungsfunktion F werden obere und untere Schranken für (λ_N / N) angegeben, welche beide unter bestimmten Bedingungen gegen die maximale Sprunghöhe von F konvergieren.

Eine leichte Modifikation dieses Grenzwertsatzes ermöglicht eine entsprechende Aussage für den Fall einer stetigen Verteilungsfunktion F. Da bei stetigem F aber in der Regel $\lambda_N = 1$ auftritt, hat dieser Fall für die (λ_N / N) - Abschätzung nur eine Randbedeutung. Als Ergänzung werden (λ_N / N) - Abschätzungen für Lipschitz-stetiges F sowie für absolut stetiges F mit gleichmäßig stetiger Dichte angegeben - letztere Abschätzung basiert auf einem Theorem von E.A.Nadaraya (1965).

Abschließend wird eine Daten-gestützte (λ_N / N) - Abschätzung konstruiert für den Fall, dass

über F nichts bekannt ist.
Zu allen vorgestellten Versionen der (λ_N / N) - Abschätzung werden stets Angaben über die Approximationsgenauigkeit zu vorgegebenem $N \in \mathbb{N}$ gemacht.

Im III.Kapitel wird zunächst ein Grenzwertsatz von F.Strobl (1995) aus der Bootstrap-Theorie vorgestellt. Für eine Verteilung $\mathcal{F}$ wird ein Parameter $\theta = T(\mathcal{F})$ geschätzt, indem für $\mathcal{F}$ das Empirische Maß $\mathcal{F}_n$ als Argument von T eingesetzt wird. Das Funktional T ist die Abbildung eines Maßraumes $\mathbb{W}$ in die reellen Zahlen; $\mathbb{W}$ enthält neben $\mathcal{F}$ zumindest noch alle Wahrscheinlichkeitsmaße mit endlichem Träger. Mit $\mathcal{F}_n^*$ wird die Bootstrap-Version von $\mathcal{F}_n$ bezeichnet. Die Anwendung konventioneller Grenzwertsätze auf eine Größe der Form $\sqrt{n} \cdot [\, T(\mathcal{F}_n^*) - T(\mathcal{F}_n)\,]$ erfordert den Nachweis der Meßbarkeit von T, und das kann in speziellen Fällen relativ aufwendig sein. Im vorliegenden Grenzwertsatz wird auf diese Meßbarkeit verzichtet, und stattdessen eine erweiterte Form der Fréchet-Differenzierbarkeit von T an $\mathcal{F}$ sowie die stochastische Beschränktheit des Empirischen $\mathbb{F}$ - Prozesses vorausgesetzt ($\mathbb{F}$ ist ein geeigneter Funktionenraum). Wegen der nicht vorausgesetzten Meßbarkeit von T ist ein erweitertes Konzept der Verteilungskonvergenz einzuführen. F.Strobl's Konzept baut auf der Hoffmann-Jørgensen-Verteilungskonvergenz auf.
Benötigt wird eine geeignete Pseudo-Metrik in $\mathbb{W}$, bezüglich welcher zugleich die (erweiterte) Fréchet-Differenzierbarkeit von T an $\mathcal{F}$ und die Beschränktheit des Empirischen $\mathbb{F}$ - Prozesses gegeben ist. Eine solche Metrik wird es aber nicht in allen Fällen geben.
Die in diesem Kapitel vorgenommenen Anmerkungen zu F.Strobl's Grenzwertsatz betreffen die vorausgesetzte Darstellbarkeit des Fréchet-Differentials $T'_{\mathcal{F}}(\mathcal{G} - \mathcal{F})$ als Integral einer Funktion $f_{\mathcal{F}}$ bezüglich des Differenz-Maßes $(\mathcal{G} - \mathcal{F}) \in M_r$. Die Menge M_r ist der Definitionsbereich der Fréchet-Ableitung $T'_{\mathcal{F}}$, welcher über den Parameter $r > 0$ variiert werden kann.
Es stellt sich die Frage, für welche Maße $\mathcal{G} \in \mathbb{W}$ die vorausgesetzte Darstellung unter welchen Bedingungen existiert, und welche Form $f_{\mathcal{F}}$ besitzt. R.J.Serfling (1980) hat den Fall $\mathbb{F} = \ell^b$ und $r = +\infty$ betrachtet. Mit ℓ^b wird die Menge aller auf $\mathbb{R}$ definierten, reellwertigen, stetigen und beschränkten Funktionen bezeichnet.
Hier wird der Fall $\mathbb{F} = \ell^b$ bei beliebigem $r > 0$ betrachtet.
Zu diesem Zweck wird der Maßraum $\mathbb{W}$ zerlegt, und dazu eine Menge $A_r(\ell^b)$ eingeführt, über welche $f_{\mathcal{F}}$ bezüglich jedes Maßes aus $\mathbb{W}$ integrierbar ist. Zu jeder Komponente der $\mathbb{W}$- Zerlegung werden (falls möglich) Bedingungen angegeben, unter denen ein entsprechendes $\mathcal{G}$ für die

o.g. Integral-Darstellung existiert.
Die zu beliebigem $r > 0$ angegebene Konstruktion von $f_{\mathcal{F}}$ stimmt im Falle $r = +\infty$ überein mit der von R.J.Serfling (1980) als $T_1(\mathcal{F}, \cdot)$ bezeichneten Funktion.

Der sehr mathematische Charakter der hier behandelten Probleme erzwang oftmals relativ formal wirkende Abhandlungen. Der Autor war bemüht, die Darstellungen durch möglichst viele verbale Erläuterungen anzureichern. Zu diesem Zweck wurden die drei Kapitel mit ausführlichen Einführungen versehen; auf jeweilige Abschlußbemerkungen wurde hingegen verzichtet.

Die Betreuung dieser extern angefertigten Dissertation wurde von Herrn Prof. Dr. Horst Skarabis übernommen. Herrn Prof. Skarabis bin ich vor allem für die Anregung zum Themenkreis „Bootstrapping“, seine anhaltende Geduld und viele wertvolle Hinweise zu großem Dank verpflichtet. Mein Dank gilt auch Herrn Prof. Dr. Johannes Gordesch für die Übernahme des Zweitgutachtens.

Desweiteren bedanke ich mich bei Frau Prof. Dr. Ursula Gather (Univ. Dortmund) für die Einladung zur o.g. Jahrestagung, und bei Herrn Dr. Franz Strobl (Univ. München) für den freundlichen Gedankenaustausch zum Thema „Empirische Prozesse“.

Unerwähnt bleiben soll nicht, dass bei den Literatur-Recherchen eine Diskussion per E-mail mit Prof. Rudolf Beran (Univ. Berkeley, USA) zum Thema „Bootstrapping“ entstand, und sich hieraus manche Erkenntnisse ergaben.

Bei Herrn Manfred Weide bedanke ich mich für ungezählte Gespräche im Zusammenhang mit dem Dissertationsvorhaben.

Doch erst das unbeschränkte Verständnis meiner Lebensgefährtin Martina dafür, lange Zeit nur an zweiter Stelle zu stehen, ermöglichte ein unbelastetes Arbeiten und letztlich die Fertigstellung dieses Manuskriptes.

Inhalt

II. Die Abschätzung der maximalen relativen Häufigkeit einer Häufigkeitsverteilung

1. Einführung

2. Die Darstellung der maximalen relativen Häufigkeit (λ_N / N)

2.1 Die Verteilung $\mathcal{F}$

2.2 Die relativen Häufigkeiten

3. Die Abschätzung von (λ_N / N)

3.1 Schranken der Größe $\sup_{t \in \mathbb{R}} \{ F_N(t+h_N) - F_N(t-h_N) \}$

3.1.1 Die Abschätzung der Kolmogorov / Smirnov - Distance $\mathbb{D}_N$

3.2 Die Schrittweite $2 \cdot h_N$

3.3 Formen der Verteilungsfunktion F und die (λ_N / N) - Abschätzung

3.3.1 Die Verteilungsfunktion F ist stückweise stetig

3.3.2 Die Verteilungsfunktion F ist stetig

3.3.3 Die (λ_N / N) - Abschätzung mithilfe eines weiteren Datensatzes bei beliebigem F

III. Anmerkungen zu einem Bootstrap-Grenzwertsatz

1. Einführung

2. Bootstrap-Variablen und -Verteilung

3. Vorbereitungen für den Grenzwertsatz

3.1 Eine verallgemeinerte Verteilungskonvergenz

3.2 Die erweiterte Fréchet-Ableitung eines Funktionals

Literaturverzeichnis