I. Zur Abhängigkeit der Stichproben-Variablen beim Ziehen ohne Zurücklegen aus einer Grundgesamtheit mit endlich vielen Objekten gleicher Auswahlchance

1. Einführung

Betrachtet wird eine Grundgesamtheit Ω_N mit N Objekten gleicher Auswahlchance.
Sei X ein reellwertiges Merkmal. Eine Stichproben-Variable ordnet einem zufällig aus Ω_N gezogenen Objekt seinen X-Meßwert zu. Stichproben-Variablen besitzen stets die selbe Wahrscheinlichkeitsverteilung wie das zu messende Merkmal.
Unter einer Stichproben-Funktion wird eine Abbildung von n Stichproben-Variablen in die reellen Zahlen verstanden.
Über die Verteilung komplexerer Stichproben-Funktionen ist relativ wenig bekannt, wenn die zugrundeliegenden Stichproben-Variablen stochastisch abhängig sind. Der weitaus größere Teil an Theoremen, Sätzen, Methoden aus Wahrscheinlichkeitstheorie und Mathematischer Statistik wurde für den sogn. „iid-case" geschaffen, deren Anwendung auf Stichproben-Funktionen also die stochastische Unabhängigkeit (kurz: Unabhängigkeit) der Stichproben-Variablen voraussetzt.
Für kompliziertere Schätzfunktionen sind aussagekräftige Angaben über Verteilungs-Kenngrößen (zentrale Momente, Quantile etc.), die wichtige Informationen über die Schätz-Qualität liefern, oftmals nur möglich, wenn die Stichproben-Variablen unabhängig sind.
In der Praxis gibt es aber Situationen, in denen eine **reale** Stichprobe vom Umfang n nur ohne Zurücklegen aus Ω_N gezogen werden kann - z.B. wenn das Meßobjekt notwendig während des Meßvorganges zerstört wird. In diesem Fall sind die zugehörigen Stichproben-Variablen also abhängig.
Zunächst ist festzuhalten, dass das Ziehen ohne Zurücklegen aus Ω_N grundsätzlich durch ein naheliegendes simuliertes Ziehen mit Zurücklegen aus Ω_N ersetzt werden kann - zumindest dann, wenn die Beobachtungsreihenfolge nicht von Interesse ist:
Als erstes werden alle N Objekte eineindeutig markiert, so dass jedes anhand seiner Markierung exakt identifizierbar ist. Dann wird jedem Objekt genau eine natürliche Zahl zwischen 1 und N zugeordnet. Per Zufallsgenerator oder Lostrommel wird n-mal durch Ziehen mit Zurücklegen jeweils eine der Zahlen $1, \ldots, N$ gezogen. Man erhält für die s paarweise verschiedenen Zahlen (unter den insgesamt n Stück, $s \leq n$) eine Häufigkeitstabelle. Über die Zuordnungsvorschrift werden die entsprechenden s Objekte in Ω_N aufgesucht, und an jedem dieser Objekte wird ein-

mal das Merkmal X gemessen.
Zusammen mit der o.g. Häufigkeitstabelle erhält man daraus n Daten (ohne Beachtung einer Reihenfolge), welche die Stichproben-Realisation eines simulierten n-maligen Ziehens mit Zurücklegen bilden.
Diese Methode liefert neben der Unabhängigkeit der zustandegekommenen Stichprobendaten und der damit besseren Zugänglichkeit der betrachteten Stichproben-Funktionen zu mathematischen Sätzen ohne Zusatzaufwand ad hoc die paarweise verschiedenen Stichprobendaten. Die zuletzt genannte Eigenschaft dieser Methode kann ebenfalls einen Vorteil bringen (s.u.).
Gegen diese Methode gibt es theoretische und praktische Einwände.
Betrachtet wird der in obigem Beispiel angegebene Fall, dass das gezogene Objekt notwendig während der Messung zerstört wird. Dann berücksichtigt eine Schätzfunktion, die allein wegen einer Simulation auf der Unabhängigkeit von n Stichproben-Variablen aufbaut, den realen Charakter der zu schätzenden Größe wohl nur unvollkommen. Wendet man mathematische Sätze an auf Schätzfunktionen, deren fundamentale Eigenschaften irreal sind, sind die resultierenden Aussagen zumindest für kleinen Stichprobenumfang vorsichtig zu beurteilen. In asymptotischem Sinne mag die Schätzfunktion durchaus positive Eigenschaften besitzen.
In der Praxis wird die beschriebene Stichproben-Simulation aus Zeit- und Kostengründen in folgenden Situationen eher nicht angewendet:
(i) Es steht eine Schätzmethode zur Verfügung, die für abhängige Stichproben-Variablen entwickelt wurde (Versionen des Bootstrap-Verfahrens, Delta-Methode s.u.).
(ii) Die Objekte von Ω_N sind homogen, und eine Kennzeichnung oder das Auffinden der gezogenen Objekte anhand ihrer Kennzeichnung ist organisatorisch schwierig (z.B. gewisse Organismen in Zoo, Labor, Großaquarium; gewisse Früchte einer Ernte; Produkte einer gewissen Serie).

Wegen $N < +\infty$ ist Ziehen ohne Zurücklegen aus Ω_N stets nach endlich vielen Einzelzügen beendet. In der Praxis bedeutet aber schon bei „mittelgroßem“ N die Vollerhebung einen unakzeptablen Aufwand. In diesen Fällen ergibt sich die Notwendigkeit von Schätzungen, und damit sofort die Frage nach der Beurteilung der Qualität einer Schätzfunktion T_n bei $n < N$.
Generell konzentrieren sich die verfügbaren Untersuchungen (einschließlich Anwendung von Bootstrap-Verfahren und Delta-Methode) über die Verteilung von Stichprobenfunktionen bzw. deren Verteilungs-Kennziffern für den Fall des Ziehens ohne Zurücklegen aus Ω_N auf solche Fälle, in denen die Stichprobenfunktionen hinreichend glatte Funktionen oder Funktionale von Arithmetischen Mitteln sind (auch die Empirische Verteilungsfunktion $\mathbb{F}_n$ ist ja ein Arithmeti-

sches Mittel).

Den theoretischen Hintergrund für die Delta-Methode liefern der Taylor'sche Satz im $\mathbb{R}^n$ und der Grenzwertsatz von P.Erdös / A.Rényi (1959) über die Verteilung einer Summe von n Stichproben-variablen beim Ziehen ohne Zurücklegen aus Ω_N .

Für die Anwendung des Bootstrap-Verfahrens auf den Fall des Ziehens ohne Zurücklegen aus Ω_N lieferten vor allem P.J.Bickel / D.A.Freedman (1981), G.J.Babu / K.Singh (1985) und M.Chao / S.Lo (1985) grundlegende Arbeiten. J.Arrenberg (1998) hat sich in diesem Zusammenhang mit Bootstrap-Methoden der Varianz-Schätzung für verschiedene Mittelwert-Schätzfunktionen befaßt.

Setzt man in der statistischen Praxis (aus Vereinfachungsgründen oder erzwungenermaßen) abhängige Stichproben-Variablen bei einer bestimmten Methode ein, obwohl diese Methode unabhängige Stichproben-Variablen voraussetzt, dann wird man in der Regel keine vertrauenswürdigen Resultate erhalten.

In manchen Anwendungen werden sich aber noch akzeptable Ergebnisse erzielen lassen, wenn der Grad der stochastischen Abhängigkeit wenigstens klein ist.

Es wird eine Ungleichung vorgestellt, die den Zusammenhang zwischen der Größe von Stichprobe und Grundgesamtheit einerseits und der stochastischen Abhängigkeit der Stichproben-Variablen beim Ziehen ohne Zurücklegen andererseits behandelt, und für eine Grundgesamtheit Ω_N mit N Objekten gleicher Auswahlchance gültig ist.

Mit der Problematik der Abhängigkeit der Einzelzüge beim n-maligen Ziehen ohne Zurücklegen aus Ω_N haben sich vor allem M.H.Hansen, W.N.Hurwitz und W.G.Madow (1953), P.Sukhatme (1956), D.Basu (1958) und W.Feller (ab 1950) eingehend beschäftigt.

W.Feller untersuchte u.a. den Zusammenhang zwischen der Hypergeometrischen Verteilung und der Binomialverteilung sowie die Auswahl-Wahrscheinlichkeiten. Seine diesbezüglichen Resultate werden im Abschnitt 2.6 vorgestellt, verbessert und erweitert. Als Folgerung aus der o.g. Ungleichung werden zwei Grenzwertsätze formuliert. Zum einen wird der Abstand der gemeinsamen Wahrscheinlichkeitsfunktionen der Stichproben-Variablen beim Ziehen mit bzw. ohne Zurücklegen aus Ω_N betrachtet, zum anderen der Abstand zwischen den Wahrscheinlichkeitsfunktionen der m-dimensionalen Hypergeometrischen Verteilung und der m-fachen Multinomialverteilung für beliebiges $m \leq N$. Es werden Bedingungen angegeben, unter denen beide Abstände mit $N \to +\infty$ gleichmäßig gegen Null konvergieren, und zwar jeweils zur Ordnung $O(N^{-r})$ mit einem $r \geq 1$.

2. Entwicklung einer Ungleichung

Gegeben ist eine Gesamtheit Ω_N aus N Objekten gleicher Auswahlchance ($N \in \mathbb{N}$).
Die hier vorgestellte Ungleichung hat das Ziel, eine mathematische Beziehung zu formulieren für den Zusammenhang zwischen der stochastischen Abhängigkeit von n Stichproben-Variablen beim Ziehen ohne Zurücklegen aus Ω_N und den Größen n und N ($n \leq N$).

DEF „Stichproben-Variablen" (kurz: SP-Variablen)
Sei $\mathbb{E}$ eine Grundgesamtheit (auch „Population" genannt), ($\mathbb{E}$, E , $\mathcal{P}$) ein Wahrscheinlichkeitsraum und ($\mathbb{R}$, B) ein Meßraum.
Mit Z wird eine Zufallsgröße bezeichnet, die jedem Element von $\mathbb{E}$ den Beobachtungs- oder Meßwert eines bestimmten reellwertigen Merkmals zuordnet, $Z : \mathbb{E} \to \mathbb{R}$ mit der Verteilung $\mathcal{P}Z^{-1} =: \mathcal{G}$. Betrachtet wird folgendes Zufallsexperiment:
Der Grundgesamtheit $\mathbb{E}$ wird n-mal zufällig ein Objekt entnommen ($n \in \mathbb{N}$), bei jedem Objekt wird der zugehörige Z-Wert registriert. Das Zufallsexperiment besteht aus n sogn. „Einzelzügen". Die mit Z_k bezeichnete Zufallsgröße repräsentiert den k-ten Einzelzug ($k \in \{ 1 , \ldots , n \}$) , d.h. Z_k ordnet dem an k-ter Stelle gezogenen Objekt aus $\mathbb{E}$ seinen Z-Wert zu.
Z_k heißt „Stichproben-Variable Nr. k " .
Jede der Stichproben-Variablen $Z_1 , \ldots , Z_n$ ist wie Z gemäß $\mathcal{G}$ verteilt.

Anmerkungen
(a) In der Definition wird nur die Zufälligkeit jeder Entnahme aus $\mathbb{E}$ gefordert, nicht aber die „globale" Unabhängigkeit aller Entnahmen. Erfolgt das Ziehen ohne Zurücklegen aus einer endlichen Grundgesamtheit, dann sind die SP-Variablen abhängig.
(b) Das betrachtete Merkmal trägt das selbe Symbol wie die Zufallsgröße, die den Elementen von $\mathbb{E}$ die Merkmalswerte zuordnet.

Betrachtet wird ein reellwertiges Merkmal X , dessen Verteilung in Ω_N mit $\mathcal{F}(N)$ bezeichnet wird. Zu jedem Objekt von Ω_N gehört genau ein X-Wert.
Mit $z_1 , \ldots , z_N$ werden die X-Werte symbolisiert, von denen $a_1 , \ldots , a_m$ paarweise verschieden sind ($1 \leq m \leq N$). Aus Ω_N werden n Objekte ohne Zurücklegen gezogen ($1 \leq n \leq N$). Jede der Stichproben-Variablen $X_1 , \ldots , X_n$ ist wie das Merkmal X gemäß $\mathcal{F}(N)$ verteilt .

Mit $x_1, \ldots, x_n$ werden die Realisationen der SP-Variablen bezeichnet, der Vektor $(x_1, \ldots, x_n)$ heißt Realisation der Stichprobe. Als nächstes werden die Häufigkeits-Variablen $Y_1, \ldots, Y_m$ definiert: Y_k gibt die Anzahl der Stichprobendaten an, die die mit a_k übereinstimmen ($k \in \{1, \ldots, m\}$); mit $n_1, \ldots, n_m$ werden die Realisationen der $Y_1, \ldots, Y_m$ symbolisiert ($1 \leq m \leq N$).

Man hat folgende Tabelle mit der Verteilung $\mathcal{F}(N)$ und einer beliebig entstandenen Stichproben-Realisation:

Paarweise verschiedene X-Werte in Ω_N	a_1	...	a_m
Absolute Häufigkeiten in Ω_N	N_1	...	N_m
Absolute Häufigkeiten in der SP-Realisation (Realisationen der $Y_1, \ldots, Y_m$)	n_1	...	n_m

(1)

Es gelten die nachfolgenden Beziehungen:

$$\left.\begin{aligned} & 1 \leq n \leq N < +\infty \;; \\ & 1 \leq N_i \leq N \;(i = 1 \ldots m), \quad N_1 + \ldots + N_m = N \;; \\ & \vec{N} := (N_1, \ldots, N_m). \\ & 0 \leq n_i \leq \min\{N_i, n\} \;(i = 1 \ldots m), \quad n_1 + \ldots + n_m = n \;; \\ & \vec{n} := (n_1, \ldots, n_m). \end{aligned}\right\} \quad (2)$$

Mit λ_N wird die absolute Häufigkeit des Modalwertes der X-Verteilung $\mathcal{F}(N)$ bezeichnet.

2.1 Die gemeinsame Verteilung der Stichproben-Variablen

Ziehen ohne Zurücklegen aus Ω_N

Die SP-Variablen $X_1, \ldots, X_n$ besitzen mit einer beliebig gewählten Stichproben-Realisation gemäß Häufigkeitsverteilung (1) die folgende gemeinsame Wahrscheinlichkeitsfunktion:

$$P_{ZoZ}\left(X_1 = x_1, \ldots, X_n = x_n\right) = \frac{\prod_{k=1}^{n} P\left(X_1 = x_1, \ldots, X_k = x_k\right)}{\prod_{j=1}^{n-1} P\left(X_1 = x_1, \ldots, X_j = x_j\right)}$$

$$= \quad P\left(X_1 = x_1\right) \cdot \prod_{k=2}^{n} P\left(X_k = x_k \;\middle|\; X_1 = x_1 \;, \ldots, \; X_{k-1} = x_{k-1}\right)$$

Mit den Größen der Häufigkeitstabelle (1) und der Konvention $\prod_{k=0}^{-1} a_k := 1$ gilt daher:

$$P_{ZoZ}(X_1 = x_1 \;, \ldots, \; X_n = x_n) \quad = \quad \frac{\prod_{j=1}^{m} \prod_{k=0}^{n_j - 1} (N_j - k)}{\prod_{i=0}^{n-1} (N - i)}$$

$$= \frac{\prod_{j=1}^{m} N_j^{\,n_j} \prod_{i=0}^{n_j-1} \left(1 - \frac{i}{N_j}\right)}{\prod_{i=0}^{n-1} (N - i)} \quad = \quad \frac{\left(\prod_{k=1}^{m} N_k^{\,n_k}\right) \left(\prod_{j=1}^{m} \prod_{i=0}^{n_j-1} \left(1 - \frac{i}{N_j}\right)\right)}{\prod_{i=0}^{n-1} (N - i)} \qquad (3)$$

Ziehen mit Zurücklegen aus Ω_N

In diesem Fall gilt für die gemeinsame Wahrscheinlichkeitsfunktion der SP-Variablen $X_1, \ldots, X_n$ mit den Größen der Häufigkeits-Tabelle (1) :

$$P_{ZmZ}(X_1 = x_1 \;, \ldots, \; X_n = x_n) \quad = \quad \prod_{k=1}^{n} P(X_k = x_k)$$

$$= \quad \prod_{k=1}^{m} \left(\frac{N_k}{N}\right)^{n_k} \quad = \quad \frac{\prod_{k=1}^{m} N_k^{\,n_k}}{N^n} \qquad (4)$$

2.2 Die Differenz D (n, N ; $\vec{N}$, $\vec{n}$)

Mit den Größen aus (1) wird definiert:

$$D(n, N; \vec{N}, \vec{n}) \quad := \quad \frac{\left(\prod_{k=1}^{m} N_k^{n_k}\right) \cdot \left(\prod_{j=1}^{m} \prod_{i=1}^{n_j-1} \left(1 - \frac{i}{N_j}\right)\right)}{\prod_{i=0}^{n-1} (N - i)} \quad - \quad \frac{\prod_{k=1}^{m} N_k^{n_k}}{N^n} \qquad (5)$$

Verglichen wird die gemeinsame Wahrscheinlichkeitsfunktion der SP-Variablen beim Ziehen ohne Zurücklegen mit jener beim Ziehen mit Zurücklegen.

Wegen (3) , (4) und (5) gilt:

Notwendig und hinreichend für eine geringe stochastische Abhängigkeit der SP-Variablen $X_1 , \ldots , X_n$ beim Ziehen ohne Zurücklegen aus Ω_N ist ein kleiner Wert des Abstandes $| D (n , N ; \vec{N} , \vec{n}) |$, der für alle gemäß (1) und (2) zulässigen Datensätze $\vec{N}$ und $\vec{n}$ gültig ist.

Im folgenden wird eine obere Schranke für $| D (n , N ; \vec{N} , \vec{n}) |$ ntwickelt, die für alle gemäß (1) und (2) zulässigen Datensätze $\vec{N}$ und $\vec{n}$ gültig ist, und nur noch von n , N und λ_N abhängt.

Ein Sonderfall: (6)

Es ist $X \equiv \text{constant}$ in Ω_N genau dann, wenn $\lambda_N = N$ ist.

Wegen (2) gilt: $\lambda_N = N \;\Rightarrow\; m = 1 ,\ n_1 = n ,\ N_1 = \lambda_N = N \;\Rightarrow D (n , N ; \vec{N} , \vec{n}) \equiv 1 - 1 = 0.$

Daher ist der Fall $\lambda_N = N$ in jede obere Abschätzung von $| D (n, N ; \vec{N} , \vec{n}) |$ mit eingeschlossen. Im folgenden wird von $1 < m$ bzw. $\lambda_N < N$ ausgegangen.

2.3 Die Abschätzung von $D (n, N ; \vec{N} , \vec{n})$

Grundlegend für die Abschätzung von $D (n, N ; \vec{N} , \vec{n})$ ist das folgende Lemma.

Lemma I.1

Es wird definiert für $x, y \in \mathbb{R}$ und $n, N \in \mathbb{N}$ mit $x \le y+1$, $1 \le y < N$, $n \le N$:

$$T(n,N,x,y) := (n-2) \cdot \min\{n-1 \,;\, y-1\} - x \cdot \left[1 - \prod_{i=0}^{n-1} \left(1 - \frac{i}{N} \right) \right] \qquad (7)$$

Sei zusätzlich eine der nachfolgenden Voraussetzungen erfüllt:

(A) $2 = y < n$, $8 < N$

(B) $3 = y < n$, $5 < N$

(C) $3 \le y < n$

(D) $3 < n \le y$

Dann gilt: $T (n , N , x , y) > 0$.

Beweis:

(A) Fallunterscheidung:

(i) $2 = y \ , \ 3 = n :$

$$T(n,N,x,y) \;=\; T(3,N,x,2) \;=\; 1\cdot 1 \;-\; x \cdot \left[1 \;-\; \prod_{i=0}^{2}\left(1 \;-\; \frac{i}{N}\right)\right]$$

$$\geq\; 1 \;-\; (y+1)\cdot\left[1 \;-\; \prod_{i=0}^{2}\left(1 \;-\; \frac{i}{N}\right)\right] \;=\; 1 \;-\; 3\cdot\left[1 \;-\; \prod_{i=0}^{2}\left(1 \;-\; \frac{i}{N}\right)\right]$$

$$=\; 1 \;-\; 3\cdot\left[1 \;-\; \left(1-\frac{1}{N}\right)\cdot\left(1-\frac{2}{N}\right)\right] \;=\; 1 \;-\; 3\cdot\left[1 \;-\; \left(1-\frac{3}{N}+\frac{2}{N^2}\right)\right]$$

$$=\; 1 \;-\; 3\cdot\left[\frac{3}{N} \;-\; \frac{2}{N^2}\right] \;=\; \frac{1}{N^2}\cdot\left[N^2 \;-\; 9\cdot N \;+\; 6\right] \;=:\; \frac{1}{N^2}\cdot q(N)$$

Es gilt: $\quad N > 8 \;\;=>\;\; q(N) > 0 \;\;=>\;\; T(3,N,x,2) > 0 \quad$ für alle $x \leq 3$.

(ii) $2 = y \ , \ 4 = n :$

$$T(n,N,x,y) \;=\; T(4,N,x,2) \;=\; 2\cdot 1 \;-\; x \cdot \left[1 \;-\; \prod_{i=0}^{3}\left(1 \;-\; \frac{i}{N}\right)\right]$$

$$\geq\; 2 \;-\; 3\cdot\left[1 \;-\; \left(1-\frac{1}{N}\right)\cdot\left(1-\frac{2}{N}\right)\cdot\left(1-\frac{3}{N}\right)\right]$$

$$=\; 2 \;-\; 3\cdot\left[1-\left(1-\frac{6}{N}+\frac{11}{N^2}-\frac{6}{N^3}\right)\right] \;=\; 2 \;-\; 3\cdot\left(\frac{6}{N}-\frac{11}{N^2}+\frac{6}{N^3}\right)$$

$$=\; \frac{2}{N^3}\cdot\left(N^3 \;-\; 9\cdot N^2 \;+\; \frac{33}{2}\cdot N \;-\; 9\right) \;=:\; \frac{2}{N^3}\cdot p(N)$$

Es gilt: $\quad N > 6 \;\;=>\;\; p(N) > 0 \;\;=>\;\; T(4,N,x,2) > 0 \quad$ für alle $x \leq 3$.

(iii) $2 = y \ , \ 4 < n \ :$

$$T(n,N,x,y) \;=\; T(n,N,x,2) \;=\; (n-2)\cdot 1 \;-\; x \cdot \left[1 \;-\; \prod_{i=0}^{n-1}\left(1 \;-\; \frac{i}{N}\right)\right]$$

$$>\; (n-2) \;-\; x\cdot 1 \;\geq\; n \;-\; 2 \;-\; 3 \;=\; n \;-\; 5 \;\geq\; 0$$

Damit ist (A) bewiesen.

(B) $3 = y = n$:

$$T(3,N,x,3) = 1\cdot 2 - x \cdot \left[1 - \prod_{i=0}^{2}\left(1 - \frac{i}{N}\right)\right]$$

$$\geq 2 - 4 \cdot \left[1 - \left(1-\frac{1}{N}\right)\cdot\left(1-\frac{2}{N}\right) \right] = 2 - 4 \cdot \left[\frac{3}{N} - \frac{2}{N^2} \right]$$

$$= \frac{2}{N^2} \cdot \left[N^2 - 6\cdot N + 4\right] =: \frac{2}{N^2} \cdot r(N)$$

Es gilt: $N > 5 \quad => \quad r(N) > 0 \quad => \quad T(3,N,x,3) > 0$ für alle $x \leq 4$.

Damit ist (B) bewiesen.

(C) Fallunterscheidung:

(i) $3 = y < n$:

$$T(n,N,x,3) = (n-2)\cdot 2 - x \cdot \left[1 - \prod_{i=0}^{n-1}\left(1 - \frac{i}{N}\right)\right]$$

$$> (n-2)\cdot 2 - x\cdot 1 \geq (n-2)\cdot 2 - 4 = 2\cdot n - 8 \geq 0$$

(ii) $4 \leq y \leq n$:

$$T(n,N,x,4) = (n-2)\cdot 3 - x \cdot \left[1 - \prod_{i=0}^{n-1}\left(1 - \frac{i}{N}\right)\right]$$

$$> (n-2)\cdot 3 - x\cdot 1 \geq (n-2)\cdot 3 - 5 = 3\cdot n - 11 > 0$$

Damit ist (C) bewiesen.

(D) Sei also $3 < n \leq y$.

Es wird definiert:

$$Q_n := (n-2)\cdot(n-1) - x \cdot \left[1 - \prod_{i=0}^{n-1}\left(1 - \frac{i}{N}\right)\right] = T(n,N,x,y)$$

Beh.: $Q_n > 0$ für alle $n \in \{ 4 , \ldots , y \}$.

Bew.: Vollständige Induktion (x , y , N werden als Konstante behandelt

mit $1 \leq y < N$, $x \leq y+1$)

$n=4$ (Induktions-Anfang) :

$$Q_4 = 6 - x \cdot \left[1 - \prod_{i=0}^{3} \left(1 - \frac{i}{N} \right) \right] = 6 - x \cdot \left[1 - \left(1-\frac{1}{N}\right)\cdot\left(1-\frac{2}{N}\right)\cdot\left(1-\frac{3}{N}\right) \right]$$

$$= 6 - x \cdot \left[\frac{6}{N} - \frac{11}{N^2} + \frac{6}{N^3} \right] = 6 - \frac{x}{N} \cdot \left[6 - \frac{11}{N} + \frac{6}{N^2} \right]$$

$$\geq 6 - 1 \cdot \left[6 - \frac{11}{N} + \frac{6}{N^2} \right] = \frac{11}{N^2} \cdot \left[N - \frac{6}{11} \right] > 0$$

$n \rightarrow n+1$ (Induktions-Schritt) :

$$Q_{n+1} = (n-1)\cdot n - x \cdot \left[1 - \prod_{i=0}^{n} \left(1 - \frac{i}{N} \right) \right]$$

$$= n^2 - n - x + x \cdot \left(1-\frac{n}{N}\right) \cdot \prod_{i=0}^{n-1} \left(1 - \frac{i}{N} \right)$$

$$= n^2 - n - x + x \cdot \prod_{i=0}^{n-1} \left(1 - \frac{i}{N} \right) - \frac{n\cdot x}{N} \cdot \prod_{i=0}^{n-1} \left(1 - \frac{i}{N} \right)$$

$$= (n-2)\cdot(n-1) - x \cdot \left[1 - \prod_{i=0}^{n-1} \left(1 - \frac{i}{N} \right) \right] - \frac{n\cdot x}{N} \cdot \prod_{i=0}^{n-1} \left(1 - \frac{i}{N} \right)$$

$$- (n-2)\cdot(n-1) + n^2 - n$$

Gemäß Induktions-Voraussetzung gilt daher:

$$Q_{n+1} = Q_n - \frac{n\cdot x}{N} \cdot \prod_{i=0}^{n-1} \left(1 - \frac{i}{N} \right) + 2\cdot(n-1) > Q_n - n\cdot 1 + 2\cdot(n-1) > Q_n$$

q.e.d.

Damit ist auch (D) bewiesen.

❑

Nun zur Abschätzung von $D(n, N; \vec{N}, \vec{n})$. Es gilt:

$$D(n,N;\vec{N},\vec{n}) = \frac{\left(\prod_{k=1}^{m} N_k^{n_k} \right) \left(\prod_{j=1}^{m} \prod_{i=1}^{n_j-1} \left(1 - \frac{i}{N_j} \right) \right)}{\prod_{i=0}^{n-1} (N - i)} - \frac{\prod_{k=1}^{m} N_k^{n_k}}{N^n}$$

$$= \left(\prod_{k=1}^{m} N_k^{n_k}\right) \cdot \left[\frac{\prod_{j=1}^{m} \prod_{i=1}^{n_j-1} \left(1 - \frac{i}{N_j}\right)}{\prod_{i=0}^{n-1} (N - i)} - \frac{1}{N^n} \right] .$$

Wegen (1) und (2) gilt für $0 \le i \le n_j - 1$, $j \in \{ 1 , . . . , m \}$:

$0 \le i \le n_j - 1 \le N_j - 1 < N_j$ für jedes $j \in \{ 1 , . . . , m \}$.

Daraus folgt: $0 \le \frac{i}{N_j} \le \frac{n_j - 1}{N_j} \le \max_k \{ \frac{n_k - 1}{N_k} \} = \frac{n_t - 1}{N_t}$ für ein $t \in \{ 1 , . . . , m \}$.

Setze $\frac{n_t - 1}{N_t} =: \gamma$, also ist $\gamma \in [0 ; 1)$ wegen (2a) .

Die Differenz $D (n , N; \vec{N}, \vec{n})$ wird zunächst nach unten, und dann nach oben abgeschätzt.

Wegen $\gamma \in [0 ; 1)$ ist für beliebige $n , m \in \mathbb{N}$ mit $2 \le m$: $(1-\gamma)^{n-m} \ge (1-\gamma)^{n-2}$.

Es gilt:

$$D(n,N;\vec{N},\vec{n}) = \left(\prod_{k=1}^{m} N_k^{n_k}\right) \cdot \left[\frac{\prod_{j=1}^{m} \prod_{i=1}^{n_j-1} \left(1 - \frac{i}{N_j}\right)}{\prod_{i=0}^{n-1} (N - i)} - \frac{1}{N^n} \right]$$

$$\ge \left(\prod_{k=1}^{m} N_k^{n_k}\right) \cdot \left[\frac{\prod_{j=1}^{m} \prod_{i=1}^{n_j-1} (1 - \gamma)}{\prod_{i=0}^{n-1} (N - i)} - \frac{1}{N^n} \right]$$

$$= \left(\prod_{k=1}^{m} N_k^{n_k}\right) \cdot \left[\frac{(1-\gamma)^{n-m}}{\prod_{i=0}^{n-1} (N - i)} - \frac{1}{N^n} \right]$$

$$\ge \left(\prod_{k=1}^{m} N_k^{n_k}\right) \cdot \left[\frac{(1-\gamma)^{n-2}}{\prod_{i=0}^{n-1} (N - i)} - \frac{1}{N^n} \right]$$

Aus (1) und (2) folgt: $n_t - 1 \le \min \{ n-1 ; \lambda_N - 1 \}$.

Aus der Bernoulli-Ungleichung ergibt sich daher für alle $n \ge 2$ die (auch für $n = 1$ noch gültige)

Beziehung:

$$(1-\gamma)^{n-2} \;\geq\; 1 - (n-2)\cdot\gamma \;=\; 1-(n-2)\cdot\frac{(n_t-1)}{N_t}$$

$$\geq\; 1 - \frac{(n-2)}{N_t}\cdot\min\{\,n-1\;;\;\lambda_N-1\,\}\,.$$

Daraus folgt:

$$D(n,N;\vec{N},\vec{n}) \;\geq\; \left(\prod_{k=1}^{m} N_k^{n_k}\right)\cdot\left[\;\frac{1-\dfrac{(n-2)}{N_t}\cdot\min\{n-1\;;\;\lambda_N-1\}}{\prod\limits_{i=0}^{n-1}(N-i)} \;-\; \frac{1}{N^n}\;\right]$$

$$= \frac{\prod\limits_{k=1}^{m} N_k^{n_k}}{\prod\limits_{i=0}^{n-1}(N-i)}\cdot\left[\,-\frac{(n-2)}{N_t}\cdot\min\{n-1\;;\;\lambda_N-1\} + 1 - \frac{\prod\limits_{i=0}^{n-1}(N-i)}{N^n}\,\right]$$

$$= \frac{N_t^{(n_t-1)}\cdot\prod\limits_{k=1,k\neq t}^{m} N_k^{n_k}}{\prod\limits_{i=0}^{n-1}(N-i)}\cdot\left[\,-(n-2)\cdot\min\{n-1\;;\;\lambda_N-1\} + N_t\cdot\left(1-\prod_{i=0}^{n-1}\left(1-\frac{i}{N}\right)\right)\right]$$

$$=:\; G_{unten}\,. \tag{8}$$

Als nächstes wird $D(n, N; \vec{N}, \vec{n})$ nach oben abgeschätzt. Wegen (2) gilt:

$$D(n,N;\vec{N},\vec{n}) \;=\; \frac{\left(\prod\limits_{k=1}^{m} N_k^{n_k}\right)\cdot\left(\prod\limits_{j=1}^{m}\prod\limits_{i=1}^{n_j-1}\left(1-\dfrac{i}{N_j}\right)\right)}{\prod\limits_{i=0}^{n-1}(N-i)} \;-\; \frac{\prod\limits_{k=1}^{m} N_k^{n_k}}{N^n}$$

$$\leqslant\; \frac{\prod\limits_{k=1}^{m} N_k^{n_k}}{\prod\limits_{i=0}^{n-1}(N-i)} - \frac{\prod\limits_{k=1}^{m} N_k^{n_k}}{N^n} \;=\; \frac{\prod\limits_{k=1}^{m} N_k^{n_k}}{\prod\limits_{i=0}^{n-1}(N-i)}\cdot\left(1-\frac{\prod\limits_{j=0}^{n-1}(N-j)}{N^n}\right)$$

$$= \frac{\prod\limits_{k=1}^{m} N_k^{n_k}}{\prod\limits_{i=0}^{n-1}(N-i)}\cdot\left(1-\prod_{j=0}^{n-1}\left(1-\frac{j}{N}\right)\right) \;=:\; G_{oben} \tag{9}$$

2.4 Die Ungleichung

Mit (7) läßt sich G_{unten} für $x = N_t$ und $y = \lambda_N$ so darstellen:

$$G_{unten} \quad = \quad (-1) \cdot \frac{N_t^{(n_t-1)} \cdot \prod_{k=1,k\neq t}^{m} N_k^{n_k}}{\prod_{i=0}^{n-1} (N - i)} \cdot T(n, N, N_t, \lambda_N)$$

Sei eine der folgenden vier Voraussetzungen erfüllt:

$2 = \lambda_N < n \leq N$, $8 < N$;

$3 = \lambda_N = n \leq N$, $5 < N$;

$3 \leq \lambda_N < n \leq N$;

$3 < n \leq \lambda_N < N$.

Wegen Lemma 1 gilt dann:

$$G_{unten} < 0 . \tag{10}$$

Für $1 \leq n \leq N$ und $1 \leq \lambda_N < N$ erhält man aus (2) zusammen mit (9) :

$$\begin{aligned} G_{oben} \quad &= \quad \frac{\prod_{k=1}^{m} N_k^{n_k}}{\prod_{i=0}^{n-1} (N - i)} \cdot \left(1 - \prod_{j=0}^{n-1} \left(1 - \frac{j}{N} \right) \right) \\ &\leq \quad \frac{(\lambda_N)^n}{\prod_{i=0}^{n-1} (N - i)} \cdot \left(1 - \prod_{j=0}^{n-1} \left(1 - \frac{j}{N} \right) \right) . \end{aligned} \tag{11}$$

Für alle gemäß (1) und (2) zulässigen Datensätze $\vec{N}$, $\vec{n}$ erhält man aus (8) , (10) und (11) unter den Voraussetzungen von Lemma I.1 (A) , (B) , (C) oder (D) die folgende Ungleichungskette, die auch noch für $1 = \lambda_N \leq n \leq N$ gültig ist:

$$\left. \begin{aligned} &\frac{(\lambda_N)^{n-1}}{\prod_{i=0}^{n-1} (N - i)} \cdot \left[-(n-2) \cdot \min\{n-1 \; ; \; \lambda_N - 1\} + 1 - \prod_{j=0}^{n-1} \left(1 - \frac{j}{N} \right) \right] \\ &\leq D(n, N; \vec{N}, \vec{n}) \leq \frac{(\lambda_N)^n}{\prod_{i=0}^{n-1} (N - i)} \cdot \left(1 - \prod_{j=0}^{n-1} \left(1 - \frac{j}{N} \right) \right) \end{aligned} \right\} \tag{12}$$

Mit dem folgenden Lemma werden die Schranken von $D(n, N; \vec{N}, \vec{n})$ verglichen.

Lemma I.2

Unter den Voraussetzungen von Lemma I.1 (A) , (B) , (C) oder (D) gilt:

$$\lambda_N \cdot \left[1 - \prod_{i=0}^{n-1} \left(1 - \frac{i}{N}\right) \right] < (n-2) \cdot \min\{n-1 \,;\, \lambda_N - 1\} - \left[1 - \prod_{i=0}^{n-1} \left(1 - \frac{i}{N}\right) \right] \tag{13}$$

Beweis:

Gemäß (7) gilt mit $y = \lambda_N < N$ und $x = \lambda_N + 1 \le N$:

$$T(n, N, \lambda_N + 1, \lambda_N) = (n-2) \cdot \min\{n-1 \,;\, \lambda_N - 1\} - (\lambda_N + 1) \cdot \left[1 - \prod_{i=0}^{n-1} \left(1 - \frac{i}{N}\right) \right]$$

Die Aussage des Lemmas ist erfüllt genau dann, wenn $T(n, N, \lambda_N + 1, \lambda_N) > 0$ ist.
Dies aber folgt unter den angegebenen Voraussetzungen aus Lemma I.1 (A) , (B) , (C) und (D) .

❑

Für Abschätzung des Betrages $| D(n, N; \vec{N}, \vec{n}) |$ werden abschließend noch einige Beziehungen benötigt.
Es gilt:

$$\min\{n-1 \,;\, \lambda_N - 1\} = \min\{n \,;\, \lambda_N\} - 1 = N \cdot \min\{ \frac{n}{N} \,;\, \frac{\lambda_N}{N} \} - 1 \;; \tag{14}$$

$$\text{aus } \frac{\lambda_N}{N} \le R \quad \text{folgt} \quad N \cdot \min\{ \frac{n}{N} \,;\, \frac{\lambda_N}{N} \} \le N \cdot \min\{ \frac{n}{N} \,;\, R \} \;. \tag{15}$$

Mitl $\delta_{k,j}$ wird das Kronecker-Symbol bezeichnet: $\delta_{k,j} := \begin{cases} 1 & \text{falls } k = j \\ 0 & \text{sonst} \end{cases}$

Aus (14) erhält man die Identität:

$$\frac{(\lambda_N)^{n-1}}{\prod_{i=0}^{n-1} (N-i)} \cdot \left[(n-2) \cdot \min\{ n-1 \,;\, \lambda_N - 1 \} - \left(1 - \prod_{j=0}^{n-1} \left(1 - \frac{j}{N}\right) \right) \cdot (-1)^{\delta_{1,\lambda_N}} \right]$$

$$= \left(\frac{\lambda_N}{N}\right)^{n-1} \cdot \prod_{i=0}^{n-1}\left(\frac{N}{N-i}\right)\cdot\frac{1}{N}\cdot\left[(n-2)\cdot\left(N\cdot\min\{\frac{n}{N}\,;\,\frac{\lambda_N}{N}\}\;-1\right) - \left(1-\prod_{j=0}^{n-1}\left(1-\frac{j}{N}\right)\right)\cdot(-1)^{\delta_{1,\lambda_N}}\right]$$

$$=: \quad B(n\,,\,N\,,\frac{\lambda_N}{N}) \quad . \tag{16}$$

Damit läßt sich nun der zentrale Satz dieses Kapitels formulieren.

Satz I.1

Eine der nachfolgenden Voraussetzungen sei erfüllt:

(i) $1 = \lambda_N \le n \le N$

(ii) $2 = \lambda_N < n \le N \;,\; 8 < N$

(iii) $3 = \lambda_N = n \le N \;,\; 5 < N$

(iv) $3 \le \lambda_N < n \le N$

(v) $3 < n \le \lambda_N \le N$

Dann gilt mit der Definition (16) für alle zugehörigen, gemäß (1) und (2) zulässigen Datensätze $\vec{N}$, $\vec{n}$ die Ungleichung:

$$|\, D(n, N;\, \vec{N}\,,\, \vec{n}\,)\,| \quad \le \quad \min\left\{\; B(n\,,\,N\,,\frac{\lambda_N}{N}) \;\;;\;\; 1 \;\right\} \quad . \tag{17}$$

Unter der Voraussetzung (i) gilt die Identität:

$$|\, D(n, N;\, \vec{N}\,,\, \vec{n}\,)\,| \;=\; B(n\,,\,N\,,\frac{1}{N}) \;=\; \frac{1}{\prod_{i=0}^{n-1}(N\,-\,i)} \;-\; \frac{1}{N^n} \quad . \tag{18}$$

Mit $2 < n \le N$ gilt:

$$\frac{\lambda_N}{N} \le R \quad \Rightarrow \quad B(n\,,\,N\,,\frac{\lambda_N}{N}) \;\le\; B(n\,,\,N\,,\,R\,) \quad .$$

Beweis:

Wegen (6) ist für $\lambda_N = N$ nichts zu zeigen - in diesem Fall ist $|\, D(n, N;\, \vec{N}\,,\, \vec{n}\,)\,| \equiv 0$.
Die Gültigkeit von (18) unter der Voraussetzung (i) folgt unmittelbar aus der Ungleichungskette (12) sowie aus der Tatsache, dass die rechte Seite von (18) nicht größer als Eins ist.
Unter den Voraussetzungen (ii) , (iii) , (iv) oder (v) gilt (13) wegen Lemma I.2 .
Zusammen mit den Umformumgen (14) und (16) folgt daher aus der Ungleichungskette (12)

unter jeder dieser Voraussetzungen die Gültigkeit von $| D (n , N ; \vec{N}, \vec{n}) | \leq B(n, N, \frac{\lambda_N}{N})$.

Nach Definition ist $| D (n , N ; \vec{N} , \vec{n} |$ der Abstand zweier Wahrscheinlichkeiten, und kann daher nicht größer als 1 sein. Daraus ergibt sich (17) .

Die wachsende Monotonie von $B(n , N , R)$ in R mit $2 < n \leq N$ folgt aus (15) .

Damit ist alles gezeigt.

❑

2.4.1 Die Beurteilung der Ungleichung

Zunächst werden einige Bemerkungen zu Satz I.1 gemacht. Danach wird aus der Ungleichung das asymptotische Verhalten von $| D (n , N ; \vec{N} , \vec{n}) |$ für große N hergeleitet sowie eine Schranke für $| D (n , N ; \vec{N} , \vec{n} |$, die nur noch von n und N abhängt.

Bemerkungen zu Satz I.1

(a) Für alle die Fälle, in denen die X-Werte $z_1 , . . . , z_N$ paarweise verschieden auftreten (d.h. $\lambda_N = 1$), gilt in der Ungleichung (17) für alle $n \in \{1, . . . , N \}$ wegen (18) das Gleichheitszeichen.

In diesem Sinne ist die mit der Ungleichung zum Ausdruck kommende Abschätzung scharf.

(b) Weil hier der Abstand von Wahrscheinlichkeitsfunktionen betrachtet wurde (anstelle von Verteilungsfunktionen), ist die Ungleichung (17) auch für solche reellwertigen Merkmale gültig, die nur nominal meßbar sind.

(c) Allgemein gilt die Beziehung $\lambda_N \leq N - m + 1$; für nicht-constante Merkmale ist $2 \leq m$ und damit $\lambda_N \leq N - 1$. Im folgenden wird $\frac{N-1}{N}$ als „triviale“ Obergrenze für $\frac{\lambda_N}{N}$ bezeichnet.

(d) Mit den Definitionen (5) und (16) ergibt sich aus (17) die nachfolgende, ausführlich formulierte Ungleichung.

Unter jeder der Voraussetzungen (i) bis (v) von Satz I.1 gilt für jede gemäß (1) und (2) zulässige Stichproben-Realisation $(x_1, \ldots, x_n)$ beim n-maligen Ziehen ohne Zurücklegen aus Ω_N :

$$\left| \; P_{ZoZ}\left(X_1 = x_1 \;, \ldots , \; X_n = x_n \right) - \prod_{k=1}^{n} P\left(X_k = x_k \right) \right| \le$$

$$\left(\frac{\lambda_N}{N}\right)^{n-1} \cdot \prod_{i=0}^{n-1} \left(\frac{N}{N-i}\right) \cdot \frac{1}{N} \cdot \left[(n-2)\cdot\left(N\cdot \min\left\{ \frac{n}{N} \; ; \; \frac{\lambda_N}{N} \right\} - 1\right) - \left(1 - \prod_{j=0}^{n-1}\left(1-\frac{j}{N}\right)\right)\cdot(-1)^{\delta_{1,\lambda_N}} \right] \tag{19}$$

Im folgenden Satz wird das asymptotische Verhalten des Abstandes $| D (n , N ; \vec{N} , \vec{n}) |$ für große N betrachtet.

Satz I.2

Sei $\vec{x} = (x_1, \ldots, x_n)$ eine Stichproben-Realisation, die sich aus Messungen des Merkmals X an n Objekten ergibt, welche ohne Zurücklegen aus Ω_N gezogen wurden. Mit $\mathbb{W}_n \subset \mathbb{R}^n$ wird die Menge aller möglichen derartigen Stichproben-Realisationen bezeichnet.
Unter jeder der Voraussetzungen von Satz I.1 gilt mit allen gemäß (1) und (2) zulässigen Datensätzen $\vec{N}$, $\vec{n}$ zu festem $n \in \mathbb{N}$ mit einem $r \ge 1$:

Der Abstand $| D (n , N ; \vec{N} , \vec{n} |$ konvergiert mit N -> +oo gleichmäßig in $\mathbb{W}_n$ gegen Null zur Ordnung $O (N^{-r})$, d.h.

$$\sup_{\vec{x}\in \mathbb{W}_n} \left| P_{ZoZ}\left(X_1 = x_1 \;, \ldots , \; X_n = x_n \right) - \prod_{k=1}^{n} P\left(X_k = x_k \right) \right| = O(N^{-r}) \quad , \qquad N \to +\infty \tag{20}$$

Beweis:
Mit $\lambda_N = 1$ folgt die Behauptung direkt aus (18) , denn die rechte Seite von (18) ist von der Ordnung $o (N^{-n})$, N -> + oo .
Sei jetzt $\lambda_N > 1$, dann folgt mit $2 < n \le N$ aus (16) wegen der fallenden Monotonie von $B (n , N , R)$ in R (siehe Satz I.1) :

$$0 \le B\left(n \; , \; N \; , \frac{\lambda_N}{N}\right) \le B(n \; , \; N \; , \; 1)$$

$$\leq \; 1 \cdot \prod_{i=0}^{n-1} \left(\frac{N}{N-i} \right) \cdot \frac{1}{N} \cdot \left[(n-2)\cdot(n-1) \; + \; 1 - \prod_{j=0}^{n-1} \left(1 - \frac{j}{N} \right) \right] \qquad (21)$$

Der letzte Term ist Daten-unabhängig und strebt zu festem $n \in \mathbb{N}$ mit $N \to +\infty$ zur Ordnung $O(N^{-1})$ gegen Null.
Zusammen mit Satz I.1 folgt daher die Behauptung.

❑

Corollar zu Satz I.2

Wegen Satz I.2 gilt zu jedem festen $n \in \mathbb{N}$ für hinreichend große $N \in \mathbb{N}$:

$$\min \left\{ \; B\left(n \, , \, N \, , \frac{\lambda_N}{N} \right) \; ; \; 1 \; \right\} \; = \; B\left(n \, , \, N \, , \frac{\lambda_N}{N} \right) \; .$$

Mit der unter bestimmten Annahmen verbesserten Abschätzung von $\frac{\lambda_N}{N}$ befaßt sich das II.Kapitel. Die „triviale“ $\frac{\lambda_N}{N}$ - Obergrenze liefert zwar eine i.a. verschlechterte, aber nicht mehr von λ_N abhängige Abschätzung (die wegen (6) auch für constante Merkmale gültig ist):

$$\begin{aligned} | D (n , N ; \vec{N} \, , \vec{n} \,) | \; &\leq \; B\left(n \, , \, N \, , \frac{\lambda_N}{N} \right) \; \leq \; B\left(n \, , \, N \, , \frac{N-m+1}{N} \right) \\ &\leq \; B\left(n \, , \, N \, , \frac{N-1}{N} \right) \; . \qquad (22) \end{aligned}$$

2.5 Der Identitätsfall für die Ungleichung
- ein Vergleich der Auswahlwahrscheinlichkeiten

Gemäß Satz I.1 gilt die Identität in der Ungleichung (17) , wenn $\lambda_N = 1$ ist.
In diesem Fall sind die Werte $z_1 \, , \, . \, . \, . \, , \, z_N$ paarweise verschieden, oder mit anderen Worten:
X bildet die Menge Ω_N bijektiv ab auf die Menge der paarweise verschiedenen Werte $z_1 \, , \, . \, . \, . \, , \, z_N$.

Dann ist die Größe $|\, D(n, N;\ \vec{N},\ \vec{n})\,|$ genau der Abstand zwischen den Auswahl-Wahrscheinlichkeiten von n Objekten, der aufgrund der Verschiedenartigkeit des Ziehens mit bzw. ohne Zurücklegen aus Ω_N zustandekommt ($n \leq N$).

W.Feller (1950, S.25) hat den Unterschied zwischen diesen beiden Auswahl-Wahrscheinlichkeiten verbal verglichen. Er stellt fest: Das n-malige Ziehen ohne Zurücklegen aus Ω_N unterscheidet sich vom n-maligen Ziehen mit Zurücklegen aus Ω_N nur relativ wenig, wenn N hinreichend groß und dazu n relativ klein ist.

Lemma I.3 liefert eine Präzisierung dieses Vergleiches.

Im Falle $\lambda_N = 1$ gilt für alle zugehörigen, gemäß (1) und (2) zulässigen Datensätze $\vec{N}$, $\vec{n}$:

$$|\, D(n, N;\ \vec{N},\ \vec{n})\,| \;=\; B(n,\ N,\ \frac{1}{N}) \;=\; \frac{1}{\prod_{i=0}^{n-1}(N-i)} - \frac{1}{N^n}\ . \tag{23}$$

Es wird nun gezeigt, dass dieser Ausdruck zu jedem festen $N > 4$ streng monoton fallend in n ist für alle $n \in \{3, \ldots, N\text{-}1\}$.

Zur Schreibvereinfachung wird abgekürzt:

$$\frac{1}{\prod_{i=0}^{n-1}(N-i)} - \frac{1}{N^n} = \frac{(N-n)!}{N!} - \frac{1}{N^n} =:\ g_N(n)\ .$$

Lemma I.3

Sei $N > 4$, dann gilt:

$g_N(n-1) > g_N(n)$ für alle $n \in \{3, \ldots, N\text{-}1\}$.

Beweis:

Es gilt:

$$g_N(n-1) > g_N(n) \iff \frac{(N-n+1)!}{N!} - \frac{1}{N^{n-1}} > \frac{(N-n)!}{N!} - \frac{1}{N^n}$$

$$\iff \frac{(N-n+1)! - (N-n)!}{N!} > \frac{1}{N^{n-1}} - \frac{1}{N^n}$$

$$\iff \frac{(N-n)!\cdot(N-n)}{N!} > \frac{N-1}{N^n} \iff (N-n)!\cdot(N-n)\cdot N^n > (N-1)\cdot N!$$

Vollständige Induktion.

$n = 3$ (Induktions-Anfang):

$$(N-3)! \cdot (N-3) \cdot N^3 > (N-1) \cdot N!$$

$$\Longleftrightarrow (N-3)\cdot N^3 > (N-1)\cdot N\cdot(N-1)\cdot(N-2) \Longleftrightarrow (N-3)\cdot N^2 > (N-1)^2\cdot(N-2)$$

$$\Longleftrightarrow N^3 - 3\cdot N^2 > N^3 - 4\cdot N^2 + 5\cdot N - 2 \Longleftrightarrow N^2 - 5\cdot N + 2 > 0 \;.$$

Dies ist genau für alle natürlichen Zahlen $N > 4$ erfüllt.

$(n-1) \to n$ (Induktions-Schritt) mit $2 \leq n-1$:

Es gilt:

$$(N-n)! \cdot (N-n) \cdot N^n = (N-n+1)! \cdot (N-n+1) \cdot N^{n-1} \cdot \frac{N \cdot (N-n)}{(N-n+1)^2} \;.$$

Gemäß Induktions-Voraussetzung folgt daraus:

$$(N-n)! \cdot (N-n) \cdot N^n > (N-1) \cdot N! \cdot \frac{N \cdot (N-n)}{(N-n+1)^2} \;. \tag{24}$$

Sei $N > n$, dann gilt: $\dfrac{N \cdot (N-n)}{(N-n+1)^2} \geq 1$.

Begründung:

$$N \cdot (N-n) \geq (N-n+1)^2 \Longleftrightarrow N^2 - n\cdot N \geq N^2 - 2\cdot(n-1) + 1$$

$$\Longleftrightarrow [\,2\cdot(n-1) - n\,]\cdot N - (n-1)^2 \geq 0$$

$$\Longleftrightarrow (n-2)\cdot N - (n-1)^2 \geq 0$$

$$\Longleftrightarrow N \geq \frac{(n-1)^2}{(n-2)} = n + \frac{1}{n-2}$$, und dies ist stets erfüllt.

Zusammen mit (24) folgt daher:

$$(N-n)! \cdot (N-n) \cdot N^n > (N-1) \cdot N! \cdot 1 \;.$$

Damit ist alles gezeigt.

❑

Bemerkungen zu Lemma I.3

Die Bedingung $n < N$ kann nicht auf $n = N$ erweitert werden.

Denn es ist mit jedem, festen $N > 1$ stets $g_N(N-1) < g_N(N)$:

$$g_N(N-1) < g_N(N) \iff \frac{1}{N!} - \frac{1}{N^{N-1}} < \frac{1}{N!} - \frac{1}{N^N}$$

$$\iff \frac{1}{N^N} < \frac{1}{N^{N-1}} \iff 1 < N \ .$$

Lemma I.3 liefert also für jedes feste $N > 4$ und alle $n \in \{ 3, \ldots, (N-1) \}$ die mathematische Begründung für folgende Interpretation:
Im Falle $\lambda_N = 1$ nimmt die stochastische Abhängigkeit der Stichproben-Variablen $X_1, \ldots, X_n$ beim Ziehen ohne Zurücklegen aus Ω_N mit wachsendem n ab. Die Abhängigkeit ist am geringsten, wenn $n = N-1$ ist, d.h. wenn die Stichprobe **fast** eine Vollerhebung ist.
Wegen der Bijektivität von X im Falle $\lambda_N = 1$ gilt diese Interpretation genauso für die Abhängigkeit von n Einzelentnahmen ohne Zurücklegen aus Ω_N.
Untersucht man z.B. die Qualität einer Schätzfunktion, wenn Ziehen ohne Zurücklegen aus Ω_N zugrundeliegt, führt die Anwendung von manchen mathematischen Sätzen bei großem Stichprobenumfang möglicherweise noch zu akzeptablen Ergebnissen, auch wenn diese Sätze die Unabhängigkeit der Stichproben-Variablen voraussetzen.
Anwender haben damit eine weitere Motivation für einen möglichst großen Stichprobenumfang neben jener, dass für viele Schätzfunktionen mit größerem Stichprobenumfang eine größere Präzision nachweisbar ist. Diese weitere Motivation ist zumindest dann gegeben, wenn die Werte des betrachteten Merkmals in Ω_N paarweise verschieden vorkommen (das ist der Regelfall bei Merkmalen wie z.B. Gewicht, Länge, Flächeninhalt, Volumen, Funktionsdauer etc.).

2.6 Die m-dimensionale Hypergeometrische Verteilung

Mit $Y_1, \ldots, Y_m$ werden die bereits in der Einführung angegebenen Häufigkeits-Variablen bezeichnet; Y_k gibt die Anzahl der Stichprobendaten unter den n Stück an, die mit dem X-Wert a_k übereinstimmen ($1 \leq k \leq m \leq N$) .

Es gibt eine enge Beziehung zwischen der gemeinsamen Verteilung der Stichproben-Variablen $X_1, \ldots, X_n$ und der gemeinsamen Verteilung der Häufigkeits-Variablen $Y_1, \ldots, Y_m$ - sowohl beim Ziehen mit wie beim Ziehen ohne Zurücklegen aus Ω_N. Die Ungleichung (17) kann genutzt werden, um einen Grenzwertsatz für die gemeinsame Wahrscheinlichkeitsfunktion der $Y_1 , \dots , Y_m$ zu formulieren, der eine Verallgemeinerung und Verbesserung eines Grenzwertsatzes von W.Feller darstellt. Zunächst werden die m-dimensionale Hypergeometrische Verteilung und die m-fache Multinomialverteilung definiert ($2 \le m \le N$).

$\vec{N}$, $\vec{n}$ seien gemäß (1) und (2) zulässige Datensätze. Die Indizes an den Wahrscheinlichkeitsmaßen P_{ZoZ} und P_{ZmZ} kennzeichen, dass Ziehen ohne bzw. mit Zurücklegen zugrundeliegt.
Beim Ziehen ohne Zurücklegen ist die gemeinsame Wahrscheinlichkeitsfunktion der $Y_1 , \dots , Y_m$ die der m-dimensionalen Hypergeometrischen Verteilung.
Für $n_i \in \{0 , 1 , \dots , N_i \}$, $i \in \{ 1 , \dots , m \}$ gilt (L.J.Bain / M.Engelhardt 1992, S. 138):

$$P_{ZoZ}(Y_1 = n_1 , \dots , Y_m = n_m) = \frac{\binom{N_1}{n_1} \cdot \dots \cdot \binom{N_m}{n_m}}{\binom{N}{n}} \tag{25}$$

$$= \frac{n!}{n_1! \cdot \dots \cdot n_m!} \cdot \frac{N_1! \cdot \dots \cdot N_m! \cdot (N-n)!}{(N_1-n_1) \cdot \dots \cdot (N_m-n_m) \cdot N!}$$

$$= \frac{n!}{n_1! \cdot \dots \cdot n_m!} \cdot \frac{\prod_{j=1}^{m} \prod_{k=0}^{n_j-1} (N_j - k)}{\prod_{i=0}^{n-1} (N - i)}$$

$$= \frac{n!}{n_1! \cdot \dots \cdot n_m!} \cdot \frac{\left(\prod_{k=1}^{m} N_k^{n_k}\right)\left(\prod_{j=1}^{m} \prod_{i=0}^{n_j-1} \left(1 - \frac{i}{N_j}\right)\right)}{\prod_{i=0}^{n-1} (N - i)} .$$

Beim Ziehen mit Zurücklegen ist die gemeinsame Wahrscheinlichkeitsfunktion der $Y_1 , \dots , Y_m$ die der m-fachen Multinomialverteilung. Es gilt (L.J.Bain / M.Engelhardt 1992, S. 138):

$$P_{ZmZ}(Y_1 = n_1 , \dots , Y_m = n_m) = \frac{n!}{n_1! \cdot \dots \cdot n_m!} \cdot \prod_{k=1}^{m} \left(\frac{N_k}{N}\right)^{n_k} \tag{26}$$

$$= \frac{n!}{n_1! \cdot \dots \cdot n_m!} \cdot \frac{\prod_{k=1}^{m} N_k^{n_k}}{N^n}$$

Mit der in (5) definierten Größe $D(n , N ; \vec{N} , \vec{n})$ ergibt sich daraus folgende Darstellung:

$$P_{ZoZ}(Y_1 = n_1 , \ldots , Y_m = n_m) - P_{ZmZ}(Y_1 = n_1 , \ldots , Y_m = n_m) = \frac{n!}{n_1! \cdot \ldots \cdot n_m!} \cdot D(n,N;\vec{N},\vec{n}) \qquad (27)$$

2.6.1 Ein Grenzwertsatz

Die Darstellung (27) ermöglicht eine direkte Anwendung von Satz I.2 auf den Abstand der Wahrscheinlichkeitsfunktionen $P_{ZoZ}(Y_1 = n_1 , \ldots , Y_m = n_m)$ und $P_{ZmZ}(Y_1 = n_1 , \ldots , Y_m = n_m)$. Es gilt der folgende Grenzwertsatz.

Satz I.3

Zu festen $n \in \mathbb{N}$ und $m \in \{2 , \ldots , N\}$ wird unter jeder der Voraussetzungen von Satz I.1 bei hinreichend großem $N \in \mathbb{N}$ die m-dimensionale Hypergeometrische Verteilung durch die m-fache Multinomialverteilung beliebig genau approximiert.

Sei $\mathbb{M}_m \subset \mathbb{R}^m$ die Menge aller gemäß (1) und (2) zulässigen $\vec{n}$. Dann gilt mit einem $r \geq 1$ für den Abstand der in (25) und (26) angegebenen Wahrscheinlichkeitsfunktionen der beiden Verteilungen:

$$\sup_{\vec{n} \in \mathbb{M}_m} \left| P_{ZoZ}\left(Y_1{=}n_1, \ldots, Y_m{=}n_m \right) - P_{ZmZ}\left(Y_1{=}n_1, \ldots, Y_m{=}n_m \right) \right| = O(N^{-r}) ,$$

N -> +oo .

Beweis:

Aus (27) ergibt sich unter jeder der Voraussetzungen von Satz I.1 mit allen gemäß (1) und (2) zulässigen Datensätzen $\vec{N}$, $\vec{n}$:

$$\left| P_{ZoZ}(Y_1 = n_1 , \ldots , Y_m = n_m) - P_{ZmZ}(Y_1 = n_1 , \ldots , Y_m = n_m) \right| = \frac{n!}{n_1! \cdot \ldots \cdot n_m!} \cdot \left| D(n,N;\vec{N},\vec{n}) \right| \leq n! \cdot \left| D(n,N;\vec{N},\vec{n}) \right| . \qquad (28)$$

Wegen Satz I.2 gilt daher zu festem $n \in \mathbb{N}$ unter jeder der Voraussetzungen von Satz I.1 mit einem $r \geq 1$ für jedes $m \in \{2, \ldots, N\}$:

$$\left| \; P_{ZoZ}(\, Y_1 = n_1 \;, \ldots , \; Y_m = n_m) \;-\; P_{ZmZ}(\, Y_1 = n_1 \;, \ldots , \; Y_m = n_m) \; \right| \;=\; O\left(N^{-r} \right) \;, \qquad (29)$$

$N \to +\infty$.

Wegen Satz I.2 hängt die Beziehung (29) nicht vom speziellen Datensatz $\vec{n} \in \mathbb{M}_m$ ab.

Daher konvergiert der Abstand beider Wahrscheinlichkeitsfunktionen mit N -> +oo gleichmäßig in $\mathbb{M}_m$ gegen Null zur Ordnung $O(N^{-r})$.

❑

2.6.2 Der Vergleich mit einem Resultat von W.Feller

Die Ergebnisse von Abschnitt 2.6.1 werden verglichen mit dem „Grenzwertsatz für die Hypergeometrische Verteilung“ von W.Feller (1953, S.47).

W.Feller's Grenzwertsatz ist äquivalent mit folgender Ungleichungskette:

$$\binom{n}{n_1} \cdot \left[\left(\frac{N_1 - n_1}{N} \right)^{n_1} \cdot \left(\frac{N_2 - n_2}{N} \right)^{n_2} - \left(\frac{N_1}{N} \right)^{n_1} \cdot \left(\frac{N_2}{N} \right)^{n_2} \right]$$

$$< \; P_{ZoZ}(\; Y_1 = n_1 \;,\; Y_2 = n_2 \;) \;-\; P_{ZmZ}(\; Y_1 = n_1 \;,\; Y_2 = n_2 \;) \qquad (30)$$

$$< \; \binom{n}{n_1} \cdot \left(\frac{N_1}{N} \right)^{n_1} \cdot \left(\frac{N_2}{N} \right)^{n_2} \cdot \left[\left(\frac{N}{N-n} \right)^{n_1} - 1 \right]$$

mit beliebig gewählten $n_k \in \{ 0, 1, \ldots, \min\{n, N_k\} \}$, $k \in \{ 1, 2 \}$; $n_1 + n_2 = n$.

Es gibt mehrere Kritikpunkte an W.Feller's Resultat:

(1.) Die angegebenen Grenzen der Differenz (30) sind Daten-bezogen und Verteilung-bezogen, genauer: sie hängen von den absoluten Häufigkeiten n_1 , n_2 der paarweise verschiedenen Werte unter den Daten $x_1, \ldots, x_n$ sowie von den Größen $\frac{N_1}{N}$, $\frac{N_2}{N}$ der Verteilung $\mathcal{F}(N)$ ab.

Aus obiger Abschätzung der Differenz (30) ergibt sich deren Konvergenz gegen Null nur punktweise. Eine Aussage über die Geschwindigkeit dieser Konvergenz macht W.Feller nicht.

(2.) Die Abschätzung von (30) ist nicht scharf.

Die Beziehungen (28) und (29) liefern eine Verbesserung hinsichtlich von Punkt **(1.)** , und zwar für beliebiges, festes $m \in \{2 , . . . , N\}$.

Eine Verbesserung hinsichtlich von Punkt **(2.)** liefert die Abschätzung (28) :

Wie in der Ungleichung (17) gilt auch in (28) unter der Voraussetzung (i) von Satz I.1 das Gleichheitszeichen. Denn aus $\lambda_N = 1$ folgt $N_i = 1$ für alle $i \in \{ 1 , . . . , m \}$, und daher ist $n_i \in \{ 0 ; 1 \}$ wegen (2) für alle $i \in \{ 1 , . . . , m \}$.

In obiger Abschätzung von (30) sind Gleichheitszeichen ausgeschlossen.

Daher stellt (28) eine Verschärfung der Abschätzung von (30) dar.

Weil (29) für beliebiges $m \in \{ 2 , . . . , N \}$ gültig ist, ist der Grenzwertsatz von W.Feller mit $m = 2$ als spezieller Fall in (28) bzw. (29) enthalten.

Mit den in Abschnitt 2.6.1 erhaltenen Ergebnissen wird daher eine Verbesserung und Verallgemeinerung des Grenzwertsatzes von W.Feller erreicht.

3. Ergänzungen

Es erfolgen nun Ergänzungen zum Fall $1 < \lambda_N$ (mit $2 < n \leq N$).

Wegen Satz I.1 gilt:

$$B\left(n , N , \frac{\lambda_N}{N} \right) \leq B\left(n , N , \frac{N-1}{N} \right)$$

$$= \left(\frac{N-1}{N}\right)^{n-1} \cdot \prod_{i=0}^{n-1} \left(\frac{N}{N-i}\right) \cdot \frac{1}{N} \cdot \left[(n-2) \cdot \left(N \cdot \min\left\{ \frac{n}{N} ; \frac{N-1}{N} \right\} -1 \right) - \left(1 - \prod_{j=0}^{n-1} \left(1 - \frac{j}{N} \right) \right) \right]$$

$$= \prod_{i=0}^{n-1} \left(\frac{N-1}{N-i}\right) \cdot \frac{1}{N} \cdot \left[(n-2) \cdot \min\{ n-1 ; N-2 \} - \left(1 - \prod_{j=0}^{n-1} \left(1 - \frac{j}{N} \right) \right) \right] .$$

Daraus erhält man unter jeder der Voraussetzungen (ii) bis (v) von Satz I.1 mit allen zugehörigen, gemäß (1) und (2) zulässigen Datensätzen $\vec{N}$, $\vec{n}$:

$$| D (n , N ; \vec{N} , \vec{n}) | \leq B\left(n , N , \frac{\lambda_N}{N} \right) \leq B\left(n , N , \frac{N-1}{N} \right) =$$

$$\min \left\{ \prod_{i=0}^{n-1} \left(\frac{N-1}{N-i}\right) \cdot \frac{1}{N} \cdot \left[(n-2) \cdot \min\{ n-1 ; N-2 \} - \left(1 - \prod_{j=0}^{n-1} \left(1 - \frac{j}{N} \right) \right) \right] ; \quad 1 \right\} \tag{31}$$

Auf der in der Einführung genannten Jahrestagung wurde vom Autor die folgende Ungleichung vorgestellt, die mit allen gemäß (1) und (2) zulässigen Datensätzen $\vec{N}$, $\vec{n}$ gültig ist:

$$| \, D(n, N; \vec{N}, \vec{n}) \, | \leq \min \left\{ \prod_{i=0}^{n-1} \left(\frac{N-1}{N-i} \right) \cdot \frac{1}{N} \cdot (n-1)^2 \; ; \; 1 \right\} . \tag{32}$$

Selbst nach Einsetzen der „trivialen“ Obergrenze $\frac{N-1}{N}$ für $\frac{\lambda_N}{N}$ in $B(n, N, \frac{\lambda_N}{N})$ erhält man mit der Obergrenze (31) eine Verbesserung gegenüber der Abschätzung (32).
Bei der Entwicklung der Beziehung (32) ging es vorrangig um die Betrachtung des asymptotischen Verhaltens von $| \, D(n, N; \vec{N}, \vec{n}) \, |$ für $N \to +\infty$.

Für eine solche Betrachtung kann (31) weiter vereinfacht werden. Man erhält mit $3 < n \leq N$:

$$\begin{aligned} | \, D(n, N; \vec{N}, \vec{n}) \, | &\leq \prod_{i=0}^{n-1} \left(\frac{N-1}{N-i} \right) \cdot \frac{1}{N} \cdot (n-2) \cdot \min\{ n-1 ; N-2 \} \\ &< \left(\frac{N-1}{N-n+1} \right)^{n-2} \cdot \frac{(n-2) \cdot (n-1)}{N} . \end{aligned} \tag{33}$$

Im II.Kapitel wird die Abschätzung von $\frac{\lambda_N}{N}$ eingehend untersucht. Unter bestimmten Annahmen läßt sich eine Verbesserung gegenüber der „trivialen“ Obergrenze $\frac{N-1}{N}$ erzielen, und damit durch Einsetzen eine Verbesserung der Abschätzung von $| \, D(n, N; \vec{N}, \vec{n}) \, |$.