Z. Wahrscheinlichkeitstheorie verw. Geb. 14, 189–199 (1970)

Eine Approximation der Binomialverteilung durch die Normalverteilung der Ordnung $1/n$

Rudolf Borges

Summary. The class (1) of transformations of a binomial variable k is studied. The famous De Moivre-Laplace local and integral limit theorems are generalized for the asymptotically standardized transformations (2). The transformation (4) respectively (3) is singled out as the only one with an error of order $1/n$. Besides it is shown, that the error term of order $n^{-\frac{1}{2}}$ of the identical transformation, the angular transformation, and the logarithmic transformation are in proportion to $\frac{1}{6}$, $-\frac{1}{12}$, and $-\frac{1}{3}$. Results on the influence of a correction of unbiasedness are mentioned in the last section. Such corrections allow a slight improvement of our new transformation.

Einleitung

Es sei k nach $B(n, p)$ binomialverteilt mit $0<p<1$ und $q=1-p$. Wir betrachten die Familie aller Transformationen

$$u_k=\varphi\left(\frac{k+d/2}{n+d}\right) \tag{1}$$

mit einer in $0<t<1$ (oder $0\leqq t\leqq 1$) dreimal stetig differenzierbaren Funktion $\varphi(t)$ mit $\varphi'(t)>0$ und einer Konstanten $d\geqq 0$. Es ist im wesentlichen nach Curtiss (1943) bekannt, daß u_k für große n näherungsweise nach $N(\varphi(p), \varphi'(p)^2\, p\, q(n+d)^{-1})$ normal verteilt ist, d.h. daß

$$x_k=\left(\varphi\left(\frac{k+d/2}{n+d}\right)-\varphi(p)\right)\frac{1}{\varphi'(p)}\sqrt{\frac{n+d}{p\,q}} \tag{2}$$

asymptotisch nach $N(0, 1)$ normalverteilt ist (bei Curtiss steht $\mathscr{E}(u_k)$ und $\mathscr{V}\,u_k$ anstelle von $\varphi(p)$ und $\varphi'(p)^2\,p\,q(n+d)^{-1}$). Wie Curtiss ausdrücklich bemerkt, liefert sein Beweis keine Möglichkeit, verschiedene u_k zu vergleichen, da dieser keine Fehlerabschätzung liefert. Die von Blom (1954) angegebenen formalen Umformungen der asymptotischen Reihen von Cornish-Fisher gestatten zwar einen Vergleich der Fehlerglieder der Ordnung $n^{-\frac{1}{2}}$ im Falle $d=1$, sind aber unbefriedigend, da die Zulässigkeit der formalen Umformungen nicht nachgewiesen wird.

Wir verallgemeinern daher die für $\varphi(t)=t$ und $d=0$ klassischen lokalen und Integral-Grenzwertsätze von De Moivre-Laplace. Daraus ergibt sich, daß es genau eine Transformation (2) gibt, deren Fehler bei beschränktem x_k von der Ordnung n^{-1} ist, nämlich:

$$y_k=(p\,q)^{-\frac{1}{6}}\sqrt{n+1/3}\int\limits_{p}^{\frac{k+1/6}{n+1/3}}[s(1-s)]^{-\frac{1}{3}}\,ds. \tag{3}$$

Es ist also in (2) $\varphi(t)=\int[t(1-t)]^{-\frac{1}{3}}\,dt$ und $d=\frac{1}{3}$ zu setzen, damit man einen Fehler der Ordnung n^{-1} erhält, während für jede andere Wahl von $\varphi(t)$ und d der Fehler von der Ordnung $n^{-\frac{1}{2}}$ ist. Benötigt man also in statistischen Anwendungen eine näherungsweise normalverteilte Transformation einer binomial verteilten zufälligen Größe, so empfiehlt es sich, die Transformation[1]

$$v_k = \int_0^{\frac{k+1/6}{n+1/3}} [s(1-s)]^{-\frac{1}{3}}\, ds \tag{4}$$

anstelle von

$$t_k = t(k, d) = \frac{k+d/2}{n+d} \tag{5}$$

oder $2 \arcsin \sqrt{t_k}$ oder gar $\log[t_k/(1-t_k)]$ zu verwenden, es sei denn, daß andere Gründe eine dieser Transformationen, z.B. in der Varianzanalyse, notwendig machen. Aus unserer Verallgemeinerung der De Moivre-Laplaceschen Grenzwertsätze ergibt sich noch, daß bei den drei zuletzt genannten Transformationen sich die Fehlerglieder der Ordnung $n^{-\frac{1}{2}}$ im wesentlichen wie $\frac{1}{6} : -\frac{1}{12} : -\frac{1}{3}$ verhalten.

Für den nummerischen Vergleich der neuen Transformation (3) mit anderen bekannten Approximationen verweisen wir auf Gebhardt (1969) in Verbindung mit Raff (1956). Für gute nummerische Approximationen mit Hilfe der Poisson-verteilung sei auf Bolshev (1963) und Molenaar (1969) hingewiesen. Molenaar gibt auch ausführliche nummerische Vergleiche an.

Für eine gewisse Literaturübersicht und Zusammenstellung über die Approximation der Binomialverteilung durch die Normalverteilung sei noch auf Govindarajulu (1963) verwiesen. Ausdrücklich möchten wir erwähnen, daß Feller (1945) ebenfalls, eine Transformation z_k mit einem Fehler der Ordnung $1/n$ bei beschränktem z_k angegeben hat, die eine Modifikation einer Approximation Bernsteins (1943) gleicher Ordnung darstellt:

$$z_k = \sqrt{\frac{n+1}{pq}} \left\{ (t(k,1)-p) - \frac{q-p}{6pq} (t(k,1)-p)^2 \right\} - \frac{q-p}{3\sqrt{(n+1)pq}}. \tag{6}$$

Zum Vergleich mit (3) beachte man, daß bei beschränktem y_k gilt:

$$\begin{aligned} z_k &= \sqrt{\frac{n+1/3}{pq}} \left\{ (t(k,\tfrac{1}{3})-p) - \frac{q-p}{6pq} (t(k,\tfrac{1}{3})-p)^2 \right\} + O(1/n) \qquad (7) \\ &= y_k + O(1/n). \end{aligned}$$

Die Fellersche Transformation hat offensichtlich eine einfachere funktionale Form, aber den Nachteil, daß ihr keine für die Anwendungen brauchbare Transformation der Gestalt (1) zugeordnet ist, die man ohne Kenntnis von p verwenden kann. Ein Vorteil der Fellerschen Transformation ist, daß er dazu den Fehler für den Integralgrenzwertsatz explizit angibt.

Wir bemerken noch, daß Bernstein (1943), Feller (1945) und Prohorov (1953) die Voraussetzung, daß z.B. z_k beschränkt ist fallen lassen und durch die Voraussetzung $|z_k| \leq C n^{\frac{1}{6}}$ ersetzen, wie wir dies in unseren Grenzwertsätzen tun werden.

Pratt (1968) gibt eine Approximation der Ordnung $n^{-\frac{3}{2}}$ der Binomialverteilung durch die Normalverteilung an und diskutiert sie zusammen mit Peizer (1968).

Tsaregradskii (1958) und Meshalkin (1960) geben Approximationen der Ordnung $1/n$ der Binomialverteilung durch unendlichteilbare Verteilungen an.

[1] Eine Tafel von $\int [s(1-s)]^{-\frac{1}{3}}\, ds$ ist von Herrn F. Gebhardt, Deutsches Rechenzentrum, 61 Darmstadt, Rheinstr. 75 erhältlich.

Ergebnisse

Als erstes geben wir eine Verallgemeinerung des klassischen lokalen und Integralgrenzwertsatzes von De Moivre-Laplace für die Binomialverteilung

$$b(k; n, p) = \binom{n}{k} p^k q^{n-k} \tag{8}$$

an.

1. Lokaler Grenzwertsatz. *Es sei* $0<p<1$ *und* $q=1-p$ *sowie* d, D *Konstanten mit* $0 \leqq d, D < +\infty$. *Erfüllt die in* (2) *definierte Transformation* x_k

$$|x_k| \leqq D(n+d)^{\frac{1}{6}}, \tag{9}$$

dann gilt

$$b(k; n, p) = \frac{d x_k}{d k} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x_k^2} \left(1 + r(x_k)\right) \tag{10}$$

mit

$$\begin{aligned} r(x) = & \frac{[(q-p)\,\varphi'(p) + 3\,p\,q\,\varphi''(p)]\,x^3 + [3(d-1)(q-p)\,\varphi'(p) - 6\,p\,q\,\varphi''(p)]\,x}{6\varphi'(p)\sqrt{(n+d)\,p\,q}} \\ & + O\left(\frac{x^6 + x^4 + x^2 + 1}{n}\right). \end{aligned} \tag{11}$$

Dabei hängt die Konstante im Restglied von D ab.

Bemerkung. Für den klassischen Fall $\varphi(t)=t$ und $d=0$ wurde dies von Prohorov (1953) auf S. 141 unten bewiesen. Feller (1957) empfiehlt auf S. 182 die Variante $d=1$.

Wir bemerken, daß mit t_k, definiert in (5)

$$\frac{d x_k}{d k} = \frac{\varphi'(t_k)}{\varphi'(p)\sqrt{(n+d)\,p\,q}}. \tag{12}$$

Mit Hilfe der Eulerschen Summenformel folgt aus dem lokalen Grenzwertsatz:

1. Integralgrenzwertsatz. *Es seien die Voraussetzungen des lokalen Grenzwertsatzes erfüllt und* $a \leqq b$ *ganze Zahlen, so daß* $A := x_{a-\frac{1}{2}}$ *und* $B := x_{b+\frac{1}{2}}$ *die Ungleichungen*

$$|A| \leqq D(n+d)^{\frac{1}{6}}, \qquad |B| \leqq D(n+d)^{\frac{1}{6}} \tag{13}$$

erfüllen. Dann gilt

$$\sum_{k=a}^{b} b(k; n, p) = \frac{1}{\sqrt{2\pi}} \int_A^B \exp\left(-\tfrac{1}{2} x^2\right) dx + R(A, B) \tag{14}$$

mit

$$\begin{aligned} R(A, B) = & \frac{(q-p)\,\varphi'(p) + 3\,p\,q\,\varphi''(p)}{6\varphi'(p)\sqrt{(n+d)\,p\,q}\sqrt{2\pi}} \left(A^2 e^{-\frac{A^2}{2}} - B^2 e^{-\frac{B^2}{2}}\right) \\ & + \frac{(3d-1)(q-p)\,\varphi'(p)}{6\varphi'(p)\sqrt{(n+d)\,p\,q}\sqrt{2\pi}} \left(e^{-\frac{A^2}{2}} - e^{-\frac{B^2}{2}}\right) + O(n^{-1}), \end{aligned} \tag{15}$$

wobei die Konstante in dem Restglied $O(n^{-1})$ *nur von D abhängt.*

1. Zusatz. *Das Fehlerglied der Ordnung $n^{-\frac{1}{2}}$ ist in dem lokalen und Integral-Grenzwertsatz genau dann Null, wenn $\varphi(t) = C\int[t(1-t)]^{-\frac{1}{3}}dt$ und $d=\frac{1}{3}$, d.h. wenn $x_k = y_k$.*

Der Beweis folgt daraus, daß das angegebene $\varphi(t)$ die allgemeine Lösung der Differentialgleichung

$$(1-2p)\,\varphi'(p) + 3p(1-p)\,\varphi''(p) = 0 \tag{16}$$

ist.

Durch einfaches Einsetzen erhalten wir

2. Zusatz. *Hat x_k die Gestalt*

$$x_k = x_{k,\alpha} = (p\,q)^{\alpha-\frac{1}{2}}\sqrt{(n+d)}\int_p^{t_k}[s(1-s)]^{-\alpha}\,ds, \tag{17}$$

so gilt für die Fehlerglieder im lokalen und Integralgrenzwertsatz

$$\begin{aligned} r_\alpha(x) = {} & \frac{(q-p)}{6\sqrt{(n+d)\,p\,q}}\,[(1-3\alpha)\,x^3 + (3d-3+6\alpha)\,x] \\ & + O\left(\frac{x^6+x^4+x^2+1}{n}\right), \end{aligned} \tag{18}$$

$$\begin{aligned} R_\alpha(A,B) = {} & \frac{(q-p)}{6\sqrt{2\pi(n+d)\,p\,q}}\,[(1-3\alpha)\left(A^2 e^{-\frac{A^2}{2}} - B^2 e^{-\frac{B^2}{2}}\right) \\ & + (3d-1)\,(e^{-\frac{A^2}{2}} - e^{-\frac{B^2}{2}})] + O(n^{-1}). \end{aligned} \tag{19}$$

Setzt man noch zusätzlich $d=\frac{1}{3}$, so ergibt sich

$$r_{\alpha,\frac{1}{3}}(x) = \frac{(q-p)\,(1-3\alpha)}{6\sqrt{(n+1/3)\,p\,q}}\,(x^3-2x) + O(n^{-1}), \tag{20}$$

$$R_{\alpha,\frac{1}{3}}(A,B) = \frac{(q-p)\,(1-3\alpha)}{6\sqrt{2\pi(n+1/3)\,p\,q}}\left(A^2 e^{-\frac{A^2}{2}} - B^2 e^{-\frac{B^2}{2}}\right) + O(n^{-1}). \tag{21}$$

Beispiele zum 2. Zusatz. Setzen wir in (17) der Reihe nach $\alpha = 0, \frac{1}{3}, \frac{1}{2}, 1$, so erhalten wir Grenzwertsätze für die Transformationen t_k, die Transformation v_k definiert in (4), die Fishersche Winkeltransformation und die logarithmische Transformation. Im Fall $d=\frac{1}{3}$ sind die Koeffizienten des Fehlergliedes der Reihe nach $\frac{1}{6}, 0, -\frac{1}{12}, -\frac{1}{3}$. Die Approximation durch die Normalverteilung ist also bei der logarithmischen Transformation nur etwa halb so gut wie bei der klassischen, während die Winkeltransformation eine doppelt so gute Approximation liefert. Bei der Transformation v_k bzw. y_k ist dagegen die Ordnung des Fehlers n^{-1} anstelle von $n^{-\frac{1}{2}}$.

Bemerkung. Für $d=1$ wurde von Blom (1954) die Familie (17) von Transformationen aus der Differentialgleichung

$$(B^2+2)\,(1-2p)\,\varphi'(p) + 3B^2\,p(1-p)\,\varphi''(p) = 0$$

abgeleitet, die man wie (16) aus (15) erhält, wenn man $A=-\infty$ setzt. Wegen $\alpha = (B^2+2)/3B^2$ erhält Blom $\alpha=\frac{1}{3}$ für $B\to\infty$.

Zur Untersuchung des Supremums des Fehlergliedes der Ordnung $n^{-\frac{1}{2}}$ im 1. Integralgrenzwertsatz und im 2. Zusatz setzen wir

$$M(s) := \sup\{(y^2+s)\exp(-\tfrac{1}{2}y^2) - (x^2+s)\exp(-\tfrac{1}{2}x^2) \colon x, y \in \mathbb{R}\}. \tag{22}$$

Man überzeugt sich leicht, daß $M(s) = s$ bzw. $2e^{-1+s/2}$ bzw. $2e^{-1+s/2} - s$, je nachdem $s \geqq 2$ bzw. $0 \leqq s \leqq 2$ bzw. $s \leqq 0$ gilt. Daher ist $M(s)$ für $s=0$ minimal. Daraus folgt

3. Zusatz. *Das Supremum des Fehlergliedes der Ordnung $n^{-\frac{1}{2}}$ in (15) über $-\infty \leqq A \leqq B \leqq +\infty$ ist bei gegebenem $\varphi(t)$ für $d = \frac{1}{3}$ minimal.*

Im Fall der Betatransformationen (17) mit $\alpha \neq \frac{1}{3}$ ist nach (19) und (22) das Supremum des Fehlergliedes der Ordnung $n^{-\frac{1}{2}}$ des Integralgrenzwertsatzes gleich

$$\frac{q-p}{6\sqrt{2\pi(n+d)\,p\,q}}\,|1-3\alpha|\,M\left(\frac{3d-1}{1-3\alpha}\right).$$

Die Tabelle illustriert die starke Abhängigkeit dieses Supremums von der Wahl von α und d und insbesondere den Vorteil der Wahl von $d = \frac{1}{3}$.

Tabelle. *Der Faktor $|1-3\alpha|\,M((3d-1)/(1-3\alpha))$ des Supremum des Fehlergliedes der Ordnung $n^{-\frac{1}{2}}$ des Integralgrenzwertsatzes für Betatransformationen*

Transformation	α	d						
		0	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{3}{4}$	1	α
t_k	0	1,45	0,90	0,74	0,94	1,37	2,00	1,45
$2\arcsin t_k^{1/2}$	$\frac{1}{2}$	1,00	0,47	0,37	0,72	1,36	2,05	0,72
$\log t_k/(1-t_k)$	1	1,89	1,57	1,47	1,80	2,33	2,89	2,89

2. Lokaler Grenzwertsatz. *Es sei $0<p<1$ und $q=1-p$ sowie y_k definiert durch (3) und D eine Konstante mit $D < +\infty$. Dann gilt*

$$b(k;n,p) = \frac{d\,y_k}{d\,k}\,\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}y_k^2}\left(1 + \frac{(1+2pq)(y_k^4-3)}{216(n+1/3)\,p\,q} + r_1(y_k)\right) \tag{23}$$

mit

$$r_1(y) = O\left(\frac{|y^5|+|y|}{n^{\frac{3}{2}}}\right) + O\left(\frac{y^8+y^4+1}{n^2}\right) \tag{24}$$

für $|y_k| \leqq D(n+1/3)^{\frac{1}{4}}$, wobei die Konstante im Restglied nur von D abhängt.

Wir bemerken, daß mit $t_k = (k+1/6)/(n+1/3)$ gilt:

$$\begin{aligned}\frac{d\,y_k}{d\,k} &= \frac{1}{[t_k(1-t_k)]^{\frac{1}{3}}\,(p\,q)^{\frac{1}{6}}\sqrt{n+1/3}}\\ &= \frac{1}{\sqrt{(n+1/3)\,p\,q}} - \frac{(1-2p)\,y_k}{3(n+1/3)\,p\,q} + O\left(\frac{y_k^2}{(n+1/3)^{\frac{3}{2}}}\right).\end{aligned} \tag{25}$$

2. Integralgrenzwertsatz. *Es sei* $0<p<1$ *und* $q=1-p$ *sowie* y_k *definiert durch* (3), *D eine Konstante mit* $D<+\infty$ *sowie* $a\leqq b$ *ganze Zahlen, so daß* $A:=y_{a-\frac{1}{2}}$ *und* $B:=y_{b+\frac{1}{2}}$ *die Ungleichungen erfüllen:*

$$|A|\leqq D(n+1/3)^{\frac{1}{4}}, \qquad |B|\leqq D(n+1/3)^{\frac{1}{4}}. \tag{26}$$

Dann gilt (14) *mit*

$$\begin{aligned} R(A,B)=&\frac{1}{216(n+1/3)\,p\,q\sqrt{2\pi}}\Big\{[(1+2pq)\,A^3-6(1-pq)\,A]\,e^{-\frac{A^2}{2}} \\ &-[(1+2pq)\,B^3-6(1-pq)\,B]\,e^{-\frac{B^2}{2}}\Big\}+O\big((n+1/3)^{-\frac{3}{2}}\big). \end{aligned} \tag{27}$$

Beweise der Grenzwertsätze

Zum Beweis des 1. lokalen Grenzwertsatzes verwenden wir die Stirlingsche Formel in der Gestalt von Barnes

$$n!=\sqrt{2\pi}\,(n+a)^{n+1/2}\exp\big(-(n+a)+\varepsilon(n,a)\big) \tag{28}$$

mit

$$\varepsilon(n,a)=O(n^{-1}) \tag{29}$$

für jede Konstante a. Genauer gilt für $0\leqq a\leqq 1$

$$\varepsilon(n,a)=\sum_{i=2}^{l}\frac{B_i(a)}{i(i-1)(n+a)^{i-1}}+O(n^{-l}) \tag{30}$$

(s. z.B. Whittaker and Watson (1927), Kap. XIII, § 13,6 mit $O(n^{-l+\frac{1}{2}})$ anstelle von $O(n^{-l})$ oder Rowe (1931) sowie Kalinin (1967) und (1967a)), wobei $B_i(t)$ das i-te Bernoullische Polynom ist. Speziell ist $B_2(a)=a^2-a+1/6$. Aus dieser Stirlingschen Formel folgt

$$\begin{aligned} \log b(k;n,p)=&(n+d)\left(t_k\log\frac{p}{t_k}+(1-t_k)\log\frac{q}{1-t_k}\right) \\ &+\frac{1-d}{2}\left(\log\frac{p}{t_k}+\log\frac{q}{1-t_k}\right)-\tfrac{1}{2}\log\big(2\pi(n+d)\,p\,q\big)+\rho \end{aligned} \tag{31}$$

mit

$$\rho=\varepsilon(n,d)-\varepsilon(k,d/2)-\varepsilon(n-k,d/2). \tag{32}$$

Für eine beliebige in $0<t<1$ monoton wachsende und dreimal stetig differenzierbare Funktion $\varphi(t)$ definieren wir Funktionen f_p und g_p durch

$$f_p(\varphi(t))=t\log(p\,t^{-1})+(1-t)\log\big(q(1-t)^{-1}\big), \tag{33}$$

$$g_p(\varphi(t))=\frac{1-d}{2}\log\frac{pq}{t(1-t)}-\log\frac{\varphi'(t)}{\varphi'(p)}. \tag{34}$$

Dann gilt wegen (12)

$$\begin{aligned} \log b(k;n,p)=&(n+d)\,f_p(\varphi(t_k))+g_p(\varphi(t_k)) \\ &+\log\frac{dx_k}{dk}-\frac{1}{2}\log 2\pi+\rho. \end{aligned} \tag{35}$$

Wir berechnen die Taylorschen Reihen von f_p und g_p an der Stelle $\varphi(p)$ mit dem Zuwachs $\varphi(t)-\varphi(p)$ unter Verwendung von (2) zu

$$\begin{aligned} f_p(\varphi(t_k)) &= -\frac{(\varphi(t_k)-\varphi(p))^2}{2\varphi'(p)^2\, p\, q} + \frac{(1-2p)\,\varphi'(p)+3\,p\,q\,\varphi''(p)}{6\varphi'(p)^4\, p^2\, q^2} \\ &\quad \cdot (\varphi(t_k)-\varphi(p))^3 + \tfrac{1}{24} f_p^{(4)}(\varphi(p'))\,(\varphi(t_k)-\varphi(p))^4 \\ &= \frac{1}{n+d}\left(-\frac{1}{2}x_k^2 + \frac{(q-p)\,\varphi'(p)+3\,p\,q\,\varphi''(p)}{6\varphi'(p)\sqrt{(n+d)\,p\,q}}\, x_k^3 \right. \\ &\quad \left. + \frac{1}{24(n+d)} f_p^{(4)}(\varphi(p'))\,\varphi'(p)^4\, p^2\, q^2\, x_k^4\right), \end{aligned} \tag{36}$$

$$\begin{aligned} g_p(\varphi(t_k)) &= +\left(\frac{(d-1)(1-2p)}{2p\; q\,\varphi'(p)} - \frac{\varphi''(p)}{\varphi'(p)^2}\right)(\varphi(t_k)-\varphi(p)) \\ &\quad + \tfrac{1}{2} g_p''(\varphi(p''))\,(\varphi(t_k)-\varphi(p))^2 \\ &= \frac{(d-1)(q-p)\,\varphi'(p)-2\,p\,q\,\varphi''(p)}{2\varphi'(p)\sqrt{(n+d)\,p\,q}}\, x_k + \frac{g_p''(\varphi(p''))\,\varphi'(p)^2\, p\, q\, x_k^2}{2(n+d)}. \end{aligned} \tag{37}$$

Dabei sind p', p'' Zwischenwerte von p und t_k. Nehmen wir noch an, daß $|x_k| \leqq D(n+d)^{\frac{1}{6}}$, so ist offensichtlich $\rho = O(n^{-1})$. Damit erhalten wir durch Einsetzen von (36) und (37) in (35) den 1. lokalen Grenzwertsatz.

Der Beweis des 2. lokalen Grenzwertsatzes verläuft wie der des ersten. Es ist lediglich ein weiteres Glied in den Reihenentwicklungen zu berechnen. Die spezielle Wahl von $d = \frac{1}{3}$ und $\varphi(t) = \int [t(1-t)]^{-\frac{1}{3}}\, dt$ vereinfacht jedoch die Rechnung wesentlich. Insbesondere ist jetzt

$$g_p = 0. \tag{38}$$

Weiter erhalten wir

$$\begin{aligned} (n+1/3)\, f_p(\varphi(t_k)) &= -\frac{1}{2}\, y_k^2 + \frac{1+2pq}{216(n+1/3)\, p\, q}\, y_k^4 \\ &\quad + \frac{f_p^{(5)}(\varphi(p'))\,(p\,q)^{\frac{5}{6}}}{120(n+1/3)^{\frac{3}{2}}}\, y_k^5. \end{aligned} \tag{39}$$

Dabei ist p' ein Zwischenwert von p und t_k. Aus der Bedingung $|y_k| \leqq D(n+1/3)^{\frac{1}{4}}$ folgt, daß es zu $\varepsilon > 0$ ein n_D mit $\varepsilon < t_k < 1-\varepsilon$ für $n \geqq n_D$ gibt. Dann gibt es ein $C < +\infty$ mit $|f^{(5)}(\varphi(p')| \leqq C$ für alle $n \geqq n_D$ und k mit $|y_k| \leqq D(n+1/3)^{\frac{1}{4}}$.

Weiter erhalten wir für den Fehler ρ mit (30)

$$\begin{aligned} \rho &= -\frac{1}{36(n+1/3)} - \frac{1}{72(k+1/6)} - \frac{1}{72(n-k+1/6)} + O((n+1/3)^{-2}) \\ &= -\frac{1}{36(n+1/3)} - \frac{1}{72(n+1/3)\, t_k(1-t_k)} + O((n+1/3)^{-2}) \\ &= -\frac{2\,p\,q+1}{72(n+1/3)\,p\,q} + \frac{(1-2p'')\,(p\,q)^{\frac{1}{4}}}{72(n+1/3)^{\frac{3}{2}}\,(p''\,q'')^2}\, y_k + O((n+1/3)^{-2}). \end{aligned} \tag{40}$$

Dabei ist p'' ein Zwischenwert von p und t_k. Setzen wir die drei letzten Gln. (38) bis (40) in (35) mit $y_k = x_k$ ein, so erhalten wir den 2. lokalen Grenzwertsatz.

Die zweite Gl. in (25) erhalten wir, indem wir

$$h_p(\varphi(t_k)) = \frac{d y_k}{dk} = \frac{1}{[t_k(1-t_k)]^{\frac{1}{3}} (p\,q)^{\frac{1}{6}} \sqrt{n+1/3}} \tag{41}$$

setzen und h_p in eine Taylorsche Reihe um die Stelle $\varphi(p)$ mit dem Zuwachs $\varphi(t_k) - \varphi(p)$ entwickeln:

$$h_p(\varphi(t_k)) = \frac{1}{\sqrt{(n+1/3)\,p\,q}} - \frac{1-2p}{3(n+1/3)\,p\,q}\, y_k + \frac{h_p^{(2)}(\varphi(p'''))\,(p\,q)^{\frac{1}{3}}}{2(n+1/3)}\, y_k^2. \tag{42}$$

Dabei ist p''' ein Zwischenwert von p und t_k. Beachtet man $h_p^{(2)}(\varphi(p''')) = O\left((n+1/3)^{-\frac{1}{2}}\right)$ unter den obigen Bedingungen, so folgt (25).

Zum Beweise des 2. Integralgrenzwertsatzes verwenden wir die Eulersche Summenformel in der Gestalt

$$\begin{aligned} \sum_{k=a}^{b} f(k) = \int_{a-1/2}^{b+1/2} f(s)\, ds - \tfrac{1}{24}\left(f'(b+\tfrac{1}{2}) - f'(a-\tfrac{1}{2})\right) \\ + \tfrac{1}{6} \int_{a-1/2}^{b+1/2} \overline{B_3(s)}\, f'''(s)\, ds. \end{aligned} \tag{43}$$

Dabei ist $\overline{B_3(s)}$ das dritte reduzierte Bernoullische Polynom, d.h. $\overline{B_3(s)} = s^3 - 3s^2/2 + s/2$ für $0 \leqq s < 1$ und $\overline{B_3(s)}$ periodisch mit Periode 1. Hier wählen wir

$$f(s) = \frac{d y_s}{ds}\, e^{-\frac{1}{2} y_s^2} \left(1 + \frac{(1+2p\,q)(y_s^4 - 3)}{216(n+1/3)\,p\,q}\right). \tag{44}$$

Mit (25) erhalten wir

$$f'(s) = -\frac{y_s}{(n+1/3)\,p\,q}\, e^{-\frac{1}{2} y_s^2} + O\left(\frac{1}{(n+1/3)^{\frac{3}{2}}}\right), \tag{45}$$

$$f'''(s) = y_s'\, e^{-\frac{1}{2} y_s^2}\, O\left(\frac{|y_s|^3}{(n+1/3)^{\frac{3}{2}}}\right). \tag{46}$$

Beachten wir noch

$$\int (t^4 - 3) \exp(-t^2/2)\, dt = -(t^3 + 3t) \exp(-t^2/2), \tag{47}$$

so erhalten wir durch Einsetzen von (44) bis (46) in (43) und eine elementare Umformung aus dem 2. lokalen Grenzwertsatz den 2. Integralgrenzwertsatz.

Unverfälschtheitskorrektur

Den Einfluß der in (48) angegebenen Unverfälschtheitskorrektur der Ordnung $1/n$ auf die Integralgrenzwertsätze wollen wir noch ohne Beweis angeben und diskutieren. Zunächst gilt unter gewissen Regularitätsvoraussetzungen

$$\mathscr{E}\,\varphi\left(\frac{k+d/2}{n+d}\right) = \varphi(p) + \frac{d(q-p)\,\varphi'(p) + p\,q\,\varphi''(p)}{2(n+d)} + O(n^{-2}). \tag{48}$$

Dies legt nahe die Transformation

$$\xi_k = \left\{ \varphi\left(\frac{k+d/2}{n+d}\right) - \varphi(p) - \frac{d(q-p)\,\varphi'(p) + p\,q\,\varphi''(p)}{2(n+c_1)} \right\} \frac{1}{\varphi'(p)} \sqrt{\frac{n+c}{p\,q}} \tag{49}$$

mit Konstanten c, c_1, $d \geqq 0$ anstelle von x_k definiert in (2) zu betrachten:

3. Integralgrenzwertsatz. *Es sei* $0<p<1$ *und* $q=1-p$ *sowie* c, c_1, d, D, *Konstanten mit* $0 \leqq c$, c_1, d, D, $<+\infty$ *und* $a \leqq b$ *ganze Zahlen, so daß* $A := \xi_{a-1/2}$ *und* $B := \xi_{b+1/2}$ *die Ungleichungen* (13) *erfüllen. Dann gilt* (14) *mit*

$$R(A,B) = \frac{(q-p)\,\varphi'(p) + 3\,p\,q\,\varphi''(p)}{6\varphi'(p)\sqrt{2\pi\, n\, p\, q}} \{(A^2-1)\, e^{-\frac{1}{2}A^2} - (B^2-1)\, e^{-\frac{1}{2}B^2}\} + O(1/n), \tag{50}$$

wobei die Konstante im Restglied nur von D abhängt.

Aus der Lösung der Differentialgleichung (16) erhalten wir

4. Zusatz. *Das Fehlerglied der Ordnung* $n^{-\frac{1}{2}}$ *im* 3. *Integralgrenzwertsatz ist genau dann Null, wenn* $\varphi(t) = C \int [t(1-t)]^{-\frac{1}{3}}\, dt$ *bei beliebiger Wahl von* c, c_1, *und* d.

Setzen wir diese Transformation in (49) ein, so erhalten wir

$$\eta_k = (p\,q)^{-\frac{1}{6}} \sqrt{n+c} \left\{ \int_p^{\frac{k+d/2}{n+d}} [s(1-s)]^{-\frac{1}{3}}\, ds - \frac{(3d-1)(q-p)}{6(n+c_1)(p\,q)^{\frac{1}{3}}} \right\} \tag{51}$$

anstelle von y_k definiert in (3). Also bedeutet die Wahl von $d=\frac{1}{3}$ im 1. Zusatz und damit in y_k bzw. v_k definiert in (4) die Vermeidung der Berechnung einer Unverfälschtheitskorrektur der Ordnung n^{-1}. Anstelle des 2. Integralgrenzwertsatzes kann man jetzt beweisen:

4. Integralgrenzwertsatz. *Es sei* $0<p<1$ *und* $q=1-p$ *sowie* c, c_1, d, D *Konstanten mit* $0 \leqq c$, c_1, d, $D<+\infty$ *und* $a \leqq b$ *ganze Zahlen, so daß für* $A := \eta_{a-1/2}$ *und* $B := \eta_{b+1/2}$ *gilt:*

$$|A| \leqq D\, n^{\frac{1}{6}}, \qquad |B| \leqq D\, n^{\frac{1}{6}}. \tag{52}$$

Dann gilt (14) *mit*

$$\begin{aligned} R(A,B) = {} & \frac{1}{216 n\, p\, q \sqrt{2\pi}} \{(1+2p\,q)(A^3\, e^{-\frac{1}{2}A^2} - B^3\, e^{-\frac{1}{2}B^2}) \\ & + [108(c-d)\, p\, q + 6(1-6d)(1-p\,q)]\,(A\, e^{-\frac{1}{2}A^2} - B\, e^{-\frac{1}{2}B^2})\} + O(n^{-\frac{3}{2}}). \end{aligned} \tag{53}$$

Ein Vergleich von (53) mit (27) zeigt, daß durch geeignete Wahl von c und d das Supremum des Fehlergliedes der Ordnung $1/n$ mit Hilfe der Unverfälschtheitskorrektur in η_k definiert in (51) gegenüber y_k definiert in (3) gesenkt werden kann.

Ein Vergleich von (50) mit (15) für $d=\frac{1}{3}$ zeigt dagegen nach dem 3. Zusatz, daß für große n bei Verwendung der Transformation ξ_k definiert in (49) anstelle von x_k mit $d=\frac{1}{3}$ definiert in (2) das Supremum des Approximationsfehlers größer ist, d.h. daß die Verwendung der Unverfälschtheitskorrektur weniger vorteilhaft ist als die Wahl von $d=\frac{1}{3}$ in (17) bzw. (2). Wie wir jedoch sogleich aus dem 5. Zusatz und der Tabelle entnehmen können, ist es günstiger im Falle der Winkeltransformation $2 \arcsin \sqrt{t_k}$ mit $d=0$ die Unverfälschtheitskorrektur zu verwenden.

Bekannt ist, daß Fisher (1930) mit Hilfe der Unverfälschtheitskorrektur in der Genetik ein für $n\to\infty$ asymptotisch besseres Resultat erzielt hat als ohne diese im Jahre 1922 und z.B. Stange (1963) in der Qualitätskontrolle mit der Korrektur nummerisch bessere Ergebnisse als ohne diese erzielte.

Im Fall der Betatransformationen $\varphi_\alpha(t)=\int[t(1-t)]^{-\alpha}\,dt$ erhalten wir als Spezialfall von (48)

$$\mathscr{E}\,\varphi_\alpha\left(\frac{k+d/2}{n+d}\right)=\varphi_\alpha(p)+\frac{(q-p)(d-\alpha)}{2(n+d)(p\,q)^\alpha}+O(n^{-2}). \tag{54}$$

Die Unverfälschtheitskorrektur der Ordnung $1/n$ verschwindet also für $d=\alpha$. Für den Fall $d=\alpha=\frac{1}{2}$ der Winkeltransformation ist dies nach Anscombe (1954) bekannt.

5. Zusatz. *Hat ξ_k die Gestalt*

$$\xi_k=\xi_{k,\alpha}=(p\,q)^{\alpha-1/2}\sqrt{n+c}\left(\int\limits_p^{\frac{k+d/2}{n+d}}[s(1-s)]^{-\alpha}\,ds-\frac{(q-p)(d-\alpha)}{2(n+d)(p\,q)^\alpha}\right), \tag{55}$$

so gilt für das Fehlerglied im 3. *Integralgrenzwertsatz*

$$R_\alpha(A,B)=\frac{(1-3\alpha)(q-p)}{6\sqrt{2\pi\,n\,p\,q}}\{(A^2-1)\,e^{-\frac{1}{2}A^2}-(B^2-1)\,e^{-\frac{1}{2}B^2}\}+O(1/n). \tag{56}$$

Das Fehlerglied der Ordnung $n^{-1/2}$ in (56) ist also genauso groß wie dasselbe Glied in (19) mit $d=\alpha$, wie zu erwarten war, da $\xi_{k,d}=x_{k,d}$. Wir haben deshalb in der Tabelle zum Vergleich die Spalte $d=\alpha$ explizit angegeben.

Literatur

Anscombe, F. J.: Discussion of the analysis of variance with various binomial transformations by R. Fisher. Biometrics **10**, 141–144 (1954).

Bernstein, S.: Retour au problème de l'évaluation de la formule limite de Laplace [Russisch]. Izvestija Akad. Nauk SSSR, Ser. mat. **7**, 3–16 (1943).

Blom, G.: Transformations of the binomial, negative binomial, Poisson and χ^2-distributions. Biometrika **41**, 302–316 (1954); **43**, 235 (1956).

Bol'shev, L. N.: Asymptotically Pearson's transformations [Russisch]. Teor. Verojatn. Primen. **8**, 129–155 (1963).

Curtiss, J. H.: On transformations used in the analysis of variance. Ann. math. Statistics **14**, 107–122 (1943).

Dyke, G. V., Patterson, H. D.: Analysis of factorial arrangements when the data are proportions. Biometrics **8**, 1–12 (1952).

Feller, W.: On the normal approximation to the binomial distribution. Ann. math. Statistics **16**, 319–329 (1945).

– An introduction to probability theory and its applications. Vol. I, 2. Aufl. New York: Wiley 1957.

Fisher, R. A.: On the dominance ratio. Proc. roy. Soc. Edinburgh **42**, 321–341 (1921–1922).

– The distribution of gene ratios for rare mutations. Proc. roy. Soc. Edinburgh, Sect. A **50**, 205–220 (1930).

Gebhardt, F.: Some numerical comparisons of several approximations to the binomial distribution. To appear in: J. Amer. statist. Assoc. December 1969.

Govindarajulu, Zakkula: Normal approximation to the classical discrete distributions. Sankhyā, Ser. A **27**, 143–167 (1965). (Abgedruckt aus: Classical and contagious distributions. Proc. International Symp. McGill Univ., Montreal, Canada, August 1963, ed. by Ganapati P. Patil.)

Kalinin, V. M.: A theorem on summation and its application to special functions [Russisch]. Doklady Akad. Nauk. SSSR **175**, 1004 – 1007 (1967).

– Convergent and asymptotic expansions for probability distributions [Russisch, englische Zusammenfassung]. Teor. Verojatn. Primen. **12**, 24 – 38 (1967a).

Meshalkin, L. D.: On the approximation of polynomial distributions by infinitely divisible laws [Russisch]. Teor. Verojatn. Primen. **5**, 114 – 124 (1960) und englische Übersetzung in: Theor. Probab. Appl. **5**, 106 – 114 (1960).

Molenaar, W.: How to poison Poisson (when approximating binomial tails). Statistica Neerlandica **23**, 19 – 40 (1969).

Peizer, D. B., Pratt, J. W.: A normal approximation for binomial, F, Beta, and other common, related tail probabilities, I. J. Amer. statist. Assoc. **63**, 1416 – 1456 (1968).

Pratt, J. W.: A normal approximation for binomial, F, beta, and other common related tail probabilities, II. J. Amer. statist. Assoc. **63**, 1457 – 1483 (1968).

Prohorov, Yu. V.: Asymptotic behavior of the binomial distribution [Russisch]. Uspehi mat. Nauk **8**, 135 – 142 (1953).

Raff, M. S.: On approximating the point binomial. J. Amer. statist. Assoc. **51**, 293 – 303 (1956).

Rowe, Ch. H.: A proof of the asymptotic series for $\log \Gamma(z)$ and $\log \Gamma(z+a)$. Ann. of Math., II. Ser. **32**, 10 – 16 (1931).

Stange, K.: Die zeichnerische Ermittlung von Plänen für Gut-Schlecht-Prüfung in einem geeigneten Funktionsnetz. Qualitätskontrolle **8**, 85 – 91 (1963).

Tsaregradskii, I. P.: On the uniform approximation of the binomial distribution [Russisch]. Teor. Verojatn. Primen. **3**, 470 – 474 (1958).

Whittaker, E. T., Watson, G. N.: A course of modern analysis. 4. Aufl. Cambridge Univ. Press 1927 (Nachdruck 1952).

Professor Dr. Rudolf Borges
Department of Mathematical Sciences
New Mexico State University
Las Cruces, New Mexico 88001, USA

(Eingegangen am 11.12.1967, in revidierter Form am 15.8.1969)