EINIGE ZWEIDIMENSIONALE VERTEILUNGS- UND GRENZVERTEILUNGSSÄTZE IN DER THEORIE DER GEORDNETEN STICHPROBEN

ISTVÁN VINCZE

Einleitung

Im Folgenden beweisen wir einige Verteilungs- und Grenzverteilungssätze, die sich auf zwei Funktionen der Elemente zweier Stichproben beziehen. Diese zweidimensionale Verteilungssätze bieten die Möglichkeit, neue Kriterien zur Entscheidung der Frage, ob zwei Stichproben aus derselben Verteilung stammen oder nicht, zu konstruieren. Solche Kriterien kann man als Verfeinerung gewisser, nur auf einer Stichprobenfunktion beruhender Tests auffassen, undzwar im folgenden Sinne: Es seien D und E zwei verschiedene Stichprobenfunktionen der Elemente zweier Stichproben, die aus zwei zufälligen Variablen ξ bzw. η mit den Verteilungsfunktionen $F(x)$, bzw. $G(x)$ stammen. Wenn eine statistische Probe nur auf der Stichprobenfunktion D beruht, so bekommen wir als beste kritische Region für eine gewisse Alternativhypothese gegen die Nullhypothese $F(x) \equiv G(x)$ „im eindimensionalen D-Raum" eine Punktmenge, welche wir einfachkeitshalber als ein Intervall $D_1 < D < D_2$ annehmen können. Diese Region ist in der (D, E)-Ebene ein Streifen $D_1 < D < D_2, -\infty < E < \infty$. Wenn wir nun eine Probe mit Hilfe der zweidimensionalen Verteilung $H(x, y) = \mathbf{P}\{D < x, E < y\}$ konstruieren, so können wir kaum erwarten — wenigstens nicht aus heuristichen Gründen —, dass wir zu unserer Gegenhypothese denselben Streifen als beste kritische Region erhalten, wenn nur das Stichprobenfunktionspaar (D, E) in gewissem Sinne mehr aussagt, als D allein. — Auf die Fragen der auf unsere Verteilungssätze beruhenden statistischen Proben wollen wir in einer weiteren Mitteilung zurückkehren; wir möchten die zur Anwendung gewisser solcher Proben nötigen Tabellen auch bei dieser Gelegenheit angeben.

Unsere Sätze schliessen sich eng den bekannten Kriterien von KOLMOGOROW, SMIRNOW und RÉNYI an. Eben deshalb leiten wir unsere Resultate mit der Analyse dieser Kriterien ein.

Weiterhin wünschen wir uns auch in der späteren Arbeit mit einigen hier nicht behandelten Fragen des Zusammenhanges unserer Verteilungssätze mit den bekannten Versteilungssätzen zu befassen.

1. §. Die Tests von Kolmogorow, Smirnow und Rényi

Es seien $\xi_1, \xi_2, \ldots, \xi_n$ die Ergebnisse unabhängiger Messungen bezüglich einer zufälligen Variablen ξ mit der stetigen Verteilungsfunktion $F(x)$; d. h. unabhängige zufällige Variablen von der gleichen Verteilung. Es seien weiterhin $\xi_1^* < \xi_2^* < \cdots < \xi_n^*$ die Elemente dieser Stichprobe, der Grösse nach geordnet. Es bezeichne $F_n(x)$ die Verteilungsfunktion der Stichprobe, d. h.

$$F_n(x) = \begin{cases} 0, & \text{wenn } x \leqq \xi_1^*, \\ \dfrac{k}{n}, & \text{wenn } \xi_k^* < x \leqq \xi_{k+1}^*, \; k=1,2,\ldots n-1, \\ 1, & \text{wenn } \xi_n^* < x. \end{cases}$$

N. V. Smirnow [14] bestimmte die Grenzverteilung von

$$\sqrt{n} \sup_{-\infty<x<\infty} (F_n(x) - F(x)),$$

wenn $n \to \infty$, A. N. Kolmogorow [8] die Grenzverteilung von

$$\sqrt{n} \sup_{-\infty<x<\infty} |F_n(x) - F(x)|.$$

Beide Grenzverteilungen sind unabhängig von der Verteilungsfunktion $F(x)$ und in Kenntnis dieser kann man bei gegebener, genügend grosser Stichprobe erwägen, ob die maximale Abweichung der empirischen Verteilungsfunktion von der theoretischen Verteilung, bzw. der Maximalwert der absoluten Abweichung noch erlaubt ist, nicht im Gegensatz zur Hypothese, dass die Stichprobenelemente $\xi_1, \xi_2, \ldots, \xi_n$ von der gleichen Verteilung mit der theoretischen Verteilungsfunktion $F(x)$ sind, steht.

Es seien nun $\eta_1, \eta_2, \ldots, \eta_n$ unabhängige Stichprobenelemente der zufälligen Variablen η mit der stetigen Verteilungsfunktion $G(x)$, und es sei $G_n(x)$ die empirische Verteilungsfunktion der Stichprobe. Smirnow hat die Grenzverteilung von

$$\sqrt{\frac{n}{2}} \sup_{-\infty<x<\infty} (F_n(x) - G_n(x)) \quad \text{und} \quad \sqrt{\frac{n}{2}} \sup_{-\infty<x<\infty} |F(x) - G_n(x)|$$

unter der Hypothese $F(x) \equiv G(x)$ bestimmt.[1]

Die Kenntnis dieser, von der Verteilungsfunktion $F(x)$ unabhängigen Grenzverteilungen ermöglicht die statistische Kontrolle der Hypothese $F(x) \equiv G(x)$ auf Grund von zwei Stichproben.

A. Rényi [11], [12] hat darauf hingewiesen, dass man obige Abweichungen bei kleinen Werten der Verteilungsfunktion $F(x)$ anders erwägen muss, als bei grösseren Funktionswerten. Deshalb hat er die Untersuchung der *relativen Abweichung* vorgeschlagen und in seiner Arbeit — unter anderem —

[1]) Smirnow hat sich nicht auf den Fall von gleich grossen Stichproben beschränkt.

folgende Grenzwertsätze bewiesen:

$$\lim_{n\to\infty} \mathbf{P}\left\{\sqrt{n} \sup_{x_a<x<\infty} \left(\frac{F_n(x)-F(x)}{F(x)}\right) < y\right\} = \begin{cases} \sqrt{\frac{2}{\pi}} \int\limits_0^{y\sqrt{\frac{a}{1-a}}} e^{-\frac{u^2}{2}} du, & \text{wenn } y > 0, \\ 0 & \text{wenn } y \leqq 0. \end{cases}$$

$$\lim_{n\to\infty} \mathbf{P}\left\{\sqrt{n} \sup_{x_a<x<\infty} \left|\frac{F_n(x)-F(x)}{F(x)}\right| < y\right\} = \begin{cases} \frac{4}{\pi} \sum\limits_{k=0}^{\infty} (-1)^k \frac{e^{-\frac{(2k+1)^2\pi^2(1-a)}{8y^2a}}}{2k+1}, & \text{wenn } y > 0, \\ 0 & \text{wenn } y \leqq 0. \end{cases}$$

Hierbei ist a eine beliebig kleine, positive durch $F(x_a)=a$ definierte Zahl. T. ANDERSON und D. DARLING [1] geben eine Methode zur Bestimmung der Grenzverteilung des Ausdruckes

$$\frac{F_n(x)-F(x)}{\sqrt{F(x)(1-F(x))}};$$

der auf diesem beruhende Test würde beide Enden der empirischen Verteilungsfunktion in gleichem Masse in Betracht ziehen. WANG SHOU-YEN [15] hat einen auf zwei Stichproben beruhenden Test entworfen, der — dem Gedanken von A. RÉNYI folgend — auf der relativen Abweichung beruht. In seiner Arbeit beschäftigt er sich jedoch nicht mit der Untersuchung des absoluten Wertes der relativen Abweichung.

Die RÉNYIschen Grenzwahrscheinlichkeiten verschwinden bei $a=0$, d.h. die relative Abweichung kann in diesem Falle mit der Wahrscheinlichkeit 1 sehr gross sein. Also lässt der RÉNYIsche Test notwendigerweise den Anfang der empirischen Verteilung ausser acht. Es ist also nötig, den Anfangspunkt der Beobachtung in Erwägung der Natur des konkreten Problems und der eventuellen Alternativhypothese zu bestimmen.

Ähnliche Schwierigkeiten treten dem Wesen nach bei jedem statistischen Test auf. Eine Möglichkeit der Überbrückung dieser Schwierigkeiten kann in der Konstruktion solcher Tests bestehen, die auf der Untersuchung der gemeinsamen Verteilung zweier oder mehrerer Statistiken beruhen. Im folgenden betrachten wir einerseits die maximale Abweichung bzw. absolute Abweichung der beiden empirischen Verteilungsfunktionen, anderseits die Stelle, wo dieser Wert zum ersten male angenommen wird. Die Tests erlauben am Anfang der Verteilung auch grössere relative Abweichungen, dort nämlich, wo solche mit grossen Wahrscheinlichkeiten auftreten; im weiteren Verlauf ist jedoch die Grösse der zulässigen relativen Abweichung beschränkt.

Wir haben auch die Verteilung des Quotienten der beiden untersuchten zufälligen Variablen bestimmt; dies ermöglicht eine Probe von ähnlichem Typ wie der RÉNYIsche Test, jedoch ohne der Notwendigkeit der Verkürzung.

Als Randverteilungen unserer Verteilungen erhält man im Falle der Grenzverteilungen die KOLMOGOROW—SMIRNOWschen Verteilungen für zwei Stichproben, im Falle endlicher Stichprobengrössen die durch B. V. GNEDENKO und V. S. KOROLJUK [7] gegebenen genauen Verteilungen, für gleich grosse Stichproben.

Es sei noch erwähnt, dass Z. W. BIRNBAUM und R. PYKE [2] bezüglich zweidimensionaler Verteilungen und Grenzverteilungen verwandte Ergebnisse erhalten haben (wovon ich durch 1. briefliche Mitteilung des ersten Verfassers Kenntnis erhalten habe).

2. §. Verteilungs- und Grenzverteilungssätze

Es sein wieder $F_n(x)$ und $G_n(x)$ die emprischen Verteilungsfunktionen von Stichproben mit der Grösse n der zufälligen Variablen ξ, bzw. η, mit den Verteilungsfunktionen $F(x)$, bzw. $G(x)$. Wir untersuchen den Verlauf der Funktion $F_n(x)-G_n(x)$ und bezeichnen im weiteren mit $\xi_0^{(n)}$ diejenige Stelle, an welcher die Differenz der beiden von links stetigen Treppenfunktionen zum erstenmal den Maximum annimt; also ist $\xi_0^{(n)}$ die genaue untere Grenze der Maximumstellen. Eine Ausnahme bildet der Fall, wenn der Maximum 0 beträgt. In diesem Falle wäre — dem Vorausgesagten entsprechend — $\xi_0^{(n)}=-\infty$, bzw. die untere Grenze der Werte der zufälligen Variablen. In solchem Falle lassen wir das erste Intervall, in welchem $F_n(x)-G_n(x)=0$ ausser Betracht und beachten zur Bestimmung von $\xi_0^{(n)}$ nur diejenige x-Stellen, die grösser sind, als eine Stelle x_0, wobei $F_n(x_0)-G_n(x_0)<0$.

Es sei weiterhin $\eta_0^{(n)}$ die genaue untere Grenze der Maximumstellen von $|F_n(x)-G_n(x)|$.

Mit diesen Bezeichnungen können folgende Sätze formuliert werden:

Satz 1. *Ist* $F(x)\equiv G(x)$, *so ist*

$$\mathbf{P}\left\{\sup_{-\infty<x<\infty}(F_n(x)-G_n(x))=\frac{k}{n},\ \tfrac{1}{2}(F_n(\xi_0^{(n)}+0)+G_n(\xi_0^{(n)}+0))=\frac{r}{2n}\right\}=$$

$$=\begin{cases} 0, \textit{ wenn } k<0 \textit{ oder } k+r \textit{ ungerade}, \\[2ex] \dfrac{1}{(r-1)(2n-r+2)}\dfrac{\binom{r}{\frac{r}{2}}\binom{2n-r}{n-\frac{r}{2}}}{\binom{2n}{n}}, \textit{ wenn } k=0,\ r=2,4,\ldots \\[4ex] \dfrac{k(k+1)}{r(2n-r+1)}\dfrac{\binom{r}{\frac{r+k}{2}}\binom{2n-r+1}{n-\frac{r+k}{2}}}{\binom{2n}{n}}, \textit{ wenn } k=1,2,\ldots,n;\ r=k,\ k+2,\ldots,2n-k. \end{cases}$$

Führen wir folgende Bezeichnungen ein:

$$A_r^{(k)}=\sum_{\nu=0}^{\left[\frac{r}{2k}\right]}(-1)^{\nu}\binom{r}{\frac{r+k}{2}+\nu\cdot k}\frac{(2\nu+1)k}{r},$$

$$B_r^{(k)}=\frac{2^r}{k}\sum_{\nu=1}^{k}(-1)^{\nu-1}\left(\cos(2\nu-1)\frac{\pi}{2k}\right)^r\sin(2\nu-1)\frac{\pi}{2k}.$$

Satz 2. *Ist $F(x) \equiv G(x)$, k und $r > 0$, ganze Zahlen, so ist*

$$\mathbf{P}\left\{\sup_{-\infty<x<\infty} |F_n(x)-G_n(x)| = \frac{k}{n},\ \tfrac{1}{2}(F_n(\eta_0^{(n)}+0)+G_n(\eta_0^{(n)}+0)) = \frac{r}{2n}\right\} =$$

$$= \begin{cases} 0, \text{ wenn } k<0 \text{ oder } r+k \text{ ungerade}, \\ \dfrac{2A_r^{(k)}A_{2n-r+1}^{(k+1)}}{\binom{2n}{n}} = \dfrac{B_r^{(k)}B_{2n-r+1}^{(k+1)}}{2\binom{2n}{n}}, \text{ wenn } k=1,2,\ldots,n;\ r=k,k+2,\ldots,2n-k \end{cases}$$

Hierbei kommt $k=0$ nicht in Frage.

In Betracht dessen, dass die Randverteilungen dieser Verteilungen durch GNEDENKO und KOROLJUK angegebene Verteilungen sind, erhalten wir die wahrscheinlichkeitstheoretischen Beweise folgender Identitäten:

$$\sum_{(r)}^{*} \frac{k(k+1)}{r(2n-r+1)} \binom{r}{\frac{r+k}{2}} \binom{2n-r+1}{n-\frac{r+k}{2}} = \frac{2k+1}{n+k+1}\binom{2n}{n-k},$$

wobei die Summierung auf $r=k, k+2, \ldots, 2n-k$ erfolgt, während $1 \leqq k \leqq n$.

Für $k=0$ ist

$$\sum_{(r)}^{*} \frac{1}{(r-1)(2n-r+2)} \binom{r}{\frac{r}{2}} \binom{2n-r}{n-\frac{r}{2}} = \frac{1}{n+1}\binom{2n}{n},$$

wobei die Summierung auf $r=2,4,6,\ldots$ erfolgt. Eine ähnliche, jedoch kompliziertere Identität kann der absoluten Abweichung entsprechend aufgeschrieben werden.

Von obigen, sich auf endliche Stichproben beziehenden Sätzen erhalten wir die Grenzverteilungssätze, indem im Falle $n \to \infty$ r in n-ter Ordnung und k in $\sqrt{n}$-ter Ordnung nach Unendlich hält; genauer

$$k \sim \sqrt{2n}\, y, \quad r \sim 2nz.$$

Satz 3. *Ist $F(x) \equiv G(x)$, so ist für $0 \leqq z \leqq 1$:*

$$\lim_{n\to\infty} \mathbf{P}\left\{\sqrt{\frac{n}{2}} \sup_{-\infty<x<\infty} (F_n(x)-G_n(x)) < y,\ \tfrac{1}{2}(F_n(\xi_0^{(n)}+0)+G_n(\xi_0^{(n)}+0)) < z\right\} =$$

$$= \begin{cases} \sqrt{\dfrac{2}{\pi}} \displaystyle\int_0^y \int_0^z \frac{u^2}{(v(1-v))^{3/2}} e^{-\frac{1}{2}\frac{u^2}{v(1-v)}}\, du\, dv \quad \text{für } y>0 \\ 0 \quad \text{für } y \leqq 0. \end{cases}$$

Aus der hier gewonnenen Verteilung erhält man durch Integration nach z von 0 bis 1 die eine SMIRNOWsche Verteilung:

$$\lim_{n\to\infty} \mathbf{P}\left(\sqrt{\frac{n}{2}} \sup_{-\infty<x<\infty} (F_n(x)-G_n(x)) < y\right) = \begin{cases} 1-e^{-2y^2}, & \text{für } y>0 \\ 0\ , & \text{für } y \leqq 0. \end{cases}$$

Durch Integration nach y von 0 bis ∞ erhält man die Grenzverteilung der zufälligen Variablen $\frac{1}{2}(F_n(\xi_0^{(n)}+0)+G_n(\xi_0^{(n)}+0))$:

$$\sqrt{\frac{2}{\pi}}\int_0^z \frac{1}{(v(1-v))^{3/2}}\left(\int_0^\infty u^2 e^{-\frac{1}{2}\frac{u^2}{v(1-v)}}\,du\right)dv = z,$$

d. h., da $F(x)\equiv G(x)$, folgt aus Satz 3. folgendes

Korollar. *Ist* $F(x)\equiv G(x)$, *so ist für* $0\leqq z\leqq 1$, $\mathbf{P}\{F(\xi_0^{(n)}+0)<z\}\to z$, d. h. die Grenzverteilung von $\xi_0^{(n)}$ stimmt mit der gemeinsamen Verteilung von ξ und η überein.

Führen wir folgende Bezeichnung ein:

$$f(y,z)=\frac{y}{z^{3/2}}\sum_{\nu=0}^{\infty}(-1)^\nu(2\nu+1)e^{-\frac{1}{2}(2\nu+1)^2\frac{y^2}{z}}.$$

Satz 4. *Ist* $F(x)\equiv G(x)$, *so ist für* $0\leqq z\leqq 1$,

$$\lim_{n\to\infty}\mathbf{P}\left\{\sqrt{\frac{n}{2}}\sup_{-\infty<x<\infty}|F_n(x)-G_n(x)|<y,\ \frac{1}{2}(F_n(\eta_0^{(n)}+0)+G_n(\eta_0^{(n)}+0))<z\right\}=$$

$$=\begin{cases}\sqrt{\frac{8}{\pi}}\int_0^y\int_0^z f(u,v)f(u,1-v)\,du\,dv, & \text{für}\quad y>0\\ 0, \quad \text{für}\quad y\leqq 0.\end{cases}$$

Durch Integration nach z von 0 bis 1 erhält man die Verteilung von SMIRNOW bezüglich der absoluten Abweichung:

$$\lim_{n\to\infty}\mathbf{P}\left\{\sqrt{\frac{n}{2}}\sup_{-\infty<x<\infty}|F_n(x)-G_n(x)|<y\right\}=\begin{cases}\sum_{\nu=-\infty}^{\infty}(-1)^\nu e^{-2\nu^2y^2}, & \text{wenn}\quad y>0\\ 0 & \text{, wenn}\quad y\leqq 0.\end{cases}$$

Satz 5. *Ist* $F(x)\equiv G(x)$, *so ist*

$$\lim_{n\to\infty}\mathbf{P}\left\{\sqrt{\frac{n}{2}}\frac{\sup_{-\infty<x<\infty}(F_n(x)-G_n(x))}{\frac{1}{2}(F_n(\xi_0^{(n)}+0)+G_n(\xi_0^{(n)}+0))}<c\right\}=$$

$$=\begin{cases}\frac{3}{\sqrt[4]{8}}\int_0^c\sqrt{t}\,e^{\frac{t^2}{4}}W_{-\frac{7}{4},\frac{1}{4}}\left(\frac{t^2}{2}\right)dt & \text{für}\quad c>0,\\ 0, & \text{für}\quad c\leqq 0,\end{cases}$$

wobei mit W die sogenannte WHITTAKER*sche Funktion bezeichnet wird:*

$$W_{\lambda,\mu}(z)=\frac{z^{\mu+\frac{1}{2}}e^{-\frac{z}{2}}}{\Gamma\left(\mu-\lambda+\frac{1}{2}\right)}\int_0^\infty e^{-zt}\,t^{\mu-\lambda-\frac{1}{2}}(1+t)^{\mu-\lambda-\frac{1}{2}}\,dt,$$

wenn

$$\Re e(\mu-\lambda)>-\frac{1}{2},\quad |\arg z|<\pi.$$

Unsere Grenzverteilungssätze 3. und 4. sind im Einklang mit den bekannten in dieser Richtung fallenden Resultaten. (Siehe [3], [9].) Wir bemerken weiter, dass man die erwähnte Grenzverteilungssätze auch durch die von DOOB [5] und DONSKER [4] ausgearbeitete Methode gewinnen kann. Auf diese Fragen wollen wir noch zurückkehren.

3. §. Beweis der Verteilungssätze

Satz 1. wird mit Hilfe eines anschaulichen kombinatorischen Lemmas bewiesen, während der Beweis von Satz 2. durch die Zurückleitung der Behauptung auf ein Irrfahrtsproblem erfolgt. Die Wahrscheinlichkeiten von Satz 2. werden auch mit Hilfe von Differenzengleichungen berechnet; dies führt zur Herstellung mit trigonometrischen Funktionen. Bei dem entsprechenden Grenzverteilungssatz werden wir uns auf die erste Herstellung berufen.

a) Vereinigen wir die vollkommen unabhängigen zufälligen Variablen $\xi_1, \xi_2, \ldots, \xi_n$ und $\eta_1, \eta_2, \ldots, \eta_n$, die alle dieselbe Verteilung besitzen, zu einer einzigen Folge $\zeta_1, \zeta_2, \ldots, \zeta_{2n}$; letztere ordnen wir der Grösse nach: $\zeta_1^* < \zeta_2^* < \cdots < \zeta_{2n}^*$. Wir interpretieren nun die Variablen $\vartheta_1, \vartheta_1, \ldots, \vartheta_{2n}$ derart, dass

$$\vartheta_i=\begin{cases}+1, & \text{wenn } \zeta_i^*=\xi_j\\ -1, & \text{wenn } \zeta_i^*=\eta_h.\end{cases}$$

Also besteht die Folge $\vartheta_1, \vartheta_2, \ldots, \vartheta_{2n}$ aus n Stück $+1$ und n Stück -1. Aus diesen lassen sich jedoch — den verschiedenen Möglichkeiten, wie man die n Stück $+1$-er in den $2n$ Stellen unterbringen kann, entsprechend — $\binom{2n}{n}$ verschiedene Folgen bilden. Jede dieser Folgen tritt mit der gleichen Wahrscheinlichkeit auf, da jede Folge je eine geordnete Stichprobe der Variablen ξ und η enthält, und die insgesamt möglichen $(n!)^2$ Permutationen der ξ und η untereinander zu den insgesamt möglichen $(2n)!$ ungeordneten Stichproben von der Grösse $2n$ führen, wobei jede Reihenfolge gleich wahrscheinlich ist.

Betrachten wir nun die Summen

$$\varsigma_i=\vartheta_1+\vartheta_2+\cdots+\vartheta_i.$$

Diese Summe gibt den Unterschied der Anzahl derjenigen ξ_j und η_h an, die nicht grösser sind als ζ_i^*. Dies ist jedoch nichts anderes, als $n(F_n(\zeta_i^*+0)-$ $-G_n(\zeta_i^*+0))$. Damit die in §2. mit $\xi_0^{(n)}$ bezeichnete zufällige Variable genau

r betragen soll, d. h. die kleinste ganze Zahl, für welche

$$\sup_{-\infty<x<\infty} n(F_n(x)-G_n(x)) = n(F_n(\zeta_r^*+0)-G_n(\zeta_r^*+0))$$

besteht, sein soll; weiterhin dieser Supremalwert genau k betragen soll, müssen folgende Bedingungen erfüllt werden:

$$\varsigma_i < k \quad \text{für} \quad 1 \leqq i \leqq r-1,$$
$$\varsigma_r = k$$
$$\varsigma_i \leqq k \quad \text{für} \quad r+1 \leqq i \leqq 2n-k.$$

In diesem Falle ist also

$$\mathbf{P}\left\{\sup_{-\infty<x<\infty}(F_n(x)-G_n(x)) = \frac{k}{n},\ \frac{1}{2}(F_n(\xi_0^{(n)}+0)+G_n(\xi_0^{(n)}+0)) = \frac{r}{2n}\right\} =$$
$$(3.1) \qquad = \mathbf{P}\left\{\max_{1\leqq i\leqq r-1}\varsigma_i < k,\ \varsigma_r = k,\ \max_{r+1\leqq i\leqq 2n-k}\varsigma_i \leqq k\right\}.$$

Zur Berechnung dieser Wahrscheinlichkeit müssen von den $\binom{2n}{n}$ möglichen Folgen $\vartheta_1, \vartheta_2, \ldots, \vartheta_{2n}$ zuerst nur diejenigen betrachtet werden, in deren r-tem Abschnitt die Anzahl der $+1$ die Anzahl der -1 genau mit k übertrifft, d. h. in denen $\frac{r+k}{2}$ Stück $+1$ und $\frac{r-k}{2}$ Stück -1 vorkommen. (Also muss $r+k$ gerade sein: $r=k, k+2, \ldots$.) Es gibt $\binom{r}{\frac{r+k}{2}}$ solche Abschnitte.

Alle diese Folgen kommen jedoch nicht in Frage. Es muss nämlich auch gesichert werden, dass in keiner Teilfolge $\vartheta_1, \vartheta_2, \ldots, \vartheta_{r'}$ $(r'<r)$ die Anzahl der $+1$ genau mit k die Anzahl der -1 übertrifft. Dies kann man dadurch erreichen, dass man diejenigen Folgen ausschliesst, die — von hinten an betrachtet — einen Abschnitt, in welchem die Anzahl der $+1$ und der -1 übereinstimmt, besitzen. Für diese gilt folgendes, bekanntes

Lemma 1.[2] *Die Wahrscheinlichkeit dessen, dass eine aus α Stück $+1$ und β Stück -1 $(\alpha>\beta)$ zufälligerweise aufgeschriebene Folge einen Abschnitt, in welchem die Anzahl der $+1$ und der -1 übereinstimmt, enthält, beträgt $\frac{2\beta}{\alpha+\beta}$.*

In unserem Fall handelt es sich um das gegensätzliche Ereignis bei $\alpha=\frac{r+k}{2}$, $\beta=\frac{r-k}{2}$, und so ist die gesuchte Wahrscheinlichkeit $1-\frac{r-k}{r}=\frac{k}{r}$. Also ergibt sich für die Anzahl der geeigneten Folgen des ersten r Abschnittes

$$(3.2) \qquad \binom{r}{\frac{r+k}{2}}\frac{k}{r}.$$

[2]) Siehe z. B. [13], Aufgabe 36., Seite 74; Lösung, Seite 694.

Hiermit ist die Erfüllung der Forderungen $\varsigma_i < k$, $i=1, 2, \ldots, r-1$ und $\varsigma_r = k$ gesichert, es muss noch die Forderung $\varsigma_i \leqq k$, $i=r+1, \ldots, 2n-k$ erfüllt werden. (Die weiteren Summen sind nämlich offensichtlich kleiner als k, da $\varsigma_{2n}=0$ ist.) Dies können wir dadurch erreichen, dass wir für die Folge $\vartheta_{r+1}, \vartheta_{r+2}, \ldots, \vartheta_{2n}$ zwar solche Abschnitte — von vorne betrachtet — in welchen die Anzahl der $+1$ und -1 übereinstimmt, zulassen, jedoch diejenigen Folgen, in derem irgendeinem beliebigen Abschnitt die Anzahl der $+1$ die der -1 übertrifft, ausschliessen. Hierzu gebrauchen wir

Lemma 2.[3] *Die Wahrscheinlichkeit dessen, dass in keinem Abschnitt einer aus α Stück -1 und β Stück $+1$ $(\alpha \geqq \beta)$ zufälligerweise aufgeschriebenen Folge die Anzahl der $+1$ die der -1 übertrifft, beträgt* $\frac{\alpha+1-\beta}{\alpha+1}$.

Für diejenigen Folgen, die $\alpha = n - \frac{r-k}{2}$ Stück -1 und $\beta = n - \frac{r+k}{2}$ Stück $+1$ enthalten, beträgt diese Wahrscheinlichkeit

$$\frac{2(k+1)}{2n-r+k+2}.$$

Da die Anzahl aller möglichen Folgen $\vartheta_{r+1}, \vartheta_{r+2}, \ldots, \vartheta_{2n}$ $\binom{2n-r}{n-\frac{r+k}{2}}$ beträgt, ist die Anzahl der für uns in Frage kommenden Folgen

$$\binom{2n-r}{n-\frac{r+k}{2}} \frac{2(k+1)}{2n-r+k+2} = \binom{2n-r+1}{n-\frac{r+k}{2}} \frac{k+1}{2n-r+1}. \tag{3. 3}$$

Da jeder in (3. 2) betrachteten Folge $\vartheta_1, \vartheta_2, \ldots, \vartheta_r$ jede in (3. 3) betrachtete Folge $\vartheta_{r+1}, \vartheta_{r+2}, \ldots, \vartheta_{2n}$ angeschlossen werden kann, und da die Anzahl der überhaupt möglichen Folgen $\binom{2n}{n}$ beträgt, ergibt sich für die in (3. 1) gesuchte Wahrscheinlichkeit

$$\frac{\binom{r}{\frac{r+k}{2}}\binom{2n-r+1}{n-\frac{r+k}{2}} \frac{k(k+1)}{r(2n-r+1)}}{\binom{2n}{n}}.$$

Es ist leicht einzusehen, dass aus Lemma 1. und 2., mit entsprechender Anwendung des obigen Gedankenganges, unsere Behauptung für den Fall $k=0$ folgt. Dadurch ist Satz 1. bewiesen.

b) Bevor wir zum Beweis von Satz 2. übergehen, machen wir eine vorbereitende Bemerkung. Betrachten wir einen Punkt, der an den ganzzahligen Stellen der Zahlengeraden mit der Wahrscheinlichkeit von je $\frac{1}{2}$ nach rechts

---

[3]) Siehe [13], Aufgabe 37. (wie in [2]).

bzw. nach links herumirrt und von 0 ausgehend, zum ersten Male beim r-ten Schritt die Stelle $+k$ erreicht. Die Zahl aller möglichen verschiedenen Wege dieses herumirrenden Punktes ist nach (3. 2)

$$\binom{r}{\frac{r-k}{2}}\frac{k}{r}.$$

Diese Wege sind nämlich alle derart, dass sie sich nach dem $r-1$-ten Schritt an der Stelle $k-1$ befunden und mit $k-1$ mehr Schritte nach rechts, als nach links, d. h. $\frac{r+k}{2}-1$ Schritte nach rechts, $\frac{r-k}{2}$ Schritte nach links enthalten haben. Die anzahl aller solcher Wege ist

$$\binom{r-1}{\frac{r+k}{2}-1}.$$

Davon muss jedoch die Anzahl derjenigen Wege abgezogen werden, die die Stelle $+k$ schon erreicht oder übertreten haben. Nach dem bekannten Spiegelungsprinzip ist dies gleich der Anzahl der aus 0 ausgehenden, in $r-1$ Schritten genau die Stelle $k+1$ erreichenden Wege. Diese Zahl beträgt, auf ähnlicher Art, wie vorhin,

$$\binom{r-1}{\frac{r+k}{2}}$$

und die gesuchte Zahl — die Anzahl der die Stelle k zum erstenmal in r Schritten erreichenden Wege — ist demnach

$$\binom{r-1}{\frac{r+k}{2}-1}-\binom{r-1}{\frac{r+k}{2}}=\binom{r}{\frac{r+k}{2}}\frac{k}{r}.$$

c) Die Bestimmung der in Satz 2. auftretenden Wahrscheinlichkeiten kann ebenfalls auf die Folgen $\vartheta_1, \vartheta_2, \ldots, \vartheta_{2n}$ und auf den Nachweis der Behauptungen bezüglich der daraus gebildeten ς_i Summen zurückgeführt werden. Dem Gedankengang von Punkt **a)** entsprechend kann nämlich auch die Behauptung

$$\mathbf{P}\left\{\sup_{-\infty<x<\infty}|F_n(x)-G_n(x)|=\frac{k}{n},\ \tfrac{1}{2}(F_n(\eta_0^{(n)}+0)+G_n(\eta_0^{(n)}+0))=\frac{r}{2n}\right\}=$$

$$=\mathbf{P}\left\{\max_{1\leqq i\leqq r-1}|\varsigma_i|<k, |\varsigma_r|=k, \max_{r+1\leqq i\leqq 2n}|\varsigma_i|\leqq k\right\}$$

eingesehen werden.

Dementsprechend muss man aus den $\binom{2n}{n}$ verschiedenen Reihenfolgen der ϑ_i $(i=1,2,\ldots,2n)$ die Anzahl derer feststellen, die den Bedingungen

$$
\begin{aligned}
&|\varsigma_i|<k, && i=1,2,\ldots,r-1,\\
&|\varsigma_r|=k, && \\
&|\varsigma_i|\leqq k, && i=r+1,r+2,\ldots
\end{aligned}
$$

genügen. Dem im Punkt **b)** gesagten gemäss ist dies gleich der Anzahl der α. aus 0 ausgehenden und zum erstenmal beim r-ten Schritt eine der Stellen $+k$ oder $-k$ erreichenden, β. weiterhin die Stellen $+k$ und $-k$ nickt übertretenden und beim $2n$-ten Schritt wieder in die Nullstelle zurückgelangenden Wege.

Wir bezeichnen die Anzahl derjenigen Wege, die α. erfüllen, mit $2A_r^{(k)}$. Dann ist offensichtlich $A_r^{(k)}$ die Anzahl derjenigen Wege, die von 0 ausgehend, zum erstenmal beim r-ten Schritt die Stelle $+k$ erreichend und bis dahin die Stelle $-(k-1)$ nicht übertreten haben. Wir bestimmen die Anzahl dieser Wege mit Hilfe des Spiegelungsprinzipes.

Die Anzahl derjenigen Wege, die von 0 ausgehend, zum erstenmal beim r-ten Schritt die Stelle k erreichen, ist — laut **b)** —

$$\binom{r}{\frac{r+k}{2}}\frac{k}{r}. \tag{3.2}$$

Dies enthält jedoch auch jene Wege, die die Stelle $-k$ erreicht oder übertreten haben. Spiegeln wir den von 0 bis $-k$ reichenden Teil dieser Wege über den Punkt $-k$. Damit gehen diese Wege in solche über, die von $-2k$ ausgehend, zum erstenmal im r-ten Schritt die Stelle $+k$ erreichen. Die Zahl aller solcher Wege beträgt — in (3. 2) an stelle von k $3k$ setzend,

$$\binom{r}{\frac{r+3k}{2}}\frac{3k}{r}. \tag{3.4}$$

Wenn wir (3. 4) von (3. 2) subtrahieren, so bringen wir damit auch solche Wege zum Abzug, die von $-2k$ ausgehend, die Stelle $-3k$ erreicht haben, also vor der Spiegelung die Stelle $+k$ übertreten haben, sind so in (3. 2) nicht enthalten. Die Zahl dieser Wege muss also zur Differenz von (3. 2) und (3. 4) addiert werden. Diese Wege gehen durch Spiegelung der ersten — von $-2k$ bis $-3k$ reichenden — Strecke über die Stelle $-3k$ in solche Wege über, die von $-4k$ ausgehend, zum erstenmal beim r-ten Schritt die Stelle $+k$ erreichen. Die Zahl dieser ist — in (3. 2) anstatt k nun $5k$ setzend —

$$\binom{r}{\frac{r+5k}{2}}\frac{5k}{r}.$$

Dieses Verfahren kann man so lange fortsetzen, bis die Stelle $+k$ in r Schritten von der entsprechenden Stelle $-2\nu k$ erreichbar ist, jedoch von der

Stelle $-(2\nu+1)k$ nicht mehr. Die Bedingung dessen ist, dass

$$2\nu k \leqq r < (2\nu+1)k,$$

d. h.

$$\nu = \left[\frac{r}{2k}\right].$$

Durch dieses Verfahren erhält man für $A_r^{(k)}$ den folgenden Ausdruck:

$$A_r^{(k)} = \sum_{\nu=0}^{\left[\frac{r}{2k}\right]} (-1)^\nu \binom{r}{\frac{r+k}{2}+\nu k} \frac{(2\nu+1)k}{r}.$$

Betrachten wir nun die in β. angegebenen Wege in entgegengesetzter Richtung. Von der Nullstelle ausgehend, sollen diese in $2n-r$ Schritten $+k$ oder $-k$ erreichen und in der Zwischenzeit diese Stellen nicht übertreten haben. Die Anzahl solcher Wege ist jedoch gleich der Zahl der von 0 ausgehenden, in $2n-r+1$ Schritten zum erstenmal die Stellen $(k+1)$ oder $-(k+1)$ erreichenden Wege. Diese Zahl beträgt also $2A_{2n-r+1}^{(k+1)}$. Damit diese Wege sich den vorigen an den Stellen $+k$ oder $-k$ anschliessen, erhält man als die Anzahl der geeigneten Wege die Hälfte des Produktes der beiden Zahlen, also $2A_r^{(k)} A_{2n-r+1}^{(k+1)}$, womit die erste Behauptung von Satz 2. bewiesen ist.

d) Zur trigonometrischen Herstellung[4] der in Satz 2. auftretenden Wahrscheinlichkeiten verfassen wir ein etwas allgemeineres Irrfahrtsproblem. Wir suchen die Wahrscheinlichkeit dessen, dass ein aus der Stelle z ausgehender Punkt ($-k<z<k$, z ganzzahlig) zum erstenmal beim r-ten Schritt eine der Stellen $+k$ oder $-k$ erreicht. Der irrfahrende Punkt tritt von jeder ganzzahligen Stelle mit den Wahrscheinlichkeiten von je $\frac{1}{2}$ an diebei den benachbarten ganzzahligen Stellen über. Später werden wir an Stelle von z 0 setzen, um auf die benötigten Wahrscheinlichkeiten zu schliessen.

Es bezeichne $p_{z,r}^{(k)} = p_{z,r}$ die gesagte Wahrscheinlichkeit. Hierfür ergibt sich die folgende Differenzengleichung:

$$p_{z,r} = \frac{1}{2} p_{z-1,r-1} + \frac{1}{2} p_{z+1,r-1}.$$

Die Anfangsbedingungen sind:

$$p_{k,0} = p_{-k,0} = 1,$$

$$p_{k,r} = p_{-k,r} = 0, \quad r \geqq 1,$$

$$p_{z,0} = 0, \quad -(k-1) \leqq z \leqq k-1.$$

Führen wir die Generatorenfunktion der Wahrscheinlichkeiten $p_{z,r}$ ein:

$$G_z(u) = \sum_{r=0}^{\infty} p_{z,r} u^r \qquad -k<z<k,$$

dabei ist

(3.5) $$G_k(u) = G_{-k}(u) = 1.$$

[4]) Siehe auch [6] S. 292.

Aus der obigen Differenzengleichung ergibt sich — mit Beachtung der Anfangsbedingungen — für die Generatorenfunktion die Differenzengleichung

$$G_z(u) = \tfrac{1}{2}\, u\, G_{z-1}(u) + \tfrac{1}{2}\, u\, G_{z+1}(u),$$

deren Lösung wir aus partikulären Lösungen von der Form

$$G_z(u) = (\varphi(u))^z$$

herstellen. Für $\varphi(u)$ gilt die Gleichung

$$u\varphi(u)^2 - 2\varphi(u) + u = 0,$$

woraus

$$\varphi_1(u) = \frac{1}{u} + \sqrt{\frac{1}{u^2} - 1}$$

$$\varphi_2(u) = \frac{1}{u} - \sqrt{\frac{1}{u^2} - 1} = \frac{1}{\varphi_1(u)}.$$

Es sei nun

$$G_z(u) = A(\varphi_1(u))^z + B(\varphi_2(u))^z,$$

wobei die Parameter A und B derart gewählt werden, dass $G_z(u)$ die Anfangsbedingungen (3. 5) erfüllt:

$$A(\varphi_1(u))^k + B(\varphi_2(u))^k = 1,$$

$$A(\varphi_1(u))^{-k} + B(\varphi_2(u))^{-k} = 1.$$

Daraus ergibt sich

$$A = B = \frac{1}{(\varphi_1(u))^k + (\varphi_2(u))^k}$$

und

$$G_z(u) = \frac{(\varphi_1(u))^z + (\varphi_2(u))^z}{(\varphi_1(u))^k + (\varphi_2(u))^k},$$

d. h. für $z = 0$

$$G_0(u) = \frac{2}{(\varphi_1(u))^k + (\varphi_2(u))^k}.$$

Es ist jedoch bekannt, dass für das k-te TSCHEBISEFFsche Polynom folgende Herstellung gilt:

$$2T_k(x) = (x + \sqrt{x^2 - 1})^k + (x - \sqrt{x^2 - 1})^k = 2^k x^k + \cdots,$$

woraus sich

$$G_0(u) = \frac{1}{T_k\left(\dfrac{1}{u}\right)} = \frac{u^k}{2^{k-1} \prod\limits_{\nu=1}^{k} \left(1 - u\cos(2\nu - 1)\dfrac{\pi}{2k}\right)}$$

ergibt. Die rechte Seite in Partialbrüche zerlegt, erhalten wir die Herstellung

$$G_0(u)=\frac{u^k}{2^{k-1}}\sum_{\nu=1}^{k}\frac{A_\nu}{1-u\cos(2\nu-1)\frac{\pi}{2k}},$$

wobei

$$A_\nu=\frac{\cos^k(2\nu-1)\frac{\pi}{2k}}{\prod_{\nu'=1}^{k}\left(\cos(2\nu-1)\frac{\pi}{2k}-\cos(2\nu'-1)\frac{\pi}{2k}\right)}=\frac{2^{k-1}\cos^k(2\nu-1)\frac{\pi}{2k}}{T_k'\left(\cos(2\nu-1)\frac{\pi}{2k}\right)}.$$

Es ist jedoch bekannt,[5] dass

$$\frac{1}{k}T_k'(\cos\psi)=\frac{\sin k\psi}{\sin\psi}$$

und so

$$A_\nu=\frac{2^{k-1}\cos^k(2\nu-1)\frac{\pi}{2k}\sin(2\nu-1)\frac{\pi}{2k}}{k\sin(2\nu-1)\frac{\pi}{2}}.$$

Dabei ist

$$\sin(2\nu-1)\frac{\pi}{2}=(-1)^{\nu'}$$

und deshalb erhalten wir den Ausdruck

$$G_0(u)=\frac{u^k}{k}\sum_{\nu=1}^{k}(-1)^{\nu-1}\frac{\cos^k(2\nu-1)\frac{\pi}{2k}\sin(2\nu-1)\frac{\pi}{2k}}{1-u\cos(2\nu-1)\frac{\pi}{2k}}.$$

Durch Entwicklung in Potenzreihe nach u erhält man als den Koeffizienten von u^r die gesuchte Wahrscheinlichkeit

$$p_{0,r}^{(k)}=\frac{1}{k}\sum_{\nu=1}^{k}(-1)^{\nu-1}\cos^r(2\nu-1)\frac{\pi}{2k}\sin(2\nu-1)\frac{\pi}{2k}\qquad (k+r \text{ gerade}).$$

Aus dieser Wahrscheinlichkeit bestimmen wir nun die Anzahl der den Bedingungen

$$|\varsigma_i|<k,\qquad i=1,2,\ldots,r-1,$$
$$|\varsigma_r|=k$$

entsprechenden Wege. Die zur Wahrscheinlichkeit $p_{0,r}^{(k)}$ gehörende Anzahl aller möglichen verschiedenen Wege beträgt 2^r, da aus dem Nullpunkt ausgehend insgesamt so viele verschiedene Wege in r Schritten gemacht werden können. Deshalb ist die Zahl der günstigen — die Bedingungen erfüllenden — Wege

$$B_r^{(k)}=2^r p_{0,r}^{(k)}.$$

[5]) Siehe z. B. [10], Band II. S. 75.

Man kann auf ähnlicher Weise, wie in Punkt **b.**, einsehen, dass die Anzahl der aus $+k$ oder $-k$ zu 0 führenden verschiedenen solcher Wege, die $+k$ oder $-k$ nicht überschreiten,

$$B_{2n-r+1}^{(k+1)} = 2^{2n-r+1} p_{0,2n-r+1}^{(k+1)}$$

beträgt. Ebenso kann man nach dem Gedankengang von Punkt **b.** einsehen, dass aus den überhaupt möglichen $\binom{2n}{n}$ verschiedenen Stichproben von der Grösse $2n$ die Anzahl derer, die den Bedingungen

$$|\varsigma_i| < k, \qquad i = 1, 2, \ldots, r-1,$$
$$|\varsigma_r| = k,$$
$$|\varsigma_i| \leqq k, \qquad i = r+1, r+2, \ldots, 2n-k$$

entsprechen,

$$\frac{1}{2} B_r^{(k)} B_{2n-r+1}^{(k+1)}$$

ist, d. h. wir gelangen zum Ergebnis

$$\frac{B_r^{(k)} B_{2n-r+1}^{(k+1)}}{2\binom{2n}{n}},$$

was die Behauptung von Satz 2. war.

4. §. Beweis der Grenzverteilungssätze

Bezeichnen wir die in den endlichen Verteilungssätzen auftretenden Wahrscheinlichkeiten einheitlich mit $P_{r,k}^{(n)}$ und definieren wir das zufällige Variablenpaar $(\varkappa^{(n)}, \varrho^{(n)})$, welches auf jedem Rechteck

$$\frac{k}{2\sqrt{2n}} \leqq \varkappa^{(n)} < \frac{k+1}{2\sqrt{2n}}, \qquad \frac{r}{2n} \leqq \varrho^{(n)} < \frac{r+2}{2n},$$
$$(k = 0, 1, 2, \ldots, n;\ r = k, k+2, \ldots, 2n-k)$$

gleichverteilt, mit der Dichte $2\sqrt{2}\, n^{3/2} P_{k,r}^{(n)}$, anderswo mit der Dichte 0 ist. Wir beweisen, dass die Folge der Dichtefunktionen in jedem Punkt des unendlichen Streifens $0 \leqq \varkappa \leqq \infty$, $0 \leqq \varrho \leqq 1$ der Ebene $(\varkappa, \varrho)$ einer solchen integrierbaren Dichtefunktion zustrebt, welche in jedem abgeschlossenen Bereich, das die punkte $(0, 0)$ und $(1, 0)$ nicht enthält, beschränkt ist. Bleiben nämlich k und r beschränkt, so wächst die Dichte $n^{3/2} P_{k,r}^{(n)}$ in der Ordnung von $\sqrt{n}$. Schliessen wir also eine beliebig kleine, kreisviertelförmige Umgebung der obigen Punkte aus. Wir können den Satz von LEBESGUE auf jedem beschränkten Teil des Streifens, der diese Umgebung nicht enthält, anwenden. Dieser sagt aus, dass wenn eine Folge nichtnegativer, integrierbarer Funktionen einer beschränkten, integrierbaren Funktion zustrebt, so strebt die Folge der Integrale dem Integral der Grenzfunktion zu. Die Wahrscheinlichkeit des ausgeschlossenen Teiles kann jedoch bei der Grenzfunktion, die auch in dieser Umgebung integrierbar ist, beliebig klein gemacht werden; ebenso können die Integrale der Glieder der Folge beliebig verkleinert werden, also werden unsere Grenzverteilungssätze — im Grenzübergang — auch dann gültig sein, wenn wir die Punkte nicht ausschliessen. Wir wünschen zu bemerken, dass es genügen würde, eine solche Umgebung der genannten Punkte auszuschliessen, die durch die Geraden $\varrho=0$, bzw. $\varrho=c\varkappa$ und $\varrho=0$, bzw. $\varrho=1-c\varkappa$ ($c>0$, beliebig klein) weiterhin durch beliebig kleine Kreisbögen um die Punkte $(0,0)$, bzw. $(1,0)$ gebildet werden. Letztere Bemerkung ist besonders für Satz 5. von Interesse, da hier das ersterwähnte kritische Winkelbereich vom Integrationsbereich von vornherein, ausgeschlossen ist.

Beim Beweis der Grenzverteilungssätze machen wir von folgenden einfachen, bekannten Zusammenhängen gebrauch:

$$\binom{2N}{N}\sim\frac{2^{2N+\frac{1}{2}}}{\sqrt{2\pi N}}. \tag{4.1}$$

Ist $c=\{u\sqrt{2N}\}$, wobei $\{a\}$ diejenige kleinste ganze Zahl bezeichnet, für welche $\geqq a$ gilt, so ist

$$\lim_{N\to\infty}\frac{\binom{2N}{N+c}}{\binom{2N}{N}}=e^{-2u^2}. \tag{4.2}$$

A) Beim Beweis von Satz 3. gehen wir von der Form

$$P_{r,k}^{(n)}=\frac{2k(k+1)}{r(2n-r+k+2)}\,\frac{\binom{r}{\frac{r}{2}+\frac{k}{2}}\binom{2n-r}{\frac{2n-r}{2}+\frac{k}{2}}}{\binom{2n}{n}}$$

aus, welche eine etwas abgeänderte Form der im Satz auftretenden, für $k=1,2,\ldots,n$ gültigen Wahrscheinlichkeiten ist. Stellen wir diese Wahrscheinlichkeit in folgender Form her:

$$P_{r,k}^{(n)}=\frac{2k(k+1)}{r(2n-r+k+2)}\,\frac{\binom{r}{\frac{r}{2}+\frac{k}{2}}}{\binom{r}{\frac{r}{2}}}\,\frac{\binom{2n-r}{\frac{2n-r}{2}+\frac{k}{2}}}{\binom{2n-r}{\frac{2n-r}{2}}}\,\frac{\binom{r}{\frac{r}{k}}\binom{2n-r}{\frac{2n-r}{2}}}{\binom{2n}{n}}.$$

Es sei nun

$$y=\frac{k}{2\sqrt{r}}\quad\text{und}\quad z=\frac{r}{2n},$$

dann ist

$$\frac{k}{2} = \bar{y}\sqrt{r} = \bar{y}\sqrt{\frac{z}{1-z}}\sqrt{2n-r},$$

$$dz = \frac{1}{n} \quad (dr = 2(!)), \quad 2n-r = 2n(1-z)$$

$$\bar{y}\sqrt{z} = \frac{k}{2\sqrt{2n}}, \quad d(\bar{y}\sqrt{z}) = \frac{1}{2\sqrt{2n}}.$$

Demgemäss ist nach (4. 1)

$$\frac{\binom{r}{\frac{r}{2}+\frac{k}{2}}}{\binom{r}{\frac{r}{2}}} \to e^{-2\bar{y}^2},$$

$$\frac{\binom{2n-r}{\frac{2n-r}{2}-\frac{k}{2}}}{\binom{2n-r}{\frac{2n-r}{2}}} \to e^{-2\bar{y}^2\frac{z}{1-z}}$$

und nach (4. 2)

$$2\sqrt{2n}\,\frac{\binom{r}{\frac{r}{2}}\binom{2n-r}{\frac{2n-r}{2}}}{\binom{2n}{n}} \sim 2\sqrt{\frac{2}{\pi}}\sqrt{\frac{4n^2}{r(2n-r)}} = 2\sqrt{\frac{2}{\pi}}\sqrt{\frac{1}{z(1-z)}},$$

weiterhin

$$n\,\frac{2k(k+1)}{r(2n-r+k+2)} \sim n\,\frac{2k^2}{r(2n-r)} = 4\bar{y}^2\frac{1}{1-z}.$$

Also ist

$$\lim_{n\to\infty} 2\sqrt{2}\,n^{3/2}P_{k,r}^{(n)} = \frac{16}{\sqrt{2\pi}}e^{-2\bar{y}^2\frac{1}{1-z}}\cdot\frac{\bar{y}^2}{\sqrt{z}\,\sqrt{(1-z)^3}}.$$

Mit Anwendung der Substitution $y = 2\bar{y}\sqrt{z}$ gelangen wir zur Dichtefunktion

$$\sqrt{\frac{2}{\pi}}\,\frac{y^2}{(z(1-z))^{3/2}}\,e^{-\frac{1}{2}\frac{y^2}{z(1-z)}}$$

und hiermit zum Beweis von Satz 3.

B) Führen wir die Bezeichnung

$$F(y, z)=\sqrt{\frac{2}{\pi}}\int_0^y\int_0^z \frac{u^2}{(v(1-v))^{3/2}}\, e^{-\frac{1}{2}\frac{u^2}{v(1-v)}}\, du\, dv$$

ein. Dann muss $F(y, 1)$ die SMIRNOWsche Verteilung ergeben. Wahrlich, wegen der Symmetrie in $v=\frac{1}{2}$ und mit der Substitution von

$$\frac{1}{v(1-v)}=t^2+4, \qquad dt=-\frac{2}{(v^2+4)^{3/2}}\,dv,$$

ergibt sich für

$$F(y, 1)=2\sqrt{\frac{2}{\pi}}\int_0^y u^2\left(\int_0^{1/2}\frac{1}{(v(1-v))^{3/2}}\, e^{-\frac{1}{2}\frac{u^2}{v(1-v)}}\, dv\right) du=$$

$$=4\sqrt{\frac{2}{\pi}}\int_0^y u^2 e^{-2u^2}\left(\int_0^\infty e^{-\frac{1}{2}u^2t^2}\, dt\right) du=4\int_0^y u e^{-2u^2}\, du=1-e^{-2y^2}.$$

C) Berechnen wir $G(z)=F(\infty, z)$.

$$G(z)=\sqrt{\frac{2}{\pi}}\int_0^z\frac{1}{(v(1-v))^{3/2}}\left(\int_0^\infty u^2 e^{-\frac{1}{2}\frac{u^2}{v(1-v)}}\, du\right) dv=z,$$

wobei wir bei der Berechnung des Integrals nach u von der Substitution

$$u^2(v(1-v))=w^2$$

gebrauch machen.

D) Der Beweis des Grenzverteilungssatzes 4. kann in analoger Weise, wie der Beweis von Satz 3., erfolgen. Da die in Satz 2. auftretenden Reihen absolut konvergent sind, kann der Grenzübergang des Produktes der beiden Reihen gliedweise durchgeführt werden. Hierbei muss man für jedes fixes ν von dem Zusammenhang

$$\frac{\binom{r}{\frac{r}{2}+\frac{(2\nu+1)k}{2}}}{\binom{r}{\frac{r}{2}}}\to e^{-2(2\nu+1)^2 y^2}$$

gebrauch machen für $r\to\infty$, $k=2\{\sqrt{r}\, y\}$.

E) Nun wollen wir zeigen, dass die Randverteilung der in Satz 4. auftretenden Verteilung mit der SMIRNOWschen Verteilung der absoluten Abweichung übereinstimmt. Deshalb integrieren wir die Dichtefunktion nach z von 0 bis 1:

$$\sqrt{\frac{8}{\pi}}\,y^2\int_0^1 \frac{1}{(z(1-z))^{3/2}}\sum_{\nu=0}^{\infty}(-1)^{\nu}(2\nu+1)e^{-\frac{1}{2}(2\nu+1)^2\frac{y^2}{z}}\sum_{\mu=0}^{\infty}(-1)^{\mu}(2\mu+1)e^{-\frac{1}{2}(2\mu+1)^2\frac{y^2}{1-z}}\,dz.$$

Wir ordnen das Produkt der beiden Reihen nach den Gliedern $\mu+\nu+1=$ $=$ const. und integrieren gliedweise. Das allgemeine Glied des Produktes ist

$$(4.3)\qquad \sqrt{\frac{8}{\pi}}(-1)^{\nu+\mu}(2\nu+1)(2\mu+1)y^2\int_0^1\frac{1}{(z(1-z))^{3/2}}e^{-\frac{1}{2}(2\nu+1)^2\frac{y^2}{z}-\frac{1}{2}(2\mu+1)^2\frac{y^2}{1-z}}\,dz.$$

Führen wir die Bezeichnung $A_s=\frac{1}{2}(2s+1)^2y^2$ ein und wenden wir die Transformation

$$\frac{1}{z}=u^2+1$$

an; dann ist

$$\frac{1}{1-z}=\frac{u^2+1}{u^2},\qquad dz=-\frac{2u\,du}{(u^2+1)^2}$$

und unser Integral geht über in

$$8\sqrt{\frac{2}{\pi}}(-1)^{\nu+\mu}\sqrt{A_\nu A_\mu}\,e^{-A_\nu-A_\mu}\int_0^{\infty}\left(1+\frac{1}{u^2}\right)e^{-A_\nu u^2-\frac{A_\mu}{u^2}}\,du.$$

Den letzten Teil dieses Ausdruckes stellen wir folgendermassen her;

$$\int_0^{\infty}\left(1+\frac{1}{u^2}\right)e^{-A_\nu u^2-\frac{A_\mu}{u^2}}\,du=\left(1+\sqrt{\frac{A_\nu}{A_\mu}}\right)\int_0^{\infty}e^{-A_\nu u^2-\frac{A_\mu}{u^2}}\,du-$$

$$-\frac{1}{\sqrt{A_\mu}}\int_0^{\infty}\left(\sqrt{A_\nu}-\frac{\sqrt{A_\mu}}{u^2}\right)e^{-A_\nu u^2-\frac{A_\mu}{u^2}}\,du.$$

Das zweite Glied der rechten Seite verschwindet, da das Integral von $u=0$ bis $u=\sqrt[4]{\frac{A_\mu}{A_\nu}}$ vom Vorzeichen abgesehen mit dem Integral von $u=\sqrt[4]{\frac{A_\mu}{A_\nu}}$ bis $u=\infty$ übereinstimmt. Mit der Substitution von $u=\sqrt{\frac{A_\mu}{A_\nu}}\,\frac{1}{v}$ ist nämlich

$$A_\nu u^2+\frac{A_\mu}{u^2}=A_\nu v^2+\frac{A_\mu}{v^2},$$

$$\left(\sqrt{A_\nu}-\frac{\sqrt{A_\mu}}{u^2}\right)du=\left(\sqrt{A_\nu}-\frac{\sqrt{A_\mu}}{v^2}\right)dv$$

und $u=0$ entspricht

$$v=\infty,\; u=\sqrt[4]{\frac{A_\mu}{A_\nu}} \text{ entspricht } v=\sqrt[4]{\frac{A_\mu}{A_\nu}}.$$

Das erste Glied der rechten Seite ist das bekannte Integral

$$\int_0^\infty e^{-A_\nu u^2-\frac{A_\mu}{u^2}}\,du=\frac{1}{2}\sqrt{\frac{\pi}{A_\nu}}\,e^{-2\sqrt{A_\nu A_\mu}}.$$

So erhält man für das Glied (4. 3)

$$(-1)^{\nu+\mu}4\sqrt{2}\left(\sqrt{A_\nu}+\sqrt{A_\mu}\right)e^{-\left(\sqrt{A_\nu}+\sqrt{A_\mu}\right)^2}=4(-1)^{\nu+\mu}(\nu+\mu+1)\,y\,e^{-2(\nu+\mu+1)^2y^2}.$$

Es sei $\nu+\mu+1=s$ fixe Zahl; durch Summation für $\nu=0,1,2,\ldots,s-1$, also insgesamt $\nu+\mu+1$ Glieder, erhält man

$$4(-1)^{s-1}s^2 y e^{-2s^2y^2},$$

d. h. man gelangt zur Reihe

$$4\sum_{s=1}^\infty(-1)^{s-1}s^2y^2e^{-2s^2y^2},$$

welche gleich der Dichtefunktion der SMIRNOWschen Verteilung ist.

F) Wir bezeichnen mit D jenes dreieckförmige Bereich der (y,z) Ebene, welches durch die Punkte $(0,0)$, $(1,0)$ und $(1,c)$ gebildet wird, wobei $c>0$. Laut dem am Anfang dieses §-s gesagten besteht folgender Zusammenhang:

$$\lim_{n\to\infty}\mathbf{P}\left\{\sqrt{\frac{n}{2}}\,\frac{\sup\limits_{-\infty<x<\infty}(F_n(x)-G_n(x))}{\frac{1}{2}\left(F_n(\xi_0^{(n)}+0)+G_n(\xi_0^{(n)}+0)\right)}<c\right\}=$$

$$=\sqrt{\frac{2}{\pi}}\iint\limits_D\frac{u^2}{(v(1-v))^{3/2}}\,e^{-\frac{1}{2}\frac{u^2}{v(1-v)}}\,du\,dv.$$

Das letztere Integral kann folgenderweise geschrieben werden:

$$\Phi(c)=\sqrt{\frac{2}{\pi}}\int_0^1\left(\int_0^{cv}\frac{u^2}{(v(1-v))^{3/2}}\,e^{-\frac{1}{2}\frac{u^2}{v(1-v)}}\,du\right)dv.$$

Gehen wir zur Ableitung über:

$$\Phi'(c)=\sqrt{\frac{2}{\pi}}\,c^2\int_0^1\frac{v^{3/2}}{(1-v)^{3/2}}\,e^{-\frac{1}{2}c^2\frac{v}{1-v}}\,dv$$

und führen wir die Substitution

$$\frac{v}{1-v}=t,\quad dv=\frac{dt}{(1+t)^2}$$

durch:

$$\Phi'(c) = \sqrt{\frac{2}{\pi}}\, c^2 \int_0^\infty \frac{t^{3/2}}{(1+t)^2} e^{-\frac{1}{2}c^2 t}\, dt.$$

Das letztere Integral ist nicht elementar, kann jedoch durch die WHITTAKERsche Funktion ausgedrückt werden:

$$\int_0^\infty e^{-at} t^{\mu-\lambda-\frac{1}{2}} (1+t)^{\mu+\lambda-\frac{1}{2}} dt = \Gamma\left(\mu-\lambda+\tfrac{1}{2}\right) a^{-\mu-\frac{1}{2}} e^{-\frac{a}{2}} W_{\lambda,\mu}(a).$$

In unserem Falle ist $a=\frac{c^2}{2}$, $\lambda=-\frac{7}{4}$, $\mu=\frac{1}{4}$ und so ist

$$\int_0^\infty e^{-\frac{c^2}{2}t} t^{3/2} (1+t)^{-2} e^{-\frac{1}{2}c^2 t}\, dt = \frac{3}{\sqrt[4]{8}} \sqrt{\pi}\, c^{-\frac{3}{2}} e^{-\frac{c^2}{4}} W_{-\frac{7}{4},\frac{1}{4}}\left(\frac{c^2}{2}\right),$$

woraus wir für $\Phi(c)$ den in Satz 5. angegebenen Ausdruck erhalten.

(Eingegangen: 6. VII. 1957.)

LITERATUR

[1] ANDERSON, T. V.—DARLING, D. A.: „Asymptotic theory of certain »goodness of fit« criteria based on stochastic processes". *Annals of Mathematical Statistics* **23** (1952) 193—212.

[2] BIRNBAUM, Z. W.—PYKE, R.: „On some distributions related to the statistic D_n^+". *Annals of Mathematical Statistics* (in Erscheinung).

[3] DARLING, D. A.—SIEGERT, A. J. F.: „The first passage problem for a continuous Markov process". *Annals of Mathematical Statistics* **24** (1953) 624—639.

[4] DONSKER, M. D.: „Justification and extension of Doob's heuristic approach to the Kolmogorov—Smirnov theorems". *Annals of Mathematical Statistics* **23** (1952), 277—281.

[5] DOOB, J. L.: „Heuristic approach to the Kolmogorov—Smirnov theorems". *Annals of Mathematical Statistics* **20** (1949), 393—403.

[6] FELLER, W.: *An introduction to probability theory and its applications, Vol. I.* Wiley, New York, 1950.

[7] Гнеденко, Б. В.—Королюк, В. С.: „О махимальном расхождении двух эмпирических распределений". Доклады Академии Наук СССР **80** (1951) 525—528.

[8] KOLMOGOROFF, A. N.: „Sulla determinazione empirica di una legge di distribuzione". *Giornale dell'Instituto Italiano degli Attuari* **4** (1933) 83—91.

[9] LEVY, P.: *Processus stochastique et mouvement Brownien.* Gauthier-Villars, Paris, 1948.

[10] PÓLYA, G.—SZEGŐ, G.: *Aufgaben und Lehrsätze aus der Analysis.* Springer, Berlin, 1925.

[11] RÉNYI, A.: „A rendezett minták elméletéről". *A Magyar Tudományos Akadémia III. (Matematikai és Fizikai) Osztályának Közleményei* **3** (1953) 467—503.

[12] RÉNYI, A.: „On the theory of order statistics". *Acta Mathematica Academiae Scientiarum Hungaricae* **4** (1953) 191—232.

[13] RÉNYI, A.: *Valószínűségszámítás.* Tankönyvkiadó, Budapest, 1954.

[14] Смирнов, Н. В.: „Оценка расхождения между эмпирическими кривыми распределения в двух независимых выборках". *Bulletin Mathématique de l'Université de Moscou* **2** (1939) 3—14.

[15] WANG SHOU-YEN: „On the limiting distribution of the ratio of two empirical distribution". *Acta Mathematica Sinica* **5** (1955) 253—267.

KÉT MINTA ÖSSZEHASONLÍTÁSA A MINTAELEMEK KÉT STATISZTIKAI FÜGGVÉNYE ALAPJÁN

VINCZE ISTVÁN

Kivonat

Az irodalomban számos kritérium ismeretes, amely annak eldöntésére szolgál, hogy két minta azonos eloszlásból származik-e. E próbák sokszor nem eléggé érzékenyek bizonyos ellenhipotézisekre. Közelfekvő gondolat illeszkedési próbák finomítására olyan statisztikai próbák konstruálása, amelyek a mintaelemek két (esetleg több) jól megválasztott statisztikai függvényén alapulnak. — Jelen dolgozatban néhány kétváltozós eloszlás- és határeloszlástételt bizonyítunk be, amelyek ilyen próbák konstruálására alkalmasak. Néhány további határeloszlástételt, összefüggéseket ismert határeloszlástételekkel, továbbá a próbák alkalmazásához szükséges táblázatokat — egyenlő darabszámú minták esetére — egy további közleményben kívánunk publikálni.

Legyenek $\xi_1, \xi_2, \ldots, \xi_n$ és $\eta_1, \eta_2, \ldots, \eta_n$ az $F(x)$, illetve $G(x)$ eloszlásfüggvénnyel bíró ξ, illetve η valószínűségi változókból vett n elemű minták. Jelölje $F_n(x)$, illetve $G_n(x)$ a megfelelő empirikus eloszlásfüggvényeket. Legyenek végül $\xi_0^{(n)}$ és $\eta_0^{(n)}$ azok az „első" helyek, ahol az

$$(F_n(x) - G_n(x)), \quad \text{illetve} \quad |F_n(x) - G_n(x)|$$

függvények maximális értéküket felveszik; vagyis a maximum-helyek alsó határai. E jelölésekkel a következő tételek állanak:

1. tétel. *Ha* $F(x) \equiv G(x)$, *akkor*

$$\mathbf{P}\left\{ \sup_{-\infty<x<\infty} (F_n(x) - G_n(x)) = \frac{k}{n},\ \tfrac{1}{2}\left(F_n(\xi_0^{(n)}+0) + G_n(\xi_0^{(n)}+0)\right) = \frac{r}{2n} \right\} =$$

$$= \begin{cases} 0, \text{ ha } k<0 \text{ vagy } k+r \text{ páratlan} \\[2mm] \dfrac{1}{(r-1)(2n-r+2)} \dfrac{\binom{r}{\frac{r}{2}}\binom{2n-r}{n-\frac{r}{2}}}{\binom{2n}{n}}, \quad \text{ha } k=0,\ r=2,4,\ldots, \\[6mm] \dfrac{k(k+1)}{r(2n-r+1)} \dfrac{\binom{r}{\frac{r+k}{2}}\binom{2n-r+1}{n-\frac{r+k}{2}}}{\binom{2n}{n}}, \quad \text{ha } k=1,2,\ldots,n; \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad r=k,\ k+2,\ldots 2n-k. \end{cases}$$

Minthogy fenti eloszlásunknak peremeloszlása a GNYEGYENKO—KOROLJUK-féle [7] egyoldalú eltérésre vonatkozó véges eloszlás, ennélfogva valószínűség-

számítási bizonyítását nyertük a következő azonosságoknak

$$\sum_{(r)}^{*} \frac{k(k+1)}{r(2n-r+1)} \binom{r}{\frac{r+k}{2}} \binom{2n-r+1}{n-\frac{r+k}{2}} = \frac{2k+1}{n+k+1} \binom{2n}{n-k},$$

ahol az összegezés $r=k, k+2, \ldots 2n-k$-ra értendő, míg $1 \leqq k \leqq n$.

A $k=0$ esetre

$$\sum_{(r)}^{*} \frac{1}{(r-1)(2n-r+2)} \binom{r}{\frac{r}{2}} \binom{2n-r}{n-\frac{r}{2}} = \frac{1}{n+1} \binom{2n}{n},$$

ahol az összegezés $r=2,4,6,\ldots$-ra vonatkozik.

Vezessük be a következő jelölést:

$$A_r^{(k)} = \sum_{\nu=0}^{\left[\frac{r}{2n}\right]} (-1)^{\nu} \binom{r}{\frac{r+k}{2}+\nu k} \frac{(2\nu+1)k}{r}.$$

2. tétel. *Ha $F(x) \equiv G(x)$, akkor*

$$\mathbf{P}\left\{ \sup_{-\infty<x<\infty} |F_n(x)-G_n(x)| = \frac{k}{n},\ \tfrac{1}{2}\left(F_n(\eta_0^{(n)}+0)+G_n(\eta_0^{(n)}+0)\right) = \frac{r}{2n} \right\} =$$

$$= \begin{cases} 0, & \text{ha } k<0 \text{ vagy } r+k \text{ páratlan,} \\ \dfrac{2A_r^{(k)} A_{2n-r+1}^{(k+1)}}{\binom{2n}{n}}, & \text{ha } k=1,2,\ldots n;\ r=k, k+2, \ldots 2n-k. \end{cases}$$

Az 1. tétel bizonyítása két egyszerű kombinatorikus lemmán alapszik, míg a 2. tétel bizonyítását bolyongási feladatra vezettük vissza.

Fenti eloszlástételekből a megfelelő határeloszlástételeket úgy nyerjük, hogy n növelésével r-et n-rendben, míg k-t $\sqrt{n}$ rendben tartatjuk végtelen felé, pontosabban

$$k \sim \sqrt{2ny}, \quad r \sim 2nz.$$

3. tétel. Ha $F(x) \equiv G(x)$, akkor — $0 \leqq z \leqq 1$-re —

$$\lim_{n\to\infty} \mathbf{P}\left\{ \sqrt{\frac{n}{2}} \sup_{-\infty<x<\infty} (F_n(x)-G_n(x)) < y,\ \tfrac{1}{2}\left(F_n(\xi_0^{(n)}+0)+G_n(\xi_0^{(n)}+0)\right) < z \right\} =$$

$$= \begin{cases} \sqrt{\dfrac{2}{\pi}} \displaystyle\int_0^y \int_0^z \frac{u^2}{(v(1-v))^{3/2}} e^{-\frac{1}{2}\frac{u^2}{v(1-v)}} \, du\, dv, & \text{ha } y>0, \\ 0, \quad \text{ha } y \leqq 0. \end{cases}$$

Eloszlásunk peremeloszlása a SZMIRNOV-féle [14] eloszlás, vagyis z szerint integrálva nyerjük

$$\lim_{n\to\infty} \mathbf{P}\left\{ \sqrt{\frac{n}{2}} \sup_{-\infty<x<\infty} (F_n(x)-G_n(x)) < y \right\} = \begin{cases} 1-e^{-2y^2}, & \text{ha } y>0, \\ 0, & \text{ha } y \leqq 0. \end{cases}$$

Ha y szerint integrálunk 0-tól ∞-ig, akkor egyenletes eloszláshoz jutunk, amiből adódik, hogy *a* $\xi_0^{(n)}$ *változó határeloszlása megegyezik* ξ *és* η *közös eloszlásával.*

Vezessük be a következő jelölést:

$$f(y,z)=\frac{y}{z^{3/2}}\sum_{\nu=0}^{\infty}(-1)^{\nu}(2\nu+1)\,e^{-\frac{1}{2}(2\nu+1)^2\frac{y^2}{z}}.$$

4. tétel. *Ha* $F(x)\equiv G(x)$, *akkor* — $0\leqq z\leqq 1$*-re* —

$$\lim_{n\to\infty}\mathbf{P}\left\{\sqrt{\frac{n}{2}}\sup_{-\infty<x<\infty}|F_n(x)-G_n(x)|<y,\ \tfrac{1}{2}\left(F_n(\eta_0^{(n)}+0)+G_n(\eta_0^{(n)}+0)\right)<z\right\}=$$

$$=\begin{cases}\sqrt{\dfrac{8}{\pi}}\displaystyle\int_0^y\int_0^z f(u,v)\,f(u,1-v)\,du\,dv, & \text{ha } y>0,\\ 0, & \text{ha } y\leqq 0.\end{cases}$$

z szerinti integráció 0-tól 1-ig az abszolút eltérésre vonatkozó KOLMOGOROV—SZMIRNOV-féle eloszláshoz vezet:

$$\lim_{n\to\infty}\mathbf{P}\left\{\sqrt{\frac{n}{2}}\sup_{-\infty<x<\infty}|F_n(x)-G_n(x)|\right\}=\begin{cases}\displaystyle\sum_{\nu=-\infty}^{\infty}(-1)^{\nu}e^{-2\nu^2y^2}, & \text{ha } y>0\\ 0, & \text{ha } y\leqq 0.\end{cases}$$

5. tétel. *Ha* $F(x)\equiv G_n(x)$, *akkor*

$$\lim_{n\to\infty}\mathbf{P}\left\{\sqrt{\frac{n}{2}}\,\frac{\sup\limits_{-\infty<x<\infty}(F_n(x)-G_n(x))}{\frac{1}{2}(F_n(\xi_0^{(n)}+0)+G_n(\xi_0^{(n)}+0))}<c\right\}=$$

$$=\begin{cases}\dfrac{3}{\sqrt[4]{8}}\displaystyle\int_0^c e^{\frac{t^4}{4}}W_{-\frac{7}{4},\frac{1}{4}}\left(\frac{t^2}{2}\right)dt, & \textit{ha}\ c>0,\\ 0, & \textit{ha}\ c\leqq 0,\end{cases}$$

ahol W *az ún. Whittaker-féle függvényt jelenti:*

$$W_{\lambda,\mu}(z)=\frac{z^{\mu+\frac{1}{2}}e^{-\frac{z}{2}}}{\Gamma(\mu-\lambda+\frac{1}{2})}\int_0^{\infty}e^{-zt}t^{\mu-\lambda-\frac{1}{2}}(1+t)^{\mu+\lambda-\frac{1}{2}}dt,$$

ha

$$\Re e(\mu-\lambda)>-\tfrac{1}{2},\quad |\arg z|<\pi.$$

ДВУМЕРНЫЕ ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ В ТЕОРИИ ВАРИАЦИОННЫХ РЯДОВ

I. VINCZE

Резюме

В литературе известен ряд критериев для того, чтобы определить происходят ли два образца из тождественного распределения. Эти пробы часто недостаточно чувствительны в отношении некоторых контргипотез. Естественно построение таких статистических критериев, которые основываются на двух (или больше) хорошо выбранных статистических функциях элементов образца. В настоящей работе доказывается несколько теорем о распределении и предельном распределении двух переменных, которые годятся для построения таких критериев. Несколько дальнейших теорем о предельном распределении, связи с известными теоремами о предельном распределенни и таблицы, нужные для применения проб — в случае образцов с одинаковым числом элементо в —, автор собирается опубликовать в дальнейшем.

Пусть $\xi_1, \xi_2, \ldots, \xi_n$ и $\eta_1, \eta_2, \ldots, \eta_n$ суть n-элементные образцы, взятые из случайных величин ξ и η с функцией распределения $F(x)$ и $G(x)$. Пусть $F_n(x)$ и $G_n(x)$ обозначают соответствующие эмпрические функции распределения. Пусть, наконец, $\xi_n^{(n)}$ и $\eta_0^{(n)}$ суть те „первые“ места, где функции $(F_n(x) - G_n(x))$ и $|F_n(x) - G_n(x)|$ принимают своё наибольшее значение, т. е. нижние границы мест максимума. Имеют место цледующие теоремы.

Теорема I. Если $F(x) \equiv G(x)$, то

$$\mathbf{P}\left\{ \sup_{-\infty<x<+\infty} (F_n(x) - G_n(x)) = \frac{k}{n}, \quad \tfrac{1}{2}(F_n(\xi_0^{(n)}+0) + G_n(\xi_0^{(n)}+0)) = \frac{r}{2n} \right\} =$$

$$= \begin{cases} 0 \text{ если } k<0 \text{ или } k+r \text{ нечётно} \\[2mm] \dfrac{1}{(r-1)(2n-r+2)} \dfrac{\dbinom{r}{\frac{r}{2}}\dbinom{2n-r}{n-\frac{r}{2}}}{\dbinom{2n}{n}}, & \text{если} \quad k=0, \quad r=2,4,\ldots \\[2mm] \dfrac{k(k+1)}{r(2n-r+1)} \dfrac{\dbinom{r}{\frac{r+k}{2}}\dbinom{2n-r+1}{n-\frac{r+k}{2}}}{\dbinom{2n}{n}}, & \text{если} \quad k=1,2,\ldots,n, \\ & \qquad r=k, k+2, \ldots, 2n-k. \end{cases}$$

Так как граничное распределение этого распределения есть конечное распределение Гнеденко—Королюка [7], относящееся к одностороннему отклонению, то мы нашли теоретико-вероятностное доказательство следующих тождеств:

$$\sum_{(r)}^{*} \frac{k(k+1)}{r(2n-r+1)} \binom{r}{\frac{r+k}{2}} \binom{2n-r+1}{n-\frac{r+k}{2}} = \frac{2k+1}{n+k+1} \binom{2n}{n-k}$$

где суммирование происходит для $r = k, k+2, \ldots, 2n-k$ и $1 \leqq k \leqq n$.

Для случая $k = 0$

$$\sum_{(r)}^{*} \frac{1}{(r-1)(2n-r+2)} \binom{r}{\frac{r}{2}} \binom{2n-r}{n-\frac{r}{2}} = \frac{1}{n+1} \binom{2n}{n}$$

где суммирования производится для $r = 2, 4, 6, \ldots$

Введём следующее обозначение:

$$A_r^{(k)} = \sum_{\nu=0}^{\left[\frac{r}{2n}\right]} (-1)^{\nu} \binom{r}{\frac{r+k}{2}+\nu k} \frac{(2\nu+1)k}{r}.$$

Теорема 2. Если $F(x) \equiv G(x)$, то

$$\mathbf{P}\left\{ \sup_{-\infty<x<\infty} |F_n(x) - G_n(x)| = \frac{k}{n}, \quad \tfrac{1}{2}(F_n(\eta_0^{(n)}+0) + G_n(\eta_0^{(n)}+0)) = \frac{r}{2n} \right\} =$$

$$\begin{cases} 0, & \text{если } k < 0 \text{ или } k+r \text{ нечётно,} \\ \dfrac{2A_r^{(k)} A_{2k-r+1}^{(k-1)}}{\binom{2n}{r}}, & \text{если } k = 1, 2, \ldots, n; \ r = k, k+2, \ldots, 2n-k. \end{cases}$$

Доказательство теоремы I основывается на двух простых комбинаторных леимах, в то время, как доказательство теоремы 2 сводится к задаче блуждания.

Из приведённых теорем распределения получатся соответствующие теоремы предельного распределения, если с увеличением стремится к бесконечности как n, а $k \sqrt{n}$, тоьнее

$$k \sim \sqrt{2n}\, y \quad r \sim 2nz$$

Теорема 3. Если $F(x) \equiv G(x)$, то для $0 \leqq z \leqq 1$

$$\lim_{n\to\infty} \mathbf{P}\left\{ \sqrt{\frac{n}{2}} \sup_{-\infty<x<\infty} (F_n(x) - G_n(x)) < y, \ \tfrac{1}{2}(F_n(\xi_0^{(n)}+0) + G_n(\xi_0^{(n)}+0)) < z \right\} =$$

$$= \begin{cases} \sqrt{\dfrac{2}{\pi}} \displaystyle\int_0^{y} \int_0^{z} \frac{u^2}{(v(1-v))^{3/2}} e^{-\frac{1}{2}\frac{u^2}{v(1-v)}} du\, dv, & \text{если } y < 0 \\ 0, & \text{если } y \leqq 0. \end{cases}$$

Граничное распределение этого распределения есть распределение Смирнова [14], т. е., интегрируя по z, получаем:

$$\lim \mathbf{P}\left\{\sqrt{\frac{n}{2}}\sup_{-\infty<x<\infty}(F_n(x)-G_n(x))<y\right\}=\begin{cases}1-e^{-2y^2}, & \text{если } y>0\\ 0, & \text{если } y\leqq 0.\end{cases}$$

Если интегрировать по y от 0 до ∞, то получается равномерное распределение, откуда следует, что граничное распределение переменного $\xi_0^{(n)}$ совпадает с общим распределением ξ и η.

Введём следующее обозначение:

$$f(x,y)=\frac{y}{z^{3/2}}\sum_{\nu=0}^{\infty}(-1)^{\nu}(2\nu+1)e^{-\frac{1}{2}(2\nu+1)^2\frac{y^2}{z}}.$$

Теорема 4. Если $F(x)\equiv G(x)$, то для $0\leqq z\leqq 1$

$$\lim_{n\to\infty}\mathbf{P}\left\{\sqrt{\frac{n}{2}}\sup_{-\infty<x<\infty}|(F_n(x)-G_n(x)|<y,\quad \tfrac{1}{2}(F_n(\eta_0^{(n)}+0)+\eta_n(\eta_0^{(n)}+0))<z\right\}=$$

$$=\begin{cases}\sqrt{\frac{8}{\pi}}\int\limits_0^y\int\limits_0^z f(u,v)f(u,1-v)\,du\,dv, & \text{если } y>0\\ 0, \quad \text{если } y\leqq 0.\end{cases}$$

Интегрирование по z от 0 до 1 приводит к распределению Колмогорова—Смирнова относящемуся к абсолютному отклонению:

$$\lim_{n\to\infty}\mathbf{P}\left\{\sqrt{\frac{n}{2}}\sup_{-\infty<x<\infty}|F_n(x)-G_n(x)|<y\right\}=\begin{cases}\sum\limits_{\nu=-\infty}^{\infty}(-1)^{\nu}e^{-2\nu^2y^2}, & \text{если } y>0\\ 0, & \text{если } y\leqq 0.\end{cases}$$

Теорема 5. Если $F(x)\equiv G(x)$, то

$$\lim_{n\to\infty}\mathbf{P}\left\{\sqrt{\frac{n}{2}}\frac{\sup\limits_{-\infty<x<\infty}(F_n(x)-G_n(x))}{\frac{1}{2}(F_n(\xi_0^{(n)}+0)+G_n(\xi_0^{(n)}+0))}<c\right\}=$$

$$=\begin{cases}\frac{3}{\sqrt[4]{8}}\int\limits_0^c\sqrt{t}\,e^{\frac{t^2}{4}}W_{-\frac{7}{4},\frac{1}{4}}\left(\frac{t^2}{2}\right)dt & \text{если } c>0\\ 0, \quad \text{если } c\leqq 0,\end{cases}$$

где W обозначает так называемую функцию Whittaker:

$$W_{\lambda,\mu}(z)=\frac{z^{\mu+\frac{1}{2}}e^{-\frac{z}{2}}}{\Gamma(\mu-\lambda+\frac{1}{2})}\int\limits_0^{\infty}e^{-zt}t^{\mu-\lambda-\frac{1}{2}}(1+t)^{\mu+\lambda-\frac{1}{2}}dt$$

если

$$\mathfrak{Re}(\mu-\lambda)>-\tfrac{1}{2},\qquad |\arg z|<\pi.$$