

Zeitschrift: Commentarii Mathematici Helvetici
Band: 40 (1965-1966)

Artikel: Beiträge zu den Theoremen von KOLMOGOROV und SMIRNOV.
Autor: Knüsel, Leo
DOI: <https://doi.org/10.5169/seals-30634>

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. [Siehe Rechtliche Hinweise.](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. [Voir Informations légales.](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. [See Legal notice.](#)

Download PDF: 06.10.2024

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

Beiträge zu den Theoremen von KOLMOGOROV und SMIRNOV

VON LEO KNÜSEL, Zürich

Im folgenden geben wir zwei Beiträge zu den klassischen Theoremen von KOLMOGOROV und SMIRNOV zur Stichprobentheorie¹).

1. Wir geben einen neuen Beweis für einen Satz von M. TAHA [8]. Wir betrachten eine Grundgesamtheit mit der Verteilungsfunktion $F(x)$. $F(x)$ sei überall stetig, außer in den Punkten x_ν ($\nu = 1, 2, \dots, n$). Es sei

$$\begin{aligned} f_{2\nu-1} &= F(x_\nu - 0) & (\nu = 1, \dots, n) \\ f_{2\nu} &= F(x_\nu + 0) & (\nu = 1, \dots, n) \\ f_0 &= 0 \\ f_{2n+1} &= 1 \\ s_k &= f_k - f_{k-1} & (k = 1, \dots, 2n + 1). \end{aligned}$$

$F(x)$ kann in jedem Intervall $[x_\nu, x_{\nu+1})$ dargestellt werden als

$$F(x) = f_{2\nu} + (f_{2\nu+1} - f_{2\nu}) \cdot G_\nu(x) \quad (\nu = 0, 1, \dots, n).$$

Dabei setzen wir

$$\begin{aligned} x_0 &= -\infty \\ x_{n+1} &= +\infty. \end{aligned}$$

Jedes $G_\nu(x)$ ist eine stetige Funktion auf $[x_\nu, x_{\nu+1})$. Wenn wir setzen

$$G_\nu^*(x) = \begin{cases} 0 & \text{für } -\infty < x < x_\nu \\ G_\nu(x) & \text{für } x_\nu \leq x < x_{\nu+1} \\ 1 & \text{für } x_{\nu+1} \leq x < +\infty \end{cases}$$

so ist jedes $G_\nu^*(x)$ eine stetige Verteilungsfunktion ($\nu = 0, 1, \dots, n$). Weiter seien X_1, X_2, \dots, X_N die Elemente einer Stichprobe aus einer Grundgesamtheit mit der Verteilungsfunktion $F(x)$ und $S_N(x)$ sei die zugehörige empirische Verteilungsfunktion, das heißt

$S_N(x) = \frac{k}{N}$ falls genau k -Werte aus der Stichprobe X_1, \dots, X_N kleiner oder gleich x sind.

¹) Resultate aus meiner Diplomarbeit an der ETH über das von Herrn Prof. Dr. W. SAXER gestellte Thema.

Wir setzen

$$\begin{aligned}\bar{f}_{2\nu-1} &= S_N(x_\nu - 0) & (\nu = 1, \dots, n) \\ \bar{f}_{2\nu} &= S_N(x_\nu + 0) & (\nu = 1, \dots, n) \\ \bar{f}_0 &= 0 \\ \bar{f}_{2n+1} &= 1 \\ \bar{s}_k &= \bar{f}_k - \bar{f}_{k-1} & (k = 1, \dots, 2n + 1).\end{aligned}$$

Nun sei weiter

$$\begin{aligned}\bar{F}(x) &= \bar{f}_{2\nu} + (\bar{f}_{2\nu+1} - \bar{f}_{2\nu}) \cdot G_\nu(x) \text{ für } x_\nu \leq x < x_{\nu+1} \quad (\nu = 0, 1, \dots, n) \\ \bar{D}_N &= \sup_{-\infty < x < +\infty} |S_N(x) - \bar{F}(x)| \\ \bar{D}_N^+ &= \sup_{-\infty < x < +\infty} [S_N(x) - \bar{F}(x)].\end{aligned}$$

Der Satz von TAHA lautet:

$$\lim_{N \rightarrow \infty} P\{N^{\frac{1}{2}} \cdot \bar{D}_N \leq z\} = \prod_{\nu=0}^n K\left(\frac{z}{\sqrt{s_{2\nu+1}}}\right) \quad (1)$$

$$\lim_{N \rightarrow \infty} P\{N^{\frac{1}{2}} \cdot \bar{D}_N^+ \leq z\} = \prod_{\nu=0}^n K^+\left(\frac{z}{\sqrt{s_{2\nu+1}}}\right) \quad (2)$$

wobei:

$$K(y) = \begin{cases} 0 & \text{für } y \leq 0 \\ \sum_{k=-\infty}^{+\infty} (-1)^k \cdot e^{-2k^2 y^2} & \text{für } y > 0 \end{cases} \quad (3)$$

$$K^+(y) = \begin{cases} 0 & \text{für } y \leq 0 \\ 1 - e^{-2y^2} & \text{für } y > 0. \end{cases} \quad (4)$$

Wir setzen

$R_{2\nu} = N \cdot \bar{s}_{2\nu}$ = Anzahl der Elemente X_j aus der Stichprobe

X_1, \dots, X_N mit $X_j = x_\nu$ ($\nu = 1, \dots, n$)

$R_{2\nu-1} = N \cdot \bar{s}_{2\nu-1}$ = Anzahl der Elemente X_j aus der Stichprobe

X_1, \dots, X_N mit $x_{\nu-1} < X_j < x_\nu$ ($\nu = 1, \dots, n + 1$).

Mit $A_{j_1 \dots j_{2n+1}}$ bezeichnen wir das Ereignis, das eintritt, wenn alle Gleichungen

$$R_k = j_k \quad (k = 1, \dots, 2n + 1)$$

erfüllt sind, und mit B meinen wir das Ereignis, das eintritt, wenn

$$N^{\frac{1}{2}} \cdot \bar{D}_N \leq z$$

ist. Es gilt

$$P\{B\} = \sum_{j_1 + \dots + j_{2n+1} = N} P\{A_{j_1 \dots j_{2n+1}}\} \cdot P\{B/A_{j_1 \dots j_{2n+1}}\}. \quad (5)$$

Die Summation ist über alle Tupel (j_1, \dots, j_{2n+1}) mit $j_1 + \dots + j_{2n+1} = N$ zu erstrecken. Mit $P\{B/A\}$ bezeichnen wir die bedingte Wahrscheinlichkeit für das Ereignis B unter der Bedingung, daß das Ereignis A eingetreten ist. Nun gilt weiter:

$$\begin{aligned} P\{B/A_{j_1 \dots j_{2n+1}}\} &= P\{N^{\frac{1}{2}} \cdot \sup_{-\infty < x < +\infty} |S_N(x) - \bar{F}(x)| \leq z/A_{j_1 \dots j_{2n+1}}\} = \\ &= \prod_{\nu=0}^n P\{N^{\frac{1}{2}} \cdot \sup_{x_\nu \leq x < x_{\nu+1}} |S_N(x) - [\bar{f}_{2\nu} + \bar{s}_{2\nu+1} \cdot G_\nu(x)]| \leq z/A_{j_1 \dots j_{2n+1}}\} = \\ &= \prod_{\nu=0}^n P\left\{(N \cdot \bar{s}_{2\nu+1})^{\frac{1}{2}} \cdot \sup_{x_\nu \leq x < x_{\nu+1}} \left| \frac{S_N(x) - \bar{f}_{2\nu}}{\bar{s}_{2\nu+1}} - G_\nu(x) \right| \leq \frac{z}{(\bar{s}_{2\nu+1})^{\frac{1}{2}}} / A_{j_1 \dots j_{2n+1}}\right\}. \end{aligned} \quad (6)$$

Wenn wir setzen

$$S_\nu^*(x) = \begin{cases} 0 & \text{für } -\infty < x < x_\nu \\ \frac{S_N(x) - \bar{f}_{2\nu}}{\bar{s}_{2\nu+1}} & \text{für } x_\nu \leq x < x_{\nu+1} \\ 1 & \text{für } x_{\nu+1} \leq x < +\infty \end{cases}$$

so kann jedes $S_\nu^*(x)$ aufgefaßt werden als empirische Verteilungsfunktion zu einer Stichprobe mit dem Umfang $N \cdot \bar{s}_{2\nu+1} = R_{2\nu+1}$ aus einer Grundgesamtheit mit der stetigen Verteilungsfunktion $G_\nu^*(x)$. Weil nun

$$(N \cdot \bar{s}_{2\nu+1})^{\frac{1}{2}} \cdot \sup_{x_\nu \leq x < x_{\nu+1}} \left| \frac{S_N(x) - \bar{f}_{2\nu}}{\bar{s}_{2\nu+1}} - G_\nu(x) \right| = (R_{2\nu+1})^{\frac{1}{2}} \cdot \sup_{-\infty < x < +\infty} |S_\nu^*(x) - G_\nu^*(x)|$$

ist, so können wir auf jeden einzelnen Faktor des Produktes (6) das bekannte Theorem von KOLMOGOROV anwenden und erhalten:

$$\lim_{N \rightarrow \infty} P\{B/A_{j_1 \dots j_{2n+1}}\} = \prod_{\nu=0}^n K\left(\frac{z}{(\bar{s}_{2\nu+1})^{\frac{1}{2}}}\right). \quad (7)$$

Weiter ist die gemeinsame Verteilung der R_k ($k = 1, 2, \dots, 2n + 1$) eine Polynominalverteilung mit den Parametern s_k . Es gilt daher (Grenzwertsatz für die Polynominalverteilung):

$$\begin{aligned} \lim_{N \rightarrow \infty} P\left\{\left| \frac{R_k - N s_k}{\sqrt{N}} \right| < \lambda_k \text{ für } k = 1, \dots, 2n + 1\right\} &= \\ &= \int_G \dots \int [(2\pi)^{2n} \cdot \prod_{i=1}^{2n+1} s_i]^{-\frac{1}{2}} \cdot e^{-\frac{1}{2} \cdot \sum_{i=1}^{2n+1} \frac{x_i^2}{s_i}} dx_1 \dots dx_{2n}. \end{aligned} \quad (8)$$

Dabei ist zu setzen

$$x_{2n+1} = - \sum_{i=1}^{2n} x_i$$

und das Integrationsgebiet G ist festgelegt durch

$$\left| \sum_{i=1}^{2n} x_i \right| < \lambda_{2n+1} \quad \text{und} \quad |x_k| < \lambda_k \quad \text{für} \quad k = 1, \dots, 2n.$$

Wir wollen jetzt die Summe (5) aufspalten und dann den Grenzübergang $N \rightarrow \infty$ machen. Wir betrachten folgende Mengen von Tupeln (j_1, \dots, j_{2n+1}) :

$$M = \{j = (j_1, \dots, j_{2n+1}) \mid \sum_{i=1}^{2n+1} j_i = N\}$$

$$M_1 = \left\{ j \in M \mid \left| \frac{j_k - N s_k}{N} \right| < \frac{1}{\sqrt[4]{N}} \quad \text{für} \quad k = 1, \dots, 2n+1 \right\}$$

$$M_2 = M - M_1.$$

Es gilt

$$P\{B\} = \sum_M P\{A_{j_1 \dots j_{2n+1}}\} \cdot P\{B/A_{j_1 \dots j_{2n+1}}\} = \sum_{M_1} \dots + \sum_{M_2} \dots$$

Für die Glieder der Summe über M_1 gilt:

$$\left| \frac{R_k - N s_k}{N} \right| = |\bar{s}_k - s_k| < \frac{1}{\sqrt[4]{N}} \rightarrow 0 \quad \text{für} \quad N \rightarrow \infty \quad (9)$$

$$\left| \frac{R_k - N s_k}{\sqrt{N}} \right| < \sqrt[4]{N} \rightarrow \infty \quad \text{für} \quad N \rightarrow \infty. \quad (10)$$

Damit können wir $P\{B\}$ wie folgt abschätzen:

$$\begin{aligned} P\{B\} &\geq \sum_{M_1} P\{A_{j_1 \dots j_{2n+1}}\} \cdot P\{B/A_{j_1 \dots j_{2n+1}}\} \geq \\ &\geq \inf_{M_1} P\{B/A_{j_1 \dots j_{2n+1}}\} \cdot \sum_{M_1} P\{A_{j_1 \dots j_{2n+1}}\}. \end{aligned}$$

Wegen (7) und (9) gilt für $N \rightarrow \infty$:

$$\inf_{M_1} P\{B/A_{j_1 \dots j_{2n+1}}\} \rightarrow \prod_{\nu=1}^n K \left(\frac{z}{(s_{2\nu+1})^{\frac{1}{2}}} \right)$$

und wegen (8) und (10):

$$\sum_{M_1} P\{A_{j_1 \dots j_{2n+1}}\} \rightarrow 1,$$

so daß wir also erhalten:

$$\lim_{N \rightarrow \infty} P\{B\} \geq \prod_{\nu=0}^n K\left(\frac{z}{(s_{2\nu+1})^{\frac{1}{2}}}\right). \quad (11)$$

Andererseits gilt

$$P\{B\} \leq \sup_{M_1} P\{B/A_{j_1 \dots j_{2n+1}}\} \cdot \sum_{M_1} P\{A_{j_1 \dots j_{2n+1}}\} + 1 \cdot \sum_{M_2} P\{A_{j_1 \dots j_{2n+1}}\}.$$

Für $N \rightarrow \infty$ gilt wegen (7), (8), (9) und (10):

$$\begin{aligned} \sup_{M_1} P\{B/A_{j_1 \dots j_{2n+1}}\} &\rightarrow \prod_{\nu=0}^n K\left(\frac{z}{(s_{2\nu+1})^{\frac{1}{2}}}\right) \\ \sum_{M_1} P\{A_{j_1 \dots j_{2n+1}}\} &\rightarrow 1 \\ \sum_{M_2} P\{A_{j_1 \dots j_{2n+1}}\} &\rightarrow 0, \end{aligned}$$

und somit ist

$$\lim_{N \rightarrow \infty} P\{B\} \leq \prod_{\nu=0}^n K\left(\frac{z}{(s_{2\nu+1})^{\frac{1}{2}}}\right). \quad (12)$$

Aus (11) und (12) folgt

$$\lim_{N \rightarrow \infty} P\{B\} = \prod_{\nu=0}^n K\left(\frac{z}{(s_{2\nu+1})^{\frac{1}{2}}}\right).$$

Dies ist das Resultat (1) von M. TAHA. Das Resultat (2) kann genau gleich bewiesen werden. (Man ersetze im obigen Beweis durchwegs \bar{D}_N durch \bar{D}_N^+ und die KOLMOGOROV'sche Verteilungsfunktion $K(y)$ durch die Verteilungsfunktion $K^+(y)$.)

II. A. N. KOLMOGOROV erwähnt in seiner Arbeit [3], daß die Testgrößen D_N und $D_{N_1 N_2}$ (Testgrößen von KOLMOGOROV und SMIRNOV) konservativen Charakter haben, wenn die Verteilungsfunktion $F(x)$ der Grundgesamtheit unstetig ist. Bewiesen wurde dies für den diskreten Fall von G. E. NOETHER [5]. Wir geben hier einen Beweis für diese Tatsache, und zwar für die Testgröße D_N und für den Fall, daß $F(x)$ endlich viele Unstetigkeitsstellen besitzt.

Es seien X_1, \dots, X_N die Elemente einer Stichprobe aus einer Grundgesamtheit mit der Verteilungsfunktion $F(x)$ und $S_N(x)$ sei die zugehörige empirische Verteilungsfunktion. Wir setzen:

$$D_N = \sup_{-\infty < x < +\infty} |S_N(x) - F(x)|.$$

Wenn die Verteilungsfunktion $F(x)$ stetig ist, so ist die Verteilungsfunktion

$$\Phi_N(z) = P\{N^{\frac{1}{2}} \cdot D_N \leq z\}$$

wohlbekannt (siehe [4]). Sie ist unabhängig von der speziellen Form der Verteilungsfunktion $F(x)$; $F(x)$ braucht nur stetig zu sein. Insbesondere kennt man die Grenzverteilung

$$\lim_{N \rightarrow \infty} \Phi_N(z) = K(z).$$

$K(z)$ ist die KOLMOGOROV'sche Verteilungsfunktion (3). Man kann nun die Größe D_N dazu benutzen, um die Hypothese zu prüfen, daß eine vorgelegte Stichprobe X_1, \dots, X_N aus einer Grundgesamtheit mit der vorgegebenen stetigen Verteilungsfunktion $F(x)$ stamme. Dazu berechnet man aus der vorgelegten Stichprobe und der vorgegebenen Verteilungsfunktion die Größe $N^{\frac{1}{2}} \cdot D_N$ und bestimmt den kritischen Wert z_0 als Lösung der Gleichung

$$\Phi_N(z_0) = 1 - P_\alpha.$$

Dabei ist P_α eine vorgegebene Sicherheitsschwelle (Risiko für einen Fehlentscheid). Wenn nun die berechnete Größe $N^{\frac{1}{2}} \cdot D_N$ größer ist als der kritische Wert z_0 , so verwerfen wir die Hypothese, daß die Stichprobe aus einer Grundgesamtheit mit der Verteilungsfunktion $F(x)$ stamme; andernfalls machen wir keine Aussage. Es stellt sich die Frage, ob sich das Risiko für einen Fehlentscheid vergrößert, wenn die vorgelegte Stichprobe aus einer Grundgesamtheit mit einer unstetigen Verteilungsfunktion $F'(x)$ stammt, wir aber den kritischen Wert z_0 wie oben bestimmen. Diese Frage kann verneint werden. Das Risiko für einen Fehlentscheid vergrößert sich nicht. Man sagt, die Testgröße D_N sei konservativ. Diese Behauptung wollen wir nun beweisen.

Es sei X_1, \dots, X_N eine Stichprobe aus einer Grundgesamtheit mit der stetigen Verteilungsfunktion $F(x)$ und X'_1, \dots, X'_N eine Stichprobe aus einer Grundgesamtheit mit der beliebigen Verteilungsfunktion $F'(x)$. Weiter seien D_N und D'_N die entsprechenden Testgrößen und es sei

$$\Phi_N(z) = P\{N^{\frac{1}{2}} \cdot D_N \leq z\}$$

$$\Phi'_N(z) = P\{N^{\frac{1}{2}} \cdot D'_N \leq z\}.$$

Wir bestimmen z_0 und z'_0 als Lösungen der Gleichungen

$$\Phi_N(z_0) = 1 - P_\alpha \quad \text{beziehungsweise} \quad \Phi'_N(z'_0) = 1 - P_\alpha.$$

Die Testgröße D_N ist konservativ, wenn für jede Wahl von P_α gilt: $z_0 \geq z'_0$, das heißt wenn

$$\Phi_N(z) \leq \Phi'_N(z) \quad \text{für } 0 \leq z < \infty. \tag{13}$$

Wir wollen zunächst die Beziehung (13) beweisen für den Fall, daß $F'(x)$ eine einzige Unstetigkeitsstelle besitzt, und zwar bei $x = 0$. Wir definieren eine Folge von Verteilungsfunktionen:

$$F^k(x) = \begin{cases} F'(x) & \text{für } -\infty < x < -\frac{1}{k} \\ F'\left(-\frac{1}{k}\right) + \left[F'\left(\frac{1}{k}\right) - F'\left(-\frac{1}{k}\right)\right] \cdot \frac{k}{2} \cdot \left(x + \frac{1}{k}\right) & \text{für } -\frac{1}{k} \leq x \leq +\frac{1}{k} \\ F'(x) & \text{für } +\frac{1}{k} < x < +\infty \end{cases}$$

($k = 1, 2, \dots$). Jedes $F^k(x)$ ist eine stetige Verteilungsfunktion. Für jedes k sei nun X_1^k, \dots, X_N^k eine Stichprobe aus einer Grundgesamtheit mit der Verteilungsfunktion $F^k(x)$, $S_N^k(x)$ die zugehörige empirische Verteilungsfunktion und D_N^k die entsprechende Testgröße. Weiter sei

$$M_k = \left\{ x \mid -\infty < x < -\frac{1}{k} \quad \text{oder} \quad +\frac{1}{k} < x < +\infty \right\}.$$

Es gilt:

$$D_N^k = \sup_{-\infty < x < +\infty} |S_N^k(x) - F^k(x)| \geq \sup_{M_k} |S_N^k(x) - F^k(x)|.$$

Damit folgt für jedes k :

$$\Phi_N(z) = P\{N^{\frac{1}{2}} \cdot D_N^k \leq z\} \leq P\{N^{\frac{1}{2}} \cdot \sup_{M_k} |S_N^k(x) - F^k(x)| \leq z\}.$$

Für $k \rightarrow \infty$ gilt offensichtlich:

$$P\{N^{\frac{1}{2}} \cdot \sup_{M_k} |S_N^k(x) - F^k(x)| \leq z\} \rightarrow \Phi'_N(z)$$

und es folgt somit:

$$\Phi_N(z) \leq \Phi'_N(z) \quad \text{für } 0 \leq z < \infty.$$

Damit ist (13) bewiesen für den Fall, daß $F'(x)$ eine einzige Unstetigkeitsstelle besitzt. Wenn $F'(x)$ endlich viele Sprünge hat, kann man den Beweis ganz analog führen, da ja die Unstetigkeitsstellen in diesem Fall isoliert liegen.

Ebenso kann man beweisen, daß die Testgröße D_N^+ für einseitige Abweichungen und die beiden entsprechenden Testgrößen $D_{N_1 N_2}$ und $D_{N_1 N_2}^+$ beim Zwei-Stichproben-Problem konservativen Charakter haben.

Die Beziehung (13) gilt für jedes N ($N = 1, 2, \dots$). Sie überträgt sich daher auch auf die Grenzverteilungen bei $N \rightarrow \infty$. Wir setzen:

$$\lim_{N \rightarrow \infty} \Phi'_N(z) = \Phi(z).$$

Es gilt also:

$$K(z) \leq \Phi(z) \quad \text{für } 0 \leq z < \infty.$$

P. SCHMID gibt in [6] eine Darstellung der Funktion $\Phi(z)$ bei endlich vielen Sprüngen der Verteilungsfunktion $F'(x)$.

LITERATURVERZEICHNIS

- [1] CARNAL H.: *Sur les théorèmes de Kolmogorov et Smirnov dans le cas d'une distribution discontinue*. Comment. Math. Helvetici 37, 19–35 (1962/63).
- [2] KOLMOGOROV A. N.: *Sulla determinazione empirica di una legge di distribuzione*. Giorn. Inst. Ital. Attuari 4, 83–91 (1933).
- [3] KOLMOGOROV A. N.: *Confidence limits for an unknown distribution function*. Annals of Mathematical Statistics 12, 461–463 (1941).
- [4] KOROLYUK V. S.: *On the discrepancy of empiric distributions for the case of two independent samples*. Izv. Akad. Nauk SSSR. Ser. Mat. 19 (1955), 81–96 (Russisch). Selected Translations in Mathematical Statistics 4, 105–121 (1964) (Engl.).
- [5] NOETHER G. E.: *Note on the Kolmogorov statistic in the discrete case*. Metrika 7, 115–116 (1963).
- [6] SCHMID P.: *On the Kolmogorov and Smirnov limit theorems for discontinuous distribution functions*. Annals of Mathematical Statistics 29, 1011–1027 (1958).
- [7] SMIRNOV N. V.: *Sur les écarts de la courbe de distribution empirique*. Mat. Sbornik 6, 3–26 (1939) (Russisch).
- [8] TAHA M.: *Über die verallgemeinerten Tests von Kolmogorov und Smirnov für unstetige Verteilungen*. Mitteilungen der Vereinigung schweiz. Versicherungsmath. 64, 145–174 (1964).

(Eingegangen den 23. April 1965)