Breakthroughs in Statistical Methodology

Kolmogorov (1933). On the Empirical Determination of a Distribution.

Hausarbeit im Rahmen des Seminars
„Breakthroughs in Statistical Methodology“
am Institut für Statistik der
Ludwig-Maximilians-Universität
zu München

Leitung:
Dr. Andrea Wiencierz, Paul Fink

eingereicht von:
Johannes Staudacher (Matr. Nr. 10391535)

München, den 28.05.2014

Gliederung:

1. Frühe Jahre von A. Kolmogorov und Entwicklung bis 1933

Bei Nennung des Names Kolmogorov fallen einem Statistikstudenten wohl zuerst die Axiome von Kolmogorov, also Rechenregeln für Wahrscheinlichkeiten, ein. Diese spielen in der Vorlesung „Einführung in die Wahrscheinlichkeitsrechnung und in die induktive Statistik“ eine grundlegende Rolle. Sie wurden in Rahmen von Kolmogorovs „Grundbegriffe der Wahrscheinlichkeitsrechnung“ 1933 veröffentlicht. Dabei ist es kein Zufall, dass das hier behandelten Paper im gleichen Jahr entstand.

Andrei Nikolajewitsch Kolmogorov veröffentlichte ebenfalls 1933 im italienischen Journal „Giornale dell'Istituto Italiano degli Attuari“ einen kurzen, aber wichtigen Artikel über die empirische Bestimmung einer Verteilungsfunktion. Auf dessen Basis entstand eine Menge Literatur über wahrscheinlichkeitstheoretische Probleme und Tests auf Verteilung.
Grundsätzlich beschäftigt sich Kolmogorov in seiner Veröffentlichungen mit der Betrachtung der Abweichung der empirischen Verteilungsfunktion, die am Beginn seines Artikels definiert wird, mit der wahren, stetigen Verteilung F(x). Aus dieser resultiert die Teststatistik D, die Kolmogorov-Statistik oder auch Kolmogorov-Smirnov-Statistik genannt wird. Es wird gezeigt, dass D beliebig klein wird mit, wachsendem Stichprobenumfang n. Darüber hinaus wird eine Methode zur Berechnung der Verteilung von D an spezifischen Punkten gegeben. So wird die asymptotische Verteilung von D bestimmt.[1]

2. Leben

Am 25.04.1903 wurde Andrei Nikolajewitsch Kolmogorov im russischen Tambow geboren. Er wuchs bei seiner Tante auf, da seine Mutter kurz nach seiner Geburt gestorben war und sein Vater sich nicht um den Sohn kümmerte, da er sich im Exil befand. Schon während seiner Zeit am Gymnasium in Moskau interessierte er sich neben Mathematik auch für Biologie und russische Geschichte. Später beschäftigte sich Kolmogorov außerdem mit Dichtung und Erziehungswissenschaft. So arbeitete er während seines Mathematikstudiums, das er 1920 an der Moskauer Universität begann, auch als Lehrer.

1 M.A. Stephens: Introduction to Kolmogorov (1933) On the Empirical Determination of a Distribution, Introduction

Schnell erlangte der Student die Aufmerksamkeit der Professoren und begann damit, eigene Ergebnisse, zum Beispiel in der Mengentheorie oder Fourier-Analysis, zu veröffentlichen.
1925, wenige Jahre später, folgte die erste Veröffentlichung im Bereich der Wahrscheinlichkeitstheorie. Im selben Jahr erlangte er seinen Abschluss und begann ein postgraduales Studium. In dessen Verlauf veröffentliche er weitere Arbeiten und konnte somit bei Ende seines Studium bereits 20 Veröffentlichungen vorweisen.
Kolmogorov wurde nur 2 Jahre nachdem er 1929 Fakultätsmitglied geworden war zum Professor ernannt. 1933 beförderte man den jungen Mathematiker sogar zum Direktor des Wissenschafts- und Forschungsinstituts.
Während seiner Tätigkeit an der Universität beschäftigte er sich mit der Anwendung von Maßtheorie an Wahrscheinlichkeit, da er bestrebt war, diesbezüglich eine axiomatische Grundlage zu schaffen. Daraus gingen unter anderem die wichtige Veröffentlichungen „Grundbegriffe der Wahrscheinlichkeitsrechnung“ (1933) und „Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung“ (1933) hervor, in welcher er das Verhältnis von Wahrscheinlichkeitstheorie zu analytischen physikalischen Methoden beschreibt. Dies war für die spätere Entwicklung der Theorie der Zufallsprozesse maßgeblich. Auch das hier betrachtete Paper wurde 1933 veröffentlicht, als Kolmogorov mit seinen 30 Jahren bereits international bekannt geworden und auf der Höhe seiner mathematischen Fähigkeiten war.[2]

Während der Kriegszeit wurden viele der von Kolmogorov angestoßenen Entwicklungen auf Eis gelegt. Auch er selbst wurde in den Krieg involviert. So arbeitete er unter anderem an Artillerieproblemen und hatte daher mehr und mehr mit statistischen Fragestellungen zu tun, was ihn in seinen späteren Interessen beeinflusst haben könnte.[3]

3. Betrachtung des Papers „On the Empirical Determination of a Distribution Funktion“

Zuerst wird die empirische Verteilungsfunktion, die im Artikel „Empirical Distribution Funktion“ oder EDF genannt wird, folgendermaßen definiert. Es werden die *n* unabhängigen Beobachtungen einer Zufallsvariable *X* betrachtet. Diese werden nach ihrer Größe aufsteigend geordnet.

2 N. L. Johnson and S. Kotz (1991). Breakthroughs in Statistics, Seite 94, 2. A.N. Kolmogorov: Early Years and Position in 1933
3 N. L. Johnson and S. Kotz (1991). Breakthroughs in Statistics, Seite 100f, 5. The War and Afterwards, erster Absatz

Es wird die Verteilungsfunktion

$$F(x)=P\{X\leqslant x\}$$

definiert.

Des weiteren wird die empirische Verteilungsfunktion durch die folgenden Relationen definiert.

$$F_n(x)=0, \quad x<X_1;$$

$$F_n(x)=\frac{k}{n}, \quad X_k<x<X_{k+1}, k=1,2,\ldots,n-1;$$

$$F_n(x)=1, \quad X_n<x;$$

Somit entspricht $nF_n(x)$ der Anzahl der Werte X_k , die den Wert x nicht übersteigt. Kolmogorov stellt heraus, es liege auf der Hand zu hinterfragen, ob $F_n(x)$ approximativ gleich $F(x)$ wird, wenn n einen sehr großen Wert annehme.

Es wird außerdem darauf eingegangen, dass eine ähnliche Frage bereits vom österreichischen Mathematiker Richard von Mises 1930 betrachtet wurde. Der zugehörige Test ist unter dem Namen Cramer-von-Mises-Test bekannt.

Im Anschluss wird gezeigt, dass die Wahrscheinlichkeit für die Erfüllung der Ungleichung

$$D=\sup_x|F_n(x)-F(x)|<\epsilon$$

gegen 1 geht, falls $n\to\infty$, für alle ϵ .

Dazu benutzt er eine Formulierung, auf die er kurz zuvor von seinem Kollegen Waleri Iwanowitsch Gliwenko gestoßen wurde.[4]

An dieser Stelle wird das erste Theorem definiert:

Theorem 1:

Die Wahrscheinlichkeit $\Phi_n(\lambda)$ *der Ungleichung*

$$D=\sup_x|F_n(x)-F(x)|<\lambda\sqrt{n}$$

4 N. L. Johnson and S. Kotz (1991). Breakthroughs in Statistics, Seite 107

geht für $n \to \infty$ *gegen* $\Phi_n(\lambda) = \sum_{-\infty}^{\infty} (-1)^k e^{-2k^2\lambda^2}$ *(1)*

für jede stetige Verteilungsfunktion F(x).

Es wurden einige Werte von $\Phi_n(\lambda)$ von n. Kogeknikov bestimmt:

λ	$\Phi_n(\lambda)$	λ	$\Phi_n(\lambda)$	λ	$\Phi_n(\lambda)$
0,0	0,0000	1,0	0,7300	2,0	0,99932
0,2	0,0000	1,2	0,8877	2,2	0,99986
0,4	0,0028	1,4	0,9603	2,4	0,999973
0,6	0,1357	1,6	0,9880	2,6	0,9999964
0,8	0,4558	1,8	0,9969	2,8	0,99999966

Für $D \leqslant 2.4/\sqrt{n}$ wird von einer faktischen Sicherheit ausgegangen.
Danach wird eine asymptotische Formel für kleine λ angegeben, bei denen die obige Reihe sehr langsam konvergiert.[5]

$$\Phi(\lambda) \simeq \frac{\sqrt{2\pi}}{\lambda} \exp\left(-\frac{\pi^2}{8\lambda}\right)$$

Eine Gegenüberstellung der beiden Formeln wird anhand des Beispiels $\lambda = 0.6$ gegeben.
Es ergeben sich die Ergebnisse:

Für die approximative Formel: $\Phi(\lambda) \simeq 0.1327$
Obere, exakte Formel: $\Phi(\lambda) = 0.1357$

Eine weitere Möglichkeit der Approximation, welche für $\lambda < 0{,}6$ sehr gute Resultate liefert, lautet[6]:

$$\Phi(\lambda) \simeq \frac{\sqrt{2\pi}}{\lambda} \sum_{k=1}^{\infty} \left[\exp\left(-(2k-1)\frac{\pi^2}{8\lambda^2}\right) \right].$$

5 N. L. Johnson and S. Kotz (1991). Breakthroughs in Statistics, Seite 107
6 N. L. Johnson and S. Kotz (1991). Breakthroughs in Statistics, Seite 95

Lemma 2 wird definiert:

Die Wahrscheinlichkeitsfunktion $\Phi_n(\lambda)$ *ist unabhängig von der Verteilungsfunktion* $F(x)$, *unter der Bedingung der Stetigkeit.*

Beweis:

Sei X eine Zufallsvariable mit der stetigen Verteilung $F(x)$. Zu der Zufallsvariable $Y = F(x)$ existiert eine Verteilungsfunktion, so dass

$$F^{(0)}(x)=0,\ \ x\leqslant 0;$$

$$F^{(0)}(x)=x,\ \ 0\leqslant x\leqslant 1;$$

$$F^{(0)}(x)=0,\ \ x\geqslant 0;$$

Da $F_n(x)$ und $F_n^{(0)}(x)$ die empirischen Verteilungsfunktionen von X und Y nach n Beobachtungen darstellen, gelten die Gleichungen:

$$F_n(x)-F(x)=F_n^{(0)}[F(x)]-F^{(0)}[F(x)]=F_n^{(0)}(y)-F^{(0)}(y),$$

$$\sup_x|F_n(x)-F(x)|=\sup_x|F_n^{(0)}(y)-F_n(y)|$$

Somit folgt, dass die Wahrscheinlichkeitsfunktion $\Phi_n(\lambda)$, die mit einer willkürlichen stetigen Verteilungsfunktion $F(x)$ korrespondiert, mit der Funktion $F_n^{(0)}(x)$ identisch ist. Q.E.D.[7]

Im Anschluss wird das Theorem unter der Einschränkung auf $F(x) = F_n^{(0)}(x)$ analysiert[8]. Daher notiert Kolmogorov nun $F_n^{(0)}(x)$ als $F(x)$ und betrachtet den Bereich $0\leqslant x\leqslant 1$, wo $F_n^{(0)}(x)=x$ gilt. Dadurch vereinfacht sich das Problem auf die Bestimmung der Wahrscheinlichkeit $\Phi_n(\lambda)$, welche definiert ist durch:

7 N. L. Johnson and S. Kotz (1991). Breakthroughs in Statistics , Seite 108
8 N. L. Johnson and S. Kotz (1991). Breakthroughs in Statistics, Seite 108ff

$$D = \sup_x |F_n(x) - x| < \lambda \sqrt{n}, \qquad 0 \leqslant x \leqslant 1. \tag{2}$$

Für λ erfolgt der Vorschlag $\lambda = \mu / \sqrt{n}$, wobei μ einer ganzen Zahl entspricht.

Einsetzen in die obere Formel ergibt also:

$$\Phi_n(\lambda) = P\{ \sup_x |F_n(x) - x| < \mu / n \}, \quad \lambda = \mu / \sqrt{n}.$$

$F_n(x)$ nimmt nur Vielfache von $1/n$ an. Dazu gibt gibt Kolmogorov beispielhaft $F_n(x) = i/n$ und $x = j/n + \epsilon, (0 \leqslant \epsilon \leqslant 1/n)$.

Aus der Monotonie von $F_n(x)$ folgen:

$$F_n(x) - x = \frac{i-j}{n} - \epsilon,$$

$$F_n(\frac{i-j}{n}) - \frac{i}{j} \leqslant F_n(x) - (x - \epsilon) = \frac{i-j}{n},$$

$$F_n(\frac{j+1}{n}) - \frac{j+1}{n} \geqslant F_n(x) - (x + \frac{1}{n} - \epsilon) = \frac{i-j-1}{n}.$$

Für die Richtigkeit der Ungleichung

$$|F_n(x) - x| = |\frac{i-j}{m} - \epsilon| \geqslant \mu / n$$

ist es notwendig, dass mindestens eine der folgenden Ungleichungen wahr ist:

$$F_n(\frac{j}{n}) - \frac{j}{n} \leqslant \frac{i-j}{n} \leqslant \frac{-\mu}{n},$$

$$F_n(\frac{j+1}{n}) - \frac{j+1}{n} > \frac{i-j-1}{n} \geqslant \frac{\mu}{n}.$$

An dieser Stelle wird der Schluss gezogen, dass

$$\Phi_n(\lambda) = P\{\sup_x |F_n(x) - x| < \mu/n\}, \qquad \text{siehe oben}$$

durch

$$\Phi_n(\lambda) = P\{\max |F_n(i/n) - i/n| < \mu/n\},\ i=0,1,\ldots,n. \tag{4}$$

Kolmogorov definiert nun die Wahrscheinlichkeit P_{ik} für das Eintreten des Ereignisses E_{ik} wird definiert.

E_{ik} tritt ein, wenn gleichzeitig gelten:

$$|F_n(\frac{j}{n}) - \frac{j}{n}| < \frac{\mu}{n}, \quad j=0,1,2,\ldots,k;$$

$$|F_n(\frac{k}{n}) - \frac{k}{n}| = \frac{i}{n}. \tag{5}$$

Es folgt die Anmerkung:

$$\Phi_n(\lambda) = P_{0n}. \tag{6}$$

Außerdem ist werden weitere Fälle gegeben.

$$P_{00}=1,\ P_{i0,}\ \ (i\neq 0). \tag{7}$$

Im Allgemeinen gilt also

$$P_{ik}=0, \quad |i| \geq \mu, \tag{8}$$

da sich dann die beiden obigen Voraussetzungen für das Eintreten von E_{ik} widersprechen.

Man betrachte

$$P_{\mathrm{i,k+1}} = \sum_j P_{jk} Q_{ji}^{(k)}, \quad |i| < \mu, \tag{9}$$

wobei $Q_{ji}^{(k)}$ die Wahrscheinlichkeit, dass $E_{\mathrm{i,k+1}}$ auftritt, nachdem E_{ik} eingetreten ist, also, dass die Wahrscheinlichkeit, dass

$$F_n(\frac{k+1}{n}) - F_n(\frac{k}{n}) = \frac{i-j+1}{n} \tag{10}$$

gilt unter der Bedingung

$$F_n(\frac{k}{n}) = \frac{k-j}{n}. \tag{11}$$

Somit sind von den n Ergebnissen der Beobachtungen $X_1, X_2, \ldots, X_n$ genau $n-k-j$ Im Bereich $k/n < x \leqslant 1$.

(11) kann nur dann richtig sein, wenn i-j+1 der n-k-j oben genannten Beobachtungen sich im Intervall $k/n<x\leqslant(k+1)n$ befinden.

Die Wahrscheinlichkeit, dass dies der Fall ist, ist, unter der Annahme, dass X_m gleichverteilt, lautet.

$$Q_{ik}^{(k)}=\binom{n-k-j}{i-j+1}\cdot\left(1-\frac{1}{n-k}\right)^{n-k-i-1}\cdot\left(\frac{1}{n-k}\right)^{i-j+1}. \tag{12}$$

Die Formeln (7),(8),(9),(12), und (6) erlauben die Berechnung der Wahrscheinlichkeit $\Phi_n(\lambda)$, wobei $\lambda=\mu/\sqrt{n}.$

Nun ist es möglich diese Formeln mit praktischeren zu ersetzen. Daher wird vorgeschlagen:

$$R_{ik}=\frac{(n-k-i)!\,n^n}{(n-k)^{n-k-i}\,n!}e^{-k}\,P_{0n} \tag{13}$$

R_{ik} transformiert (7) und (8) also in

$$R_{00}=1,\ \ R_{i0}=0,\quad i\neq 0; \tag{14}$$

$$R_{ik}=0,|i|\geqslant\mu. \tag{15}$$

Die Relation (9) wird überführt in

$$R_{\mathrm{i,k+1}}=\sum_j R_{jk}\frac{1}{(i-j+1)!}e^{-1}, \tag{16}$$

während man für (6) und (13)

$$\Phi_n(\lambda)=\frac{n!}{n^n}e^n R_{0n} \tag{17}$$

erhält.

Man kann also $\Phi_n(\lambda)$ mit $\lambda=\mu/\sqrt{n}.$ auch aus (14), (15), (16) und (17) berechnen.

An dieser Stelle wird eine Folge von untereinander unabhängigen Zufallsvariablen $Y_1,Y_2,...,Y_n$ betrachtet. Zu diesen existiert eine Verteilungsfunktion, die durch die folgenden Formeln beschrieben wird.

$$P\left\{Y_k=\frac{i-1}{\mu}\right\}=\frac{1}{i!}e^{-1},\qquad i=0,1,2,... \tag{18}$$

Mit

$$S_k = Y_1 + Y_2 + ... + Y_k$$

kann garantiert werden, dass die Wahrscheinlichkeit $\bar{R}_{ik}$, dass die Beziehungen

$$|S_j| < \mu, \quad j = 0,1,2,...,k;$$
$$S_k = \frac{i}{\mu},$$

gleichzeitig erfüllt sind, die selben Bedingungen wie R_{ik} erfüllt, also (14), (15) und (16). Demnach resultiert $\bar{R}_{ik} = R_{ik}$.

Darüber hinaus wird darauf hingewiesen, dass möglich ist, für $n \to \infty$ einen asymptotischen Ausdruck für R_{ik}, zu dessen Bestimmung sich des nachfolgenden Theorems zu nutze gemacht.

Theorem 2:

Sei $Y_1, Y_2, ..., Y_n$ eine Folge von unabhängigen Zufallsvariablen, die lediglich Vielfache von ϵ

Seien außerdem

$$E(Y_k) = 0, \quad E(Y_k^2) = 2b_k, \quad E(|Y_k^3|) = d_k,$$
$$S_k = Y_1 + Y_2 + ... + Y_k,$$
$$t_k = b_1 + b_2 + ... + b_k$$

und a(t) und b(t) zwei stetige und differenzierbare Funktionen, welche den Ungleichungen

$$a(t) < b(t)$$
$$a(0) < 0 < b(t)$$

genügen.

Die Wahrscheinlichkeit R_{in}, dass

$$a(t_k) < S_k < b(t_k), \quad k = 1,2,..,n,$$

$$S_n = i\epsilon$$

gleichzeitig wahr sind, wird definiert.

Außerdem wird die Greensche Funktion $u(\sigma,\tau,s,t)$, mithilfe welcher man

$$R_{\text{in}}=\epsilon\cdot\{u(0,0,i\epsilon,t_n)+\Delta\}$$

erhält, definiert, wobei Δ gleichmäßig mit ϵ gegen null geht, falls folgende Bedingungen erfüllt sind:

(i) a(t) und b(t) bleiben für konstantes n zwischen den festen Grenzen $0 < T_1 < t_n < T_2$;

(ii) es existiert eine Konstante $C > 0$ mit $d_k/b_k \leqslant C\epsilon$;

(iii) es existiert eine Konstante $K > 0$, so dass für jedes k und ein passend gewähltes i_k die folgenden Ungleichungen erfüllt sind:

$$P\{Y_k=i_k\cdot\epsilon\} > K, \qquad P\{Y_k=(i_k+1)\cdot\epsilon\} > K;$$

(iv) es existiert eine Konstante A, so dass

$$a(t_n)\ A < i\epsilon < b(t_n) - A.$$

Kolmogorov weist darauf hin, dass ein ähnliches Problem wie im Theorem 2 bereits in seiner 1931 veröffentlichten Schrift „Eine Verallgemeinerung des Laplace-Liapounoffischen Satzes“ behandelt wurde.[9]

Dort wurde die Betrachtung

$$\sum_{i=p}^{i=q} R_{\text{in}}=\int_{p\epsilon}^{q\epsilon} u(0,0,z,t_n)\,dz+\Delta',$$

wobei Δ' eine Infinitesimalzahl mit ϵ, falls die Bedingungen (i) und (ii) erfüllt sind.

$$\epsilon=1/\mu,$$
$$E(Y)=0, E(Y_k^2)=2b_k=1/\mu^2, E(|Y_k^3|)=d_k=C/\mu^3,$$
$$d_k/b_k=c/\mu=C\epsilon, t_n=n/(2\mu^2)=1/(2\lambda^2),$$
$$a(t)=-1, b(t)=+1,$$
$$R_{0n}=\epsilon\cdot\{u(0,0,0,1/(2\lambda^2))+\Delta\},$$

9 N. L. Johnson and S. Kotz (1991). Breakthroughs in Statistics, 112 unten

$$u(0,0,s,t)=\frac{1}{2\sqrt{\pi t}}\sum_{-\infty}^{\infty}(-1)^k e^{-(s-2k)^2/(4t)},$$

$$\Theta_n(\lambda)=P_{0n}=\frac{n!}{n^n}e^n R_{0n} \quad = \quad \{\sqrt{2\pi n}+\delta\cdot 1/\mu\cdot\left\{\frac{\lambda}{\sqrt{2\pi}}\sum_{-\infty}^{\infty}(-1)^k e^{-2k^2\lambda^2}+\Delta\right\},$$

$$a=\sum_{-\infty}^{\infty} d(-1)^k e^{-2k^2\lambda^2}+R \quad = \quad \Theta(\lambda)+R.$$

Der Restterm R wird hier für den Fall, dass λ größer als ein vorgegebenes $\lambda_0 > 0$ ist, gegen null, wenn $n\to\infty$. In der Tat gilt also $\epsilon=1/\mu=1/(\lambda\sqrt{n})$ und dies geht für $n\to\infty$ gegen 0.

Für Werte λ der Form $\mu/\sqrt{n}$ und unter der Bedingung $\lambda > \lambda_0$.

4. Einfluss des Papers und Gegenüberstellung mit anderen Methoden

4.1 Zeitgenössische Entwicklungen

Obwohl Kolmogorov in seiner Veröffentlichung die Annäherung von $F_n(x)$ an $F(x)$ betrachtete, schlug er nicht vor, dass die empirische Verteilungsfunktion für einen Test, ob $F(x)$ die tatsächliche Verteilung von x ist, geeignet sein könnte. Genau diese Betrachtung wurde jedoch eine der wichtigsten Resultate, die aus dem Paper hervorgingen.
Bereits in den Jahren, in denen in denen sich Kolmogorov mit der Thematik beschäftigte, wurden einige Vorschlage zur Lösung des Problems gemacht. Schon 1928 hatte Harald Cramer eine Methode vorgestellt mit den Teststatistiken[10]

$$I_j = \int \{\Delta_j(x)\}^2\, dx,$$

wobei $\Delta_j(x)=F_n(x)-\hat{F}_j(x)$ und $\hat{F}_j(x)$ der Erweiterung von $F(x)$ um den j-ten Term entspricht. Folglich kann $\Delta_j(x)$ als j-te Komponente als die Differenz $F_n(x)-F(x)$ interpretiert werden. Auf dieser Betrachtung beruhen die wenige Jahre später von Neyman angestellten Überlegungen bezüglich glatten Tests.

10 N. L. Johnson and S. Kotz (1991). Breakthroughs in Statistics, 97

1931 wurde vom österreichischen Mathematiker von Mises ein Test auf Grundlage der Teststatistik

$$\omega = n\int \lambda(x)[F_n(x)-F(x)]^2\,d\,x$$

vorgeschlagen. Dabei entspricht $\lambda(x)$ der passend gewählten Gewichtsfunktion. Diese wird als Konstante angenommen, welche derart gewählt wird, dass $E(\omega^2)=1.$ Damit ergibt sich eine Berechnungsformel für ω^2. Die Verteilung von ω^2 variiert in $F(x)$ und $\lambda(x)$, auch bei kompletter Spezifizierung. Von Mises postulierte zwar keine Theorie über die Verteilung, ging jedoch auf Varianzen im Fall von Gleich- und Normalverteilung ein.
Einige Jahre nach von Mises veröffentlichte der sowjetische Mathematiker Vladimir Ivanowitsch Smirnov eine alterierte Variante der Definition von ω^2 :

$$\omega = n\int \lambda(F(x))[F_n(x)-F(x)]^2 dF(x)$$

Es wird hier also ein Lesbegue-Integral betrachtet und der Wert der Teststatistik basiert hier auf den Werten von $Z_i=F(X_i),$ welche gleichverteilt zwischen 0 und 1, und ist daher verteilungsfrei sind. Er hängt also nicht von der wahren Verteilung $F(x)$ ab. Diese Version der Statistik, bei der gilt, dass $\lambda(F(x))=1,$ erlangte Bekanntheit unter dem Namen Cramer-von-Mises Statistik W^2.
Smirnov erweiterte Kolmogorovs Arbeit um ein- und zweiseitige Tests.
Dazu werden

$$D^+=\sup_x\{F_n(x)-F(x)\}$$

und

$$D^-=\sup_x\{F(x)-F_n(x)\}$$

definiert.[11]
Diese haben die von Smirnov vorgeschlagene, asymptotische Verteilung

$$\lim_{n\to\infty} P(\sqrt{n}\,D^+<\lambda)=1-e^{-2\lambda^2}.$$

11 N. L. Johnson and S. Kotz (1991). Breakthroughs in Statistics, Seite 98

Seien $F_n(x)$ und $G_m(x)$ empirische Verteilungsfunktionen zweier unabhängiger Zufallsstichproben mit n und m Beobachtungen. Sei $N = mn/(m+n)$ und

$D^{+}_{\mathrm{m,n}} = \sup_x \{F_n(x) - G_m(x)\}$, $D^{-}_{\mathrm{m,n}} = \sup_x \{G_m(x) - F_n(x)\}$, sowie

$D_{\mathrm{m,n}} = \sup_x |F_n(x) - G_m(x)|$.

Smirnov zeigt, dass $\sqrt{N}\, D_{\mathrm{m,n}}$ die gleiche asymptotische Verteilung besitzt wie $\sqrt{n}D$ im ersten Theorem in Kolmogorovs Veröffentlichung. Die Statistiken der Form $D^{+}, D^{-} und D$ werden auch Kolmogorov-Smirnov Statistiken genannt.[12]

4.2 Der Kolmogorov-Smirnov-Test als eines der wichtigsten Ergebnisse

Jetzt sollte auf die Thematik des Kolmogorov-Smirnov-Anpassungstest näher eingegangen werden, da dieser die Quintessenz aus dem Paper darstellt und für die Anwendung eine zentrale Rolle spielt. Der Kolmogorov-Smirnov-Test überprüft also, ob die Beobachtungen $x_1, x_2, \ldots, x_n$ einer gewissen Verteilung entsprechen. Die empirische Verteilungsfunktion kann definiert werden, durch

$$E_n = n(i)/n.$$

Bei dieser Betrachtung entspricht $n(i)$ der Anzahl der Beobachtungen, welche x_i nicht übersteigen. Deswegen entspricht die Verteilungsfunktion einer Stufenfunktion, die mit einer Stufenhöhe von $1/n$ an den beobachteten Werten ansteigt.[13]

Der maximale Abstand zwischen empirischer und echter Verteilungsfunktion bildet die Grundlage für den Test. Auch wenn die Teststatistik exakt bestimmt werden kann und sie den großen Vorteil besitzt, dass sie unabhängig von der beobachteten Verteilung ist, müssen jedoch die Einschränkungen gemacht werden, dass nur stetige und vollständig spezifizierte Verteilungen verwendet werden dürfen. Falls Erwartungswert und Varianz nicht bekannt sind, sondern aus der Empirik geschätzt werden müssen, so sind liegen keine exakten kritischen Werte vor und die Ablehnung der Nullhypotese erfolgt erst bei größeren Werten der Teststatistik. Außerdem ist der Test in der Mitte der Verteilung sensitiver als an deren Rändern.

12 N. L. Johnson and S. Kotz (1991). Breakthroughs in Statistics, Seite 98
13 L. Sachs, J. Hedderich, Angewandte Statistik (2006): 7.2.5 Kolmogoroff-Smirnoff-Anpassungstest Seite 337

Beim der Kolmogorov-Smirnov-Test nimmt man unter der Nullhypothese an, dass die beobachteten Daten der vorgegebenen Verteilung folgen, während diese bei gegenteiliger Situation abgelehnt würde.
Im Statistikprogramm R existiert die Funktion ks.test(), mithilfe derer eine Überprüfung des Verteilungsmodelles vorgenommen werden.
Außerdem können damit zwei unabhängige Stichproben jeder Verteilungsform gegenübergestellt werden.[14]

4.3 Vergleich von Cramer-von-Mises- und Kolmogorov-Smirnov Test anhand eines Zweistichprobenproblems

In diesem Abschnitt sollen die oben genannten Statistiken gegenübergestellt werden. Die Frage die sich stellt, lautet, ob sie aus der selben Grundgesamtheit stammen und kann durch den Kolmogorov-Smirnov-Test am trennschärfsten beantwortet werden. Für die beiden Verteilungsfunktionen $\hat{F}_1$ und $\hat{F}_2$ wird die Prüfgröße

$$\hat{D} = max\left|(\hat{F}_1 - \hat{F}_2)\right|$$

definiert, welche für genügend große Stichprobenumfänge $(n_1 + n_2 > 35)$ auch durch

$$D_\alpha = K_{(\alpha)}\sqrt{(n_1 + n_2)/(n_1 \cdot n_2)}$$

approximiert werden.[15]

$K_{(\alpha)}$ Entspricht einer von dem α -Niveau abhängigen Konstante.

α	0,20	0,15	0,10	0,05	0,01	0,001
$K_{(\alpha)}$	1,07	1,14	1,22	1,36	1,63	1,95

Ausgewählte Konstanten für den Kolmogorov-Smirnov Test;[16]

14 L. Sachs, J. Hedderich, Angewandte Statistik (2006): 7.2.5 Kolmogoroff-Smirnoff-Anpassungstest Seite 337ff
15 L. Sachs, J. Hedderich, Angewandte Statistik (2006): Seiten 405f
16 L. Sachs, J. Hedderich, Angewandte Statistik (2006): Tabelle 7.31

Falls einer der aus den beiden Stichproben ermittelten Werte $\hat{D}$ den kritischen Wert D_α annimmt oder übersteigt, so unterscheiden sich die Verteilungsfunktionen signifikant.

$n_1 =$	6				9				12			
$n_2 =$	6	12	18	24	9	12	15	18	12	16	18	20
$D_{0,10}$	0,667	0,583	0,556	0,500	0,556	0,500	0,489	0,444	0,417	0,438	0,417	0,417
$D_{0,05}$	0,667	0,583	0,611	0,583	0,556	0,556	0,533	0,500	0,500	0,479	0,472	0,467
$D_{0,01}$	0,833	0,750	0,722	0,667	0,667	0,667	0,644	0,611	0,583	0,583	0,556	0,567

$n_1 = n_2 =$	7	8	10	11	13	14	15	16	17	18	19	20
$D_{0,10}$	0,571	0,500	0,500	0,454	0,462	0,429	0,400	0,375	0,412	0,389	0,368	0,350
$D_{0,05}$	0,714	0,625	0,600	0,545	0,462	0,500	0,467	0,438	0,412	0,444	0,421	0,400
D_{001}	0,714	0,750	0,700	0,636	0,615	0,571	0,533	0,563	0,529	0,500	0,473	0,500

Einige Werte $D_{n1;n1;\alpha}$ für die zweiseitige Fragestellung; siehe L. Sachs, J. Hedderich (2006): Seite 405; Tabelle 7.32

$$\hat{D} = max_x \left| \hat{F}_{1,n1}(x) - \hat{F}_{2,n2}(x) \right| = max_x \left| \hat{P}(X_1 \leqslant x) - \hat{P}(X_2 \leqslant x) \right|$$

Nun sollen 2 unterschiedliche Messreihen verglichen werden. Dazu wird die Nullhypothese, dass die beiden Verteilungsfunktionen gleich sind gegen die Alternativhypothese, dass diese sich unterscheiden zum Konfidenzniveau $\alpha = 0{,}05$ getestet werden. Wir betrachten einen zweiseitigen Test.

Messreihe 1: 2,1 3,0 1,2 2,9 0,6 2,8 1,6 1,7 3,2 1,7

Messreihe 2: 3,2 3,8 2,1 7,2 2,3 3,5 3,0 3,1 4,6 3,2

Die 10 Messwerte der Reihen werden der Größe nach geordnet:

Messreihe 1: 0,6 1,2 1,6 1,7 1,7 2,1 2,8 2,9 3,0 3,2

Messreihe 2: 2,1 2,3 3,0 3,1 3,2 3,2 3,5 3,8 4,6 7,2

Die Werte der Teststatistik aus dem Beispiel werden veranschaulicht[17]

Bereich	0,0-0,9	1,0-1,9	2,0-2,9	3,0-3,9	4,0-4,9	5,0-5,9	6,0-6,9	7,0-7,9
f_1	1	4	3	2	0	0	0	0
f_2	0	0	2	6	1	0	0	1
$\hat{F}_1$	1/10	5/10	8/10	10/10	10/10	10/10	10/10	10/10
$\hat{F}_2$	0/10	0/10	2/10	8/10	9/10	9/10	9/10	10/10
$\hat{F}_1-\hat{F}_2$	1/10	5/10	6/10	2/10	1/10	1/10	1/10	0

Die absolut größte Differenz ist hier bei $\hat{D}=6/10,$ was nach der oberen Tabelle den kritischen Wert $D_{10;10;0,05}=0{,}6000$ erreicht, weswegen die Homogenitätshypothese auf dem 5%-Niveau abgelehnt wird. Somit wird nicht angenommen, dass eine einzige Verteilung den beiden Verteilungen zugrunde liegt.

Nun wird der zugehörige Test mit R durchgeführt. Dazu wird die Funktion ks.test().

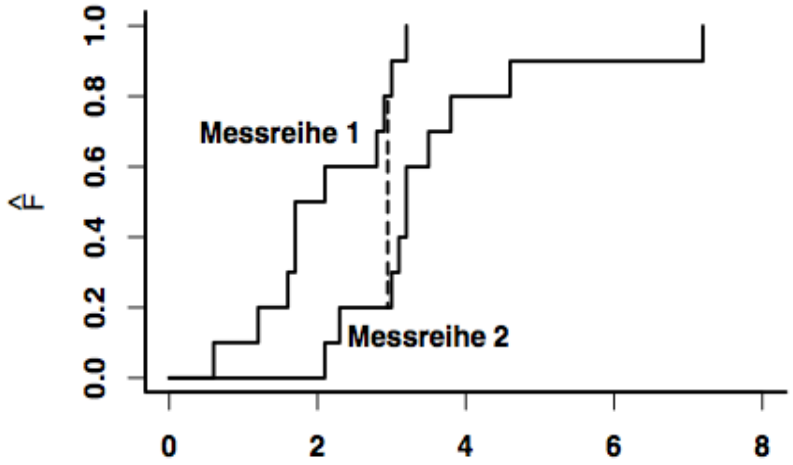
R-Code:

```
> m1<− c(2.1, 3.0, 1.2, 2.9, 0.6, 2.8, 1.6, 1.7, 3.2, 1.7)
> m2<− c(3.2, 3.8, 2.1, 7.2, 2.3, 3.5, 3.0, 3.1, 4.6, 3.2)
> ks.test(m1, m2, alternative="two.sided")

        Two-sample Kolmogorov-Smirnov test

data:  m1 and m2
D = 0.6, p-value = 0.05465
alternative hypothesis: two-sided

Warning message:
In ks.test(m1, m2, alternative = "two.sided") :
  cannot compute correct p-values with ties
```

Das Ergebnis D = 0,6 entspricht dem aus der oberen Tabelle. Allerdings tritt hier im Output eine Warnung auf, wonach beim Vorliegen von Bindungen keine exakten P-Werte berechnet werden können. Dieses Problem tritt bei n < 1000 auf. Das Verwenden der Werte aus der obigen Tabelle ist also insbesondere bei kleinen Stichprobenumfängen besser.[18]

17 L. Sachs, J. Hedderich, Angewandte Statistik (2006): Seite 407, Tabelle 7.33

18 L. Sachs, J. Hedderich, Angewandte Statistik (2006): 7.4.8 Vergleich zweier unabha¨ngiger Stichproben nach Kolmogoroff und Smirnoff Seiten 405-407



Der größte Abstand zwischen den beiden kumulierten Häufigkeitsverteilungen; siehe L. Sachs, J. Hedderich (2006): Seite 408; Tabelle 7.15

Im folgenden soll nun der Vergleich zum Cramer-von-Mises Test angestellt werden. Beim Zweistichprobenfall basiert die Testgröße auf der Summe der quadrierten Differenzen zwischen den beiden empirischen Verteilungsfunktionen. Sie wird definiert durch

$$\hat{C} = \frac{n_1 \cdot n_2}{(n_1+n_2)^2} \sum_{i=1}^{n_1+n_2} D_i^2$$

$$= \frac{n_1 \cdot n_2}{(n_1+n_2)^2} \left\{ \sum_{i=1}^{n_1} (F(x_i)-G(x_i))^2 + \sum_{j=1}^{n_2} (F(y_i)-G(y_i))^2 \right\}.$$

Die asymptotischen oberen Schranken C_α sind in der Tabelle gegeben:

α	0,30	0,20	0,10	0,05	0,01	0,001
C_α	0,184	0,241	0,347	0,461	0,743	1,168

In R soll mit den beiden Messreihen aus der vorigen Berechnung ein zweiseitiger Carmer-von-Mises Test für $\alpha = 0{,}05$ mit identischen Hypothesen durchgeführt werden.

R-Code:

```
> m1<− c(0.6, 1.2, 1.6, 1.7, 1.7, 2.1, 2.8, 2.9, 3.0, 3.2)
> m2<− c(2.1, 2.3, 3.0, 3.1, 3.2, 3.2, 3.5, 3.8, 4.6, 7.2)
> n1<−10; n2<−10;
> x<−seq(0, 8, by=0.1)
> hm1 <−hist (m1, breaks=x, plot=F); F <− cumsum(hm1$counts ) /n1
> hm2 <−hist(m2, breaks=x, plot=F); G<− cumsum(hm2$counts)/n2
> CM <− (n1∗n2) / (n1+n2)^2 ∗sum((hm1$counts+hm2$counts )∗((F−G)^2));
>CM
[1] 0.875
```

Dazu werden, wie bereits angesprochen die beiden Messreihen, die bereits beim Kolmogorov-Smirnov-Test verwendet wurden, verglichen. Der Wert der Teststatistik entspricht in diesem Fall, wie im R-Code abzulesen, $CM = 0{,}875$. Da $CM > 0{,}461 = C_{0,05}$ aus der wird hier die Nullhypothese, wie schon zuvor beim Kolmogorov-Smirnov-Test, abgelehnt.[19]

Generell besteht bei Kolmogorov-Smirnov-Statistiken das Problem der Bestimmung von Asymptoten von D, im Fall dass Parameter von $F(x;\theta)$ geschätzt werden müssen. Beim Pearson X^2 -Test besteht dieses Problem nicht, da sich die Freiheitsgrade der asymptotischen Verteilung X^2 einfach ändern. Bei D hängt die Verteilungstheorie aber vom getesteten $F(x;\theta)$ ab. Auch effiziente Methoden zur Schätzung der unbekannten Parameter, wie die ML-Methode, schaffen hier keine Abhilfe.[20]

4.4 Test-Power im Vergleich

Die Power des Kolmogorov-Smirnov Tests liegt in der Regel zwischen der des X^2 und Cramer-von-Mieses Test. Die Überlegenheit gegenüber dem X^2- Test resultiert unter anderem daraus, dass dieser durch die Gruppierung der Daten an Information verliert. Speziell bei kleinem Stichprobenumfang n ist D vorzuziehen.

Im Vergleich zu Cramer-von-Mises Statistiken ist wiederum von Gegenteiligem auszugehen. Hier wird nämlich der Vergleich von $F_n(x)$ und $F(x)$ für alle Werte von x angestellt, während beim Kolmogorov-Smirnov-Statistiken nur einzelne Punkte betrachtet werden. Wenn die Differenzen $F_n(x) - F(x)$ hauptsächlich positiv oder negativ sind, können die einseitigen D^+ und D^- eine sehr hohe Power aufweisen.[21]

Ein weiterer Vorteil von D ist, dass sich die Möglichkeit bietet, zu $F_n(x)$ eine Konstante zu Addieren und Subtrahieren, um ein Konfidenzintervall für $F(x)$ zu erhalten. Gerade für heute übliche, graphische Darstellungen ist dies sehr praktisch. [22]

19 L. Sachs, J. Hedderich, Angewandte Statistik (2006): 7.4.8 Vergleich zweier unabhängiger Stichproben nach Kolmogoroff und Smirnoff Seiten 408f

20 M.A. Stephens: Introduction to Kolmogorov (1933) On the Empirical Determination of a Distribution, 6. The Problem of Unknown Parameters

21 N. L. Johnson and S. Kotz (1991). Breakthroughs in Statistics, Seite 102, 8. Power

22 N. L. Johnson and S. Kotz (1991). Breakthroughs in Statistics, Seite 102, 9. Concluding Remarks

4.5. Weitere Entwicklungen

Die Kolmogorov-Smirnov Tests wurden für sowohl rechts- als auch linkszensierte Daten, wobei meist D benutzt wird. Nach dem Zufall zensierte Daten werden beispielsweise in der Survival-Analyse betrachtet. Bei dieser Datenlage wird oft der Kaplan-Meier Schätzer für $F(x)$ verwendet.[23]

Kolmogorovs Paper legt die Grundlage für die Benutzung der empirischen Verteilungsfunktion $F_n(x)$ als Schätzer für $F(x)$. Sein Artikel war der erste, der eine Statistik, die unter der Annahme der Nullhypothese, die nicht von der getesteten Verteilung $F(x)$ abhängen. Ebenso war es die erste, dessen asymptotischen Verteilung tabelliert werden kann. Die Findung der Verteilung für finite Stichproben wurde in dem Artikel ebenfalls gegeben. Das Interesse an den von Kolmogorov begründeten Überlegungen ist nach wie vor überwältigend, vor allem im Zuge der modernen Bootstrap-Techniken, die von den immer schneller werdenden Computern möglich gemacht werden.[24]

5. Fazit

23 N. L. Johnson and S. Kotz (1991). Breakthroughs in Statistics, Seite 101, 7.Further Developments
24 N. L. Johnson and S. Kotz (1991). Breakthroughs in Statistics, Seite 103

6. Literaturverzeichnis

N. L. Johnson and S. Kotz (1991): Breakthroughs in Statistics 1890-1989. Vol. 1-2. Springer.

A.N. Kolmogorov(1933): On the Empirical Determination of a Distribution

L. Sachs, J. Hedderich(2006):Angewandte Statistik: Methodensammlung mit R