

247. Statistische Beurteilung von Koinzidenzen bei Schwingungsspektren

von E. Billeter und Hs. H. Günthard

(22. VIII. 59)

1. Einleitung

Beim Vergleich von Schwingungsspektren hat man oft die Signifikanz der Anzahl von Koinzidenzen zu beurteilen. Fragen dieser Art treten etwa auf beim Vergleich von Infrarot- und RAMAN-Spektrum einer Molekel im gleichen Aggregatzustand oder beim Vergleich der Schwingungsspektren einer Substanz in verschiedenen Aggregatzuständen. Im ersten Fall bildet die Diskussion der Anzahl der Koinzidenzen den Ausgangspunkt für Rückschlüsse auf die Existenz eines Symmetriezentrums in der Symmetriegruppe der Molekel. Im zweiten Fall versucht man, aus dem Vorhandensein oder dem Fehlen von Koinzidenzen die Existenz gleicher bzw. verschiedener Konstellationen in verschiedenen Aggregatzuständen abzuleiten. Auch bei Infrarotspektren komplizierterer Molekeln hat man oft die Zugehörigkeit von Absorptionsbanden zu empirisch festgestellten Frequenzbändern zu beurteilen.

Um Korrelationen dieser Art zwischen Schwingungsspektren statistisch auf Grund gewisser Hypothesen festzustellen, lassen sich verschiedenartige Tests benützen.

Am naheliegendsten scheinen Tests, die die auf Korrelation zu prüfenden Merkmale x bzw. y in zwei Spektren bewerten, z. B. das Vorhandensein bzw. Fehlen mit 1 bzw. -1 , und in der üblichen Weise den Cross-Korrelationskoeffizienten

$$r = [\sum x_i y_i] [(\sum x_i^2) (\sum y_i^2)]^{-1/2}$$

als Mass für die Korrelation definieren¹⁾. Ferner ist die Anwendung von Tests für die Auffindung von Rangkorrelationen nach KENDALL²⁾ oder SPEARMAN³⁾ denkbar.

Da hierbei eine Rangordnung (z. B. nach der Grösse der Frequenz) benützt werden muss, scheinen uns diese Tests weniger natürlich. Zudem werden sie in gewissen Fällen unbestimmt.

2. K-Test

In dieser Arbeit beschreiben wir einen Test, welcher auf folgenden Annahmen beruht.

1. Die Beobachtbarkeit eines Schwingungsübergangs mittels einer bestimmten experimentellen Anordnung darf als Merkmal des Übergangs betrachtet werden.

Es wird also angenommen, dass sowohl Infrarot- wie RAMAN-Aktivität eines Schwingungsübergangs je ein Merkmal sind, oder dass die Infrarotabsorptionen ein und desselben Schwingungsüberganges einer Molekel in verschiedenen Aggregatzuständen beobachtbare Merkmale sind. Praktisch setzt dies voraus: ein Schwingungs-

¹⁾ Einen Test dieser Art benützten Hs. H. GÜNTARD, H. HEUSSER & A. FÜRST, *Helv.* **36**, 1900 (1953), bei Infrarotspektren komplexer Molekeln mit Bewertung 1 bzw. 0. Diese Bewertung verletzt jedoch eine zu fordernde «common sense»-Eigenschaft des Korrelationskoeffizienten, nämlich $-1 \leq r \leq +1$.

²⁾ M. G. KENDALL, *Rank Correlation Methods*, Griffin & Co., London 1955.

³⁾ C. SPEARMAN, *Amer. J. Psychol.* **15**, 88 (1904).

übergang kann anhand der Frequenz identifiziert werden. Die Genauigkeit der Frequenzmessung der verschiedenen experimentellen Anordnungen erhält somit bei der Identifikation von Schwingungsübergängen einen starken Einfluss auf die statistische Folge, auf welche der Test angewendet werden soll. Werden verschiedene Phasen verglichen, muss der Effekt des Aggregatzustandes auf die Auswahlregeln und die Frequenz der Übergänge berücksichtigt werden. Spezielle Vorsicht wird beim Vorhandensein von intermolekularen H-Brücken und stark polaren Gruppen geboten sein.

2. Ein Schwingungsübergang gelte als beobachtet, wenn er mittels irgendeiner der beim Test benützten experimentellen Anordnungen beobachtet wurde (Anzahl der dem Test zugrunde gelegten Schwingungsübergänge ist Vereinigungsmenge der in den verschiedenen experimentellen Anordnungen beobachteten und eventuell identifizierten Übergänge). Als wichtige Konsequenz dieser Annahme erwähnen wir die Willkür in der Wahl des Frequenzbereichs, in welchem man den Test anwenden will. Bei vielen Fragen wird man sich, auch zufolge Rücksicht auf Annahme 1, auf das Fingerprintgebiet beschränken. So werden etwa im Frequenzband 1370–4000 cm^{-1} in sehr vielen Fällen Koinzidenzen zwischen Infrarot- und RAMAN-Spektren festgestellt werden, deren Interpretation zweifelhaft ist⁴⁾.

3. Dem Test sei als Hypothese zugrunde gelegt: die in verschiedenen experimentellen Anordnungen festgestellten Merkmale seien zufällige Eigenschaften der beobachteten Schwingungsübergänge.

Sind also in zwei verschiedenartigen Experimenten n_1 bzw. n_2 Schwingungsübergänge festgestellt worden, so besitzen die Fälle 0, 1, 2, 3, . . . Koinzidenzen eine *a priori* Wahrscheinlichkeit proportional zur Zahl der Permutationen von n_1-k Elementen der Art 1 (z. B. RAMAN-inaktive Übergänge) n_2-k Elementen der Art 2 (z. B. infrarot-inaktive Übergänge) und k Elementen der Art 3 (z. B. infrarot- und RAMAN-aktive Übergänge).

4. Sind in zwei Experimenten n_1 bzw. n_2 Schwingungsübergänge beobachtet worden, folgt aus den Annahmen 1, 2 und 3 zunächst als Wahrscheinlichkeit für das Auftreten von k Koinzidenzen

$$p(k) = P(k) / \sum_{k=0}^{\min(n_1, n_2)} P(k) \quad (2-1)$$

$$\text{mit } P(k) = \frac{(n_1 + n_2 - k)!}{k!(n_1 - k)!(n_2 - k)!} = \binom{n_1 + n_2}{n_1} \binom{n_2}{k} / \binom{n_1 + n_2}{k}. \quad (2-2)$$

Die diskrete Dichte (1) kann mittels der Erzeugenden

$$f(x) = \sum_0^{n_2} p(k)x^k = \sum_0^{n_2} \frac{\binom{n_1}{k} \binom{n_2}{k}}{\binom{n_1 + n_2}{k}} x^k / \sum_0^{n_2} \frac{\binom{n_1}{k} \binom{n_2}{k}}{\binom{n_1 + n_2}{k}} \\ = F(-n_1, -n_2; -n_1 - n_2; -x) / F(-n_1, -n_2; -n_1 - n_2; -1) \quad (2-3)$$

⁴⁾ Dieser Test wurde von E. BILLETER & HS. H. GÜNTARD, *Helv.* **41**, 686 (1958), zur Beurteilung von Koinzidenzen zwischen Infrarot- und RAMAN-Spektrum von flüssigem Cyclodecan benützt. Infolge eines Versehens wurde in dieser Arbeit der 1. Teil der Bedingung (3-7') nicht angeführt, wohl aber verwendet.

⁵⁾ Für Definitionen und Formeln siehe z. B. W. MAGNUS & F. OBERHETTINGER, *Formeln und Sätze*, Springer, 1948.

⁶⁾ Siehe z. B. W. FELLER, *Probability Theory*, Wiley, New York (1950).

dargestellt werden, wo $F(a, b; c; x)$ die hypergeometrische Funktion⁵⁾ ist und für das folgende $n_2 \leq n_1$ vorausgesetzt wird. Aus (3) folgt für die Momente der Verteilung⁶⁾

$$M_1 = \langle k \rangle = f'(1) = \frac{\sum k P(k)}{\sum P(k)} \quad (2-4)$$

$$= \frac{[n_1 n_2 / (n_1 + n_2)] \cdot F(-n_1 + 1, -n_2 + 1; -n_1 - n_2 + 1; -1) / F(-n_1, -n_2; -n_1 - n_2; -1)}$$

$$M_2 = \langle k^2 \rangle = f'(1) + f''(1) = \frac{\sum k^2 P(k)}{\sum P(k)}$$

$$M_3 = \langle k^3 \rangle = f'(1) + 3 f''(1) + f'''(1) = \frac{\sum k^3 P(k)}{\sum P(k)}$$

Man erhält hieraus die zentrierten Momente

$$\mu_1 = \langle k - \langle k \rangle \rangle = 0,$$

$$\mu_2 = \sigma^2 = \langle (k - \langle k \rangle)^2 \rangle = M_2 - M_1^2 \text{ (Varianz),}$$

$$\mu_3 = \langle (k - \langle k \rangle)^3 \rangle = M_3 - 3 M_1 M_2 + 2 M_1^3 \text{ (Schiefe).} \quad (2-5)$$

Es gibt verschiedene Möglichkeiten, die Verteilung (1) asymptotisch zu approximieren (s. Anhang 5). Die von uns gefundenen Approximationen erwiesen sich für die Bestimmung der Momente bei n -Werten ≤ 50 als zu ungenau. Wie die numerische Berechnung einiger Verteilungen mit n -Werten der Grösse 10–50 zeigten, sind die Verteilungen (1) nahezu GAUSS'sche Fehlerkurven mit Mittelpunkt $\langle k \rangle$ und Dispersion $\sigma = \sqrt{\mu_2}$.

In Fig. 1, 2 und 3 sind die Resultate der numerischen Berechnung von $\langle k \rangle$, σ , μ_3 für $n_1 = 10$ (5) 50, $n_2 = 5$ (5) n_1 graphisch dargestellt; man sieht daraus, dass stets $0 < \mu_3 \lesssim \mu_2/4$. Die Verteilungen werden gut durch Fehlerkurven approximiert:

$$p(k) \approx w(k) = (\sigma \sqrt{2\pi})^{-1} \exp[-(k - \langle k \rangle)^2 / 2\sigma^2] \quad (2-6)$$

die Fehler $\frac{|p(k) - w(k)|}{p(k)}$ sind meist kleiner als 10% (s.w.u.).

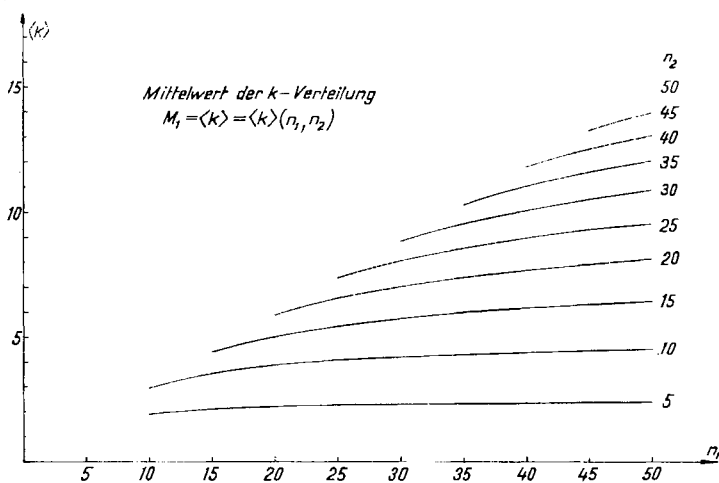


Fig. 1

3. Prüfung der Hypothese

Zur Prüfung der Hypothese 2.3 stehen verschiedene Möglichkeiten offen:

3.1. Die Hypothese ist zu verwerfen, wenn die Zahl K der beobachteten Koinzidenzen den Bedingungen genügt

$$K > k, w(K) \leq S_\sigma, \quad (3-7')$$

wo S_σ eine feste Schranke bedeutet.

3.2. Die Hypothese ist falsch, wenn K die Bedingung erfüllt

$$\sum_{k=K}^{n_2} p(k) < T_\sigma, \tag{3-7''}$$

wo T_σ eine feste Zahl $0 < T < 1$ bedeutet.

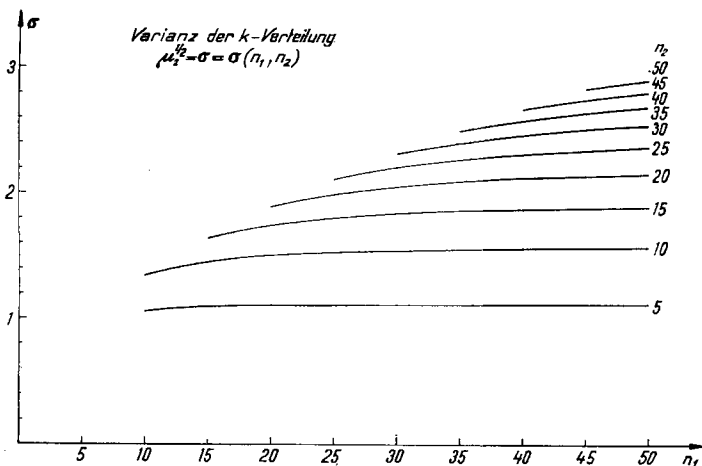


Fig. 2

Approximiert man in 1. Näherung die Summe in (3-7'') durch die Fehlerfunktion⁷⁾ so erhält man mittels der EULER'schen Summenformel⁸⁾ nach einiger Rechnung

$$\begin{aligned} \sum_{k=K}^{n_2} p(k) &\sim \frac{1}{2} [H((n_2 - \langle k \rangle) / \sigma \sqrt{2}) - H((K - \langle k \rangle) / \sigma \sqrt{2})] \\ &+ (4 \sqrt{2} \sigma)^{-1} [H'((n_2 - \langle k \rangle) / \sigma \sqrt{2}) + H'((K - \langle k \rangle) / \sigma \sqrt{2})] \\ &+ (48 \sigma^2)^{-1} [H''((n_2 - \langle k \rangle) / \sigma \sqrt{2}) - H''((K - \langle k \rangle) / \sigma \sqrt{2})]. \end{aligned} \tag{3-8''}$$

Benützt man die Approximation (6), so hat man an Stelle von (3-7')

$$p(k) \approx (\sigma \sqrt{8})^{-1} H'((K - \langle k \rangle) / \sigma \sqrt{2}). \tag{3-8'}$$

Mittels der Gleichungen (8') und (8''), der numerischen Werte von $\langle k \rangle$ und σ bei gegebenem K, n_1, n_2 (s. Fig. 1, 2) lassen sich die Grössen $p(k), \sum p(k)$ angenähert bestimmen und damit die oben beschriebenen Entscheidungen auf Grund von Beobachtungen treffen.

7) Für die Definition der Fehlerfunktionen $\text{Erfc } x$,

$$H(x) = (2/\sqrt{\pi}) \int_0^x e^{-t^2} dt \quad \text{und} \quad \Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^x e^{-t^2/2} dt$$

siehe ERDÉLYI, MAGNUS, OBERHETTINGER & TRICOMI, Higher Transcendental Functions, Vol. 2, p. 147, McGraw-Hill, New York 1953, bzw. B. L. VAN DER WAERDEN, Mathematische Statistik, p. 10, Springer, Berlin 1957: Zur numerischen Rechnung ist die Funktion $H(x)$ bequemer, siehe z. B. die Tabellen für $H(x)$ und dessen Ableitungen bei JAHNKE-EMDE, Funktionentafeln, Dover, New York 1945, p. 24 ff. (dort mit $\Phi(x), \Phi_1(x) \dots$ bezeichnet).

8) Siehe z. B. WHITAKER & WATSON, Modern Analysis, Cambridge 1950, p. 127.

4. Diskussion und Beispiele

4.1. Wir wenden den oben beschriebenen Test auf einige Beispiele an.

4.11. *Propan*⁹⁾. In den Schwingungsspektren des Propans wurden im Fingerprintgebiet $n_2 = 5$ RAMAN-Linien und $n_1 = 11$ Infrarotbanden gefunden, von denen $K = 4$ Koinzidenz zeigen («alle» IR.-Banden sind hierbei gezählt und die Koinzidenz ist anhand der Literaturangaben fixiert). Nach Fig. 1 und 2 ist $\langle k \rangle(11,6) = 2,15$, $\sigma = 1,12$, somit ist gemäss (3-7') $w(4) \approx 0,09$. Demnach ist die Wahrscheinlichkeit, dass zufälligerweise 4 der 6 RAMAN-Linien Koinzidenz zeigen, 0,09, und man kann die Hypothese mit einer Vertrauensgrenze von 90% verwerfen.

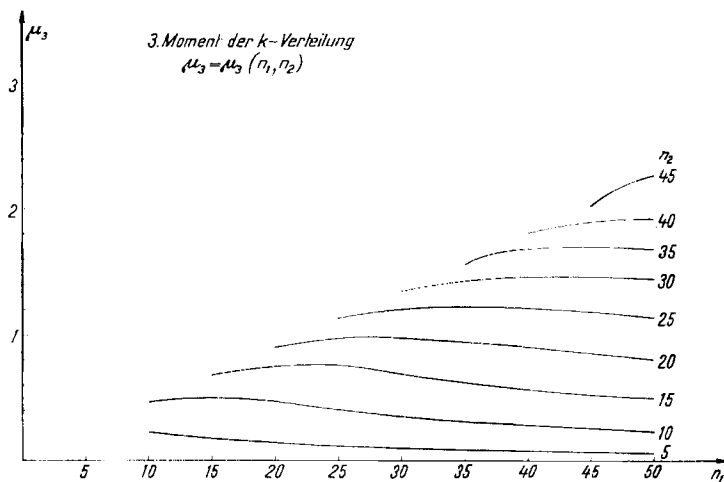


Fig. 3

4.12. Die Untersuchung der Schwingungsspektren des Cyclododecans¹⁰⁾ ergab für die flüssige Phase $n_1 = 46$, $n_2 = 35$, $K = 20$. Folglich ist $\langle k \rangle = 11,60$, $\sigma = 2,65$, $K > \langle k \rangle$, $w(K) \approx 1,0 \cdot 10^{-3}$ (die exakte numerische Rechnung ergab $w(20) \approx 1,25 \cdot 10^{-3}$). Die Koinzidenzen sind somit (im Rahmen einer Vertrauensgrenze von mindestens 0,998) signifikant.

4.13. *Dimethylacetylen*¹¹⁾. In den Schwingungsspektren des Dimethylacetylens wurden beobachtet: RAMAN-Spektrum (l): 7 Linien zwischen 520...1300 cm^{-1} Infrarotspektrum (g): 8 Linien zwischen 520...1300 cm^{-1} Zahl der Koinzidenzen: 1 (evtl. 2).

Somit $\langle k \rangle(8,7) \approx 2,2$; $\sigma = 1,08$ und, da $K < \langle k \rangle$, ist die Hypothese richtig, d. h. die Koinzidenzen zwischen Infrarot und RAMAN-Spektren sind zufällig.

4.2. Es ist vielleicht angebracht, einige Bemerkungen zu machen im Anschluss an die oben sub 2 gemachten Feststellungen.

⁹⁾ G. HERZBERG, *Infrared-Ramanspectra of Polyatomic Molecules*, Van Nostrand Comp., New York 1945, p. 361.

¹⁰⁾ E. BILLETER & Hs. H. GÜNTARD, *Helv.* **41**, 686 (1958).

¹¹⁾ G. HERZBERG, *loc. cit.*⁹⁾, p. 358, samt weiteren Literaturzitate.

¹²⁾ Wir danken Prof. RUTISHAUSER auch an dieser Stelle für die Ausführung der Rechnungen.

¹³⁾ Zum Beweis verwendet man z. B. Formel 5.19.2 bei ERDÉLYI, MAGNUS, OBERHETTINGER & TRICOMI, *Integral Transforms*, Bd. 1, p. 291.

¹⁴⁾ *J. Math. Phys.* **34**, 182 (1955).

Im allgemeinen wird im Infrarotspektrum zufolge der leichten Beobachtbarkeit von Kombinations- und Obertönen eine zu hohe Zahl von Schwingungsübergängen gezählt. Im RAMAN-Spektrum sind meist nur Grundtöne beobachtbar. Eine der Zahlen n_1 oder n_2 und evtl. K wird in diesen Fällen zu hoch angenommen. Intermolekulare Wechselwirkungen können Durchbrechungen der Auswahlregeln zur Folge haben, wodurch die Zahl der Koinzidenzen zu hoch gezählt wird. Dann läuft man Gefahr, dass der Test Signifikanz für die Koinzidenzen ergibt. Bei Molekeln mit vielen Schwingungsübergängen im Fingerprintgebiet wird die beschränkte Genauigkeit der Frequenzmessung zu zu hohen Koinzidenzzahlen K führen. In Gebieten mit starker Anhäufung von Schwingungsübergängen (Frequenzunterschied \approx Linienbreite) ist der Test mit Vorsicht anzuwenden.

5. Anhang

5.1. Die für die Herstellung der Fig. 1, 2, 3 benötigten Werte von

$$\Sigma P(k), \Sigma kP(k), \Sigma k^2P(k), \Sigma k^3P(k)$$

wurden von Herrn Prof. Dr. H. RUTISHAUSER vom Institut für angewandte Mathematik auf der ERMETH numerisch berechnet¹²⁾.

5.2. Die Summe $\Sigma P(k)$ lässt sich auf viele Arten in bestimmte Integrale umwandeln, z. B. solche mit LAGUERRE-Polynomen¹³⁾, wie

$$\Sigma P(k) = \binom{n_1 + n_2}{n_1} \Sigma \binom{n_1}{k} \binom{n_2}{k} / \binom{n_1 + n_2}{k} = (-)^{n_2} [2^{n_2+1} / n_1!] \int_0^\infty e^{-t/2} t^{n_1} L_{n_2}(t) dt$$

die jedoch wegen der Vorzeichenwechsel im Integranden für numerische Rechnung nicht bequem schienen.

5.3. Summiert man nach einer von A. D. WHEELER¹⁴⁾ angegebenen Methode unter Benützung von $\int_0^\infty e^{-u} u^n du = n!$, einer Hilfsvariablen und zweimaliger LAPLACE-Transformation, so kann man zeigen, dass

$$\Sigma P(k) = (1/n_2!) [(-\partial/\partial\lambda)^{n_2} \{(2-\lambda)^{n_1} \lambda^{-1-n_1}\}]_{\lambda=1+0} .$$

Diese Gleichung lässt sich aus einer Erzeugenden¹⁵⁾ erhalten: mit $x = 1-\lambda$ wird aus

$$\Sigma P(k) = (1/n_2!) [(d/dx)^{n_2} \{(1+x)^{n_1} (1-x)^{-1-n_1}\}]_{x=-0} \tag{5-9}$$

und somit

$$(1+x)^{n_1} (1-x)^{-1-n_1} = \sum_{k=0}^\infty x^k \frac{(n_1+k)!}{k! n_1!} F(-n_1, -k; -n_1-k; -1) . \tag{5-10}$$

Mittels des Residuensatzes folgt hieraus, dass

$$\frac{(n_1+n_2)!}{n_1! n_2!} F(-n_1, -n_2; -n_1-n_2; -1) = \frac{1}{2\pi i} \int_{\mathfrak{K}} x^{-1-n_2} (1+x)^{n_1} (1-x)^{-1-n_1} dx, \tag{5-11}$$

wo \mathfrak{K} einen ganz im Einheitskreis der x -Ebene gelegenen Kreis mit Mittelpunkt Null bedeutet. Der Integrand von (5-11) kann als $\exp [-(1+n_2)\ln x + n_1 \ln(1+x) - (1+n_1)\ln(1-x)]$ geschrieben werden und hat die durch Gleichung $h'(x) = -[(1+n_2)/x] + [n_1/(1+x)] + [(1+n_1)/(1-x)] = 0$ bestimmten Sattelpunkte bei

$$x_0 = [(1+2n_1)/2(2+n_2)] \left\{ -1 \pm \left[1 + \left((1+n_2)(2+n_2) / \left(\frac{1}{2} + n_1 \right)^2 \right) \right]^{1/2} \right\} .$$

¹⁵⁾ Ähnliche Formeln siehe ERDÉLYI, MAGNUS, OBERHETTINGER, Higher Transcendental Functions, Bd. 3, p. 248, (22), (23).

Deformiert man den Integrationsweg in (5-11) zur Gerade $x = x_0 + iy$, so erhält man in der üblichen Weise

$$\sum_0^{n_2} P(k) \sim x_0^{-1-n_2} (1+x_0)^{n_1} (1-x_0)^{-1-n_1} (2\pi h''(x_0))^{-1/2} \quad (5-12)$$

mit

$$h''(x_0) = (1+n_2) x_0^{-2} - n_1 (1+x_0)^{-2} + (1+n_1)(1-x_0)^{-2} > 0.$$

In dem Bereich von $n_1 = 10(5)50$, $n_2 = 5(5)50$; $n_2 \leq n_1$, erwies die numerische Rechnung, dass die Sattelpunktsformel (5-12) zu wenig genau ist.

5.4. Die mittels der Normalverteilung (2-6) erzielbare Approximation von $p(k)$ führt in den von uns kontrollierten Fällen zu Fehlern, welche für $|(k - \langle k \rangle) / \sigma \sqrt{2}| < 1,9$ unter ca. 10% liegen. Dies zeigt etwa die folgende Tabelle des Spezialfalls $n_1 = 41$, $n_2 = 35$, $\langle k \rangle = 11,06$, $\sigma = 2,63$.

k	p(k)	w(k)	k	p(k)	w(k)
3	0,0007	0,0013	13	0,1132	0,1155
4	28	41	14	793	813
5	88	110	15	484	494
6	223	223	16	258	260
7	460	460	17	120	118
8	795	771	18	49	47
9	1156	1159	19	17	16
10	1436	1398	20	5	5
11	1533	1517	21	1	1
12	1414	1423	22	—	—

Für $(k - \langle k \rangle) / \sigma \sqrt{2} < -2$ besonders können grössere Abweichungen auftreten, die für die praktische Anwendung jedoch nicht erheblich sein sollten, da dann $w(k) > p(k)$.

Die vorliegende Arbeit wurde durch den *Schweizerischen Nationalfonds* (Projekt Nr. 1284) unterstützt.

SUMMARY

A statistical test (K-test) has been devised to allow decisions to make about the significance of coincidences between different spectra. Tables for average number, root mean square and third moment of coincidence distribution are given. A few examples and precautions are discussed.

Org.-chemisches Laboratorium
der Eidg. Technischen Hochschule, Zürich