# Eine moderne Methode zur Untersuchung von Zahlenreihen

Von Dr. *Charles Willigens*, Bern

---

In einer früheren Abhandlung [1]) habe ich versucht zu zeigen, wie man mit Hilfe der Pearsonschen Darstellungsmethode wichtige Merkmale statistischer Zahlenreihen bestimmen kann, und zwar auf einfache Art, ohne die von Pearson aufgestellten «Tables for Statisticians and Biometricians» zu benötigen. In der vorliegenden Arbeit beabsichtige ich eine Reihe von Zahlenbeispielen zu behandeln, bei welchen es meistens darauf ankommt, zu bestimmen, ob die Verteilung dem Gaussschen Gesetze entspricht oder nicht. Verteilungen nach dem Gaussschen Gesetze weisen nämlich eine Reihe von einfachen Eigenschaften auf, welche vielfach auf jede beliebige Zahlenreihe übertragen werden, ohne die Zulässigkeit eines solchen Vorgehens irgendwie zu begründen.

Die wichtigsten Formeln, welche zur Anwendung gelangen können, sind folgende.

Die Verteilungsfunktion genügt einer Differentialgleichung erster Ordnung der Form

$$\frac{y'}{y} = \frac{x - a}{b_0 + b_1\, x + b_2\, x^2} \tag{1}$$

Nimmt man als Anfangspunkt das arithmetische Mittel $m$ der gegebenen Zahlenwerte, so bedeuten $\mu_2$, $\mu_3$ und $\mu_4$ die Mittelwerte der zweiten, dritten und vierten Potenzen der Differenzen $x - m$. Unter dieser Voraussetzung genügen die konstanten Koeffizienten, welche in der Differentialgleichung (1) auftreten, dem Gleichungensystem.

$$\begin{cases} b_1 = a \\ b_0 + 3\,b_2\,\mu_2 = -\,\mu_2 \\ 3\,b_1\,\mu_2 + 4\,b_2\,\mu_3 = -\,\mu_3 + a\,\mu_2 \\ 3\,b_0\,\mu_2 + 4\,b_1\,\mu_3 + 5\,b_2\,\mu_4 = -\,\mu_4 + a\,\mu_3 \end{cases} \tag{2}$$

Setzt man

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \qquad \beta_2 = \frac{\mu_4}{\mu_2^2}$$

---

[1]) Methode zur Bestimmung der wichtigsten Merkmale einer statistischen Zahlenreihe, Zeitschrift für schweizerische Statistik und Volkswirtschaft, 68. Jahrgang, 1932, S. 445.

so erhält man durch Auflösung des Systems (2) die Ausdrücke

$$
(2) \qquad \begin{cases} b_0 = \mu_2 \dfrac{3\beta_1 - 4\beta_2}{10\beta_2 - 12\beta_1 - 18} \\[2ex] b_1 = a = -\dfrac{\mu_3}{\mu_2} \dfrac{\beta_2 + 3}{10\beta_2 - 12\beta_1 - 18} \\[2ex] b_2 = \dfrac{-2\beta_2 + 3\beta_1 + 6}{10\beta_2 - 12\beta_1 - 18} \end{cases}
$$

Ist die Verteilungskurve symmetrisch zur Ordinatenachse, so ist $\mu_3 = 0$ und folglich auch $\beta_1 = 0$. Die Gleichungen (3) nehmen die vereinfachte Form an

$$
(3^1) \qquad \begin{cases} a = b_1 = 0 \\[1ex] b_0 = \mu_2 \dfrac{2\beta_2}{9 - 5\beta_2} \\[2ex] b_2 = \dfrac{\beta_2 - 3}{9 - 5\beta_2} \end{cases}
$$

Im Falle der symmetrischen Verteilung nimmt für die verschiedenen Werte von $\beta_2$ folgende Hauptformen an.

$1 < \beta_2 < 1{,}8$. Die Kurve weist ein Minimum auf.

$\beta_2 = 1{,}8$. Die Kurve ist eine horizontale Gerade.

$1{,}8 < \beta_2 < 3$. Die Kurve hat ein Maximum und schneidet die $ox$-Achse in zwei Punkten im Endlichen.

$\beta_2 = 3$ Gausssche Verteilungskurve.

$\beta_2 > 3$ Die Kurve hat ein Maximum und nähert sich asymptotisch der $ox$-Achse.

Der Abstand der Wendepunkte von der $oy$-Achse ist durch den Ausdruck gegeben

$$
(4) \qquad w = \pm \sqrt{\frac{b_0}{b_2 - 1}} = \pm \sqrt{\frac{\mu_2 \beta_2}{3(\beta_2 - 2)}}
$$

Für $\beta_2 = 3$ ist $w = \pm \sqrt{\mu_2}$

ist

$$
\beta_2 < 3 \quad |w| > \sqrt{\mu_2}
$$
$$
\beta_2 > 3 \quad |w| < \sqrt{\mu_2}
$$

Die Verteilung für $\beta_2 < 3$ bezeichnen wir als übernormal, für $\beta_2 > 3$ als unternormal, wobei zu beachten ist, dass dieser Begriff keineswegs mit dem den über- oder unternormalen Streuung bei Lexis identisch ist, weshalb ich den Ausdruck von Verteilung anstatt Streuung einführen möchte. Im Falle der asymmetrischen Kurve liegen die Wendepunkte, sofern sie reell sind, in gleichem Abstande der Ordinate durch das Maximum der Kurve. In diesem Falle ist dieser Abstand durch den Ausdruck gegeben.

$$w = \pm \sqrt{\frac{b_0 + a^2 (b_2 + 1)}{b_2 - 1}} \tag{4¹}$$

welcher für $a = o$ die Form (4) annimmt.

Der Wert von $\beta_1$ bestimmt die Asymmetrie die Zahl $\beta_2$ die Wölbung der Kurve.

## Anwendung auf die Fehlertheorie

Beispiel 1 (Czuber, Wahrscheinlichkeitsrechnung I, S. 335). Bestimmungen der Polhöhe von Kapstadt in den Jahren 1892 bis 1894 haben zu 15 Werten geführt, die um folgende Werte $x$ von ihrem arithmetischen Mittel differieren:

| $x$ | $x$ |
|---|---|
| — 24 | + 1 |
| — 24 | + 6 |
| — 22 | + 15 |
| — 6 | + 17 |
| — 4 | + 19 |
| — 4 | + 28 |
| — 2 | |
| — 1 | |
| — 1 | |

Die Zahlen sind Hundertstelsekunden. Berechnet man die Mittelwerte der zweiten, dritten und vierten Potenzen dieser Zahlen, so erhält man:

$$\mu_2 = \frac{3.406}{15} = 227{,}06$$

$$\mu_3 = -\frac{1334}{15} = -88{,}93$$

$$\mu_4 = \frac{1.780.054}{15} = 118.670{,}26$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0{,}00067 \qquad \beta_2 = \frac{\mu_4}{\mu_2^2} = 2{,}85545$$

Den häufigsten Wert erhält man durch die Formel

$$a = -\frac{\mu_3}{\mu_2} \frac{\beta_2 + 3}{10\beta_2 - 12\beta_1 - 18} = 0{,}2165$$

$$\sqrt{\mu_2} = 15{,}0687 \qquad \frac{a}{\sqrt{\mu_2}} = 0{,}01436$$

Das Mass der Asymmetrie $\frac{a}{\sqrt{\mu_2}}$ ist sehr klein und $\beta_2$ vom Werte 3 wenig verschieden, so dass die Fehlerverteilung gut dem Gaussschen Gesetz entspricht.

Beispiel 2 (Bertrand, Calcul des probabilités, S. 182). Auf Grund von 470 Beobachtungen der Deklination eines Sternes durch Bradley gelangte Bessel zu folgenden Abweichungen, indem er die Resultate auf 100 Beobachtungen zurückführte.

| Fehlergrenzen | $x$ | $n$ beobachtet | bei normaler Verteilung $n$ berechnet |
|---|---|---|---|
| 0,″0—0,″4 | 2 | 22,0 | 19,5 |
| 0,″4—0,″8 | 6 | 19,3 | 18,3 |
| 0,″8—1,″2 | 10 | 18,3 | 16,2 |
| 1,″2—1,″6 | 14 | 9,3 | 13,6 |
| 1,″6—2,″0 | 18 | 9,0 | 10,6 |
| 2,″0—2,″4 | 22 | 7,7 | 7,9 |
| 2,″4—2,″8 | 26 | 3,3 | 5,5 |
| 2,″8—3,″2 | 30 | 5,0 | 3,6 |
| 3,″2—3,″6 | 34 | 2,7 | 2,2 |
| 3,″6—4,″0 | 38 | 1,3 | 1,3 |
| 4,″0— — | 42 | 2,0 | 1,4 |

Als Wert für $x$ wird der mittlere Wert des Fehlerintervalls angenommen, die entsprechende Zahl $n$ gibt an, wie oft der betreffende Wert vorkommt.

Die Einheit beträgt $\frac{1}{10}$ Sekunde.

Da nur die absoluten Beträge der Fehler bekannt sind, ist es nicht möglich, den Wert von $\beta_1$ und den Betrag der Asymmetrie zu bestimmen.

Für die Mittelwerte der Potenzen 2 und 4 erhält man die Werte

$$\mu_2 = 263{,}356 \qquad \mu_4 = 214.143{,}856$$
$$\beta_2 = 3{,}09$$

Die Übereinstimmung mit dem Gaussschen Fehlergesetz darf als sehr gut angesehen werden, die entsprechenden Werte von $n$ bringen auch die leicht unternormale Verteilung zum Ausdruck.

Beispiel 3 (Bertrand, Calcul des probabilités, S. 183). Für die Rektaszension desselben Sternes erhielt Bessel folgende Zahlen:

| Fehlergrenzen | $x$ | $n$ beobachtet | bei normaler Verteilung $n$ berechnet |
|---|---|---|---|
| 0,″0—0,″1 | 0,5 | 38,0 | 33,5 |
| 0,″1—0,″2 | 1,5 | 28,0 | 28,0 |
| 0,″2—0,″3 | 2,5 | 17,7 | 19,2 |
| 0,″3—0,″4 | 3,5 | 8,0 | 10,9 |
| 0,″4—0,″5 | 4,5 | 4,7 | 5,1 |
| 0,″5—0,″6 | 5,5 | 2,0 | 2,0 |
| 0,″6—0,″7 | 6,5 | 1,0 | 0,7 |
| 0,″7—0,″8 | 7,5 | 0,3 | 0,0 |

$x$ ist wiederum die Mitte des Fehlerintervalls, jeder Wert von $x$ kommt $n$-mal vor, die Einheit beträgt $\frac{1}{10}$ Sekunde.

$$\mu_2 = 4{,}95.925 \qquad \mu_4 = 85{,}27731$$
$$\beta_2 = 3{,}47$$

Die Verteilung ist ausgesprochen unternormal. Die Fehler sind also kleiner, als dem Gaussschen Gesetz entspricht. Dies geht auch aus den berechneten Werten von $n$ nach dem Gaussschen Gesetze, welche ich nach Bertrand wiedergebe.

Aus dem Werte von $\beta_2$ ist man berechtigt auf sehr genaue Messungen zu schliessen, unter der Voraussetzung, dass systematische Konstruktionsfehler der Messapparate ausgeschaltet sind.

## Anwendung auf statistische Reihen

Beispiel 4. Während der 10 Monate von Oktober 1931 bis Juli 1932 wurden in jedem Monat pro Tag und pro Person folgende Zahlen von Karten der eidgenössischen Volkszählung numeriert.

| $x$ | $x-m$ |
|---|---|
| 100 | — 51,4 |
| 130 | — 21,4 |
| 138 | — 13,4 |
| 143 | — 8,4 |
| 145 | — 6,4 |
| 154 | + 2,6 |
| 154 | + 2,6 |
| 169 | + 17,6 |
| 188 | + 36,6 |
| 193 | + 41,6 |
| 1514 | |

$$m = 151{,}4 \qquad \mu_2 = 678{,}44 \qquad \mu_3 = -2.235{,}192$$
$$\mu_4 = 1.211.387{,}9792 \qquad \sqrt{\mu_2} = 26{,}0485$$
$$\beta_1 = 0{,}016 \qquad \beta_2 = 2{,}6318 \qquad a = +2{,}236$$
$$\frac{a}{\sqrt{\mu_2}} = 0{,}0858$$

Die Kartenzahlen $x$ sind der Grösse nach, nicht etwa zeitlich geordnet.

Die Asymmetrie ist schwach, die Verteilung übernormal, da $\beta_2$ kleiner ist als 3. Zum Vergleich soll hier die Verteilung nach den Vielfachen von $\frac{\sqrt{\mu_2}}{4}$ gegeben werden, nach der Gaussschen Verteilung.

Es wurde zwischen den Grenzen $\pm t\sqrt{\mu_2}$ folgende Anzahl der gegebenen Werte beobachtet und berechnet.

| $t$ | $n$ beobachtet | $n$ berechnet |
|---|---|---|
| 0,25 | 3 | 1,97 |
| 0,5 | 4 | 3,83 |
| 0,75 | 6 | 5,47 |
| 1 | 7 | 6,83 |
| 1,25 | 7 | 7,89 |
| 1,5 | 8 | 8,66 |
| 1,75 | 9 | 9,20 |
| 2 | 10 | 9,54 |

Auf den ersten Blick scheint die Übereinstimmung der beiden beobachteten und berechneten Reihen gut zu sein, man muss aber bedenken, dass eine Differenz um eine Einheit zwischen den beobachteten und berechneten Zahlen, zwischen den Frequenzen einer Differenz von der Grössenordnung einer Einheit der ersten Dezimalstelle entspricht, also einem relativ grossen Unterschiede.

Endlich ist noch zu beachten, dass der häufigste Wert grösser ist als der Mittelwert und folglich höheren Arbeitsleistungen entspricht.

Beispiel 5. In den Jahren von 1906 bis 1930 wurden in Basel folgende mittlere Jahrestemperaturen festgestellt.

$x =$ mittlere Temperatur, $n$-mal festgestellt.

| $x$ | $n$ | $x-m$ | $x$ | $n$ | $x-m$ |
|---|---|---|---|---|---|
| 8,9 | 3 | $-0,85$ | 9,8 | 2 | $+0,05$ |
| 9,1 | 1 | $-0,65$ | 10,0 | 4 | $+0,25$ |
| 9,2 | 1 | $-0,55$ | 10,2 | 1 | $+0,45$ |
| 9,4 | 3 | $-0,35$ | 10,4 | 2 | $+0,65$ |
| 9,5 | 1 | $-0,25$ | 10,5 | 1 | $+0,75$ |
| 9,6 | 1 | $-0,15$ | 10,6 | 2 | $+0,85$ |
| 9,7 | 3 | $-0,05$ | | | |

$$m = 9{,}75 \qquad \mu_2 = 0{,}2665 \qquad \sqrt{\mu_2} = 0{,}5162$$

$$\mu_3 = -0{,}0031 \qquad \mu_4 = 0{,}1464$$

$$\beta_1 = 0{,}0005 \qquad \beta_2 = 2{,}0611$$

$$a = +0{,}0228 \qquad \frac{a}{\sqrt{\mu_2}} = 0{,}0441$$

Wir haben ein Beispiel ausgesprochen übernormaler Verteilung, bei sehr schwacher Asymmetrie. Die entsprechende Verteilungskurve ist von der Ellipse, welche dem Werte $\beta_2 = 2$ entspricht, nicht sehr verschieden, wenn man von der Asymmetrie absieht.

Beispiel 6. Dieses Beispiel ist den Mitteilungen des statistischen Bureaus des Kantons Bern entnommen [1]). Es handelt sich darum, festzustellen, ob die

[1]) Vieh- und Geflügelbestandsermittlung des Kantons Bern vom 19. April 1929 (Untersuchungen über die Verwendbarkeit repräsentativer Erhebungsmethoden bei Viehbestandsermittlung). Mitteilungen des statistischen Bureaus des Kantons Bern, neue Folge Nr. 1.

Verhältniszahlen der Bestände an Pferden unter 4 Jahren in dem Jahre 1926 zu denen des Jahres 1921 in den 30 Amtsbezirken des Kantons Bern sich nach dem Gesetze von Gauss verteilen.

Das arithmetische Mittel $m$ beträgt 62,7, die Abweichungen $x-m$ von dieser Zahl sind:

| $x-m$ | $x-m$ |
|---|---|
| $-24,3$ | $-6,1$ |
| $-19,4$ | $-4,5$ |
| $-19,3$ | $-2,2$ |
| $-19,1$ | $+2,1$ |
| $-16,9$ | $+3,1$ |
| $-15,8$ | $+7,7$ |
| $-15,1$ | $+8,3$ |
| $-15,1$ | $+13,1$ |
| $-12,8$ | $+20,6$ |
| $-11,7$ | $+29,5$ |
| $-10,1$ | $+32,4$ |
| $-9,3$ | $+37,3$ |
| $-9,1$ | $+37,3$ |
| $-8,7$ | $+45,9$ |
| $-8,3$ | |
| $-8,2$ | |

$$x = \frac{\text{Bestand von 1926}}{\text{Bestand von 1921}}$$

Die Rechnung führt zu folgenden Werten.

$$\mu_2 = 369{,}638 \qquad \mu_3 = +7.079{,}10 \qquad \mu_4 = 382.814{,}36$$

$$\beta_1 = 0{,}9923 \qquad \beta_2 = 2{,}8018 \qquad a = +58{,}802$$

Bemerkenswert ist, dass der Wert von $a$, welcher dem Extremum von $y$ der Kurve entspricht, der gegebenen Zahlenreihe nicht mehr angehören kann. Berechnen wir mit Hilfe der Formeln (3) die Koeffizienten $b_0$ und $b_2$, so erhalten wir für die Konstanten der Differentialgleichung folgende Werte:

$$b_0 = +1611{,}06 \qquad a = b_1 = +58{,}802 \qquad b_2 = -1{,}7862$$

Die Wurzeln der Gleichung $b_0 + b_1 x + b_2 x^2 = 0$ sind $x_1 = -18{,}3$ $x_2 = +50{,}7$, der Wert von $a$ liegt also ausserhalb des Zwischenraumes $x_1$, $x_2$. Für $x=0$ ist $\frac{y_0'}{y_0} = \frac{-a}{b_0} < 0$, da $y_0$ gewiss positiv ist. Der Wert von $y$ ist für $x = -18{,}3$ unendlich gross und positiv, $y$ nimmt dann beständig ab und wird für $x = 50{,}7$ gleich Null. Die Verteilung der Zahlen ist also von der normalen weit entfernt.

Untersuchen wir, wie schon früher, die Verteilung innerhalb der Grenzen $\pm t\sqrt{\mu_2}$, wo $\sqrt{\mu_2} = 19{,}226$.

| $t$ | berechnet | beobachtet | davon $-$ | davon $+$ |
|---|---|---|---|---|
| 0,25 | 5,9 | 4 | 2 | 2 |
| 0,5 | 11,5 | 12 | 8 | 4 |
| 0,75 | 16,4 | 16 | 11 | 5 |

| $t$ | berechnet | beobachtet | davon — | davon + |
|---|---|---|---|---|
| 1 | 20,5 | 21 | 16 | 5 |
| 1,25 | 23,7 | 24 | 18 | 6 |
| 1,5 | 26,0 | 25 | 18 | 7 |
| 1,75 | 27,6 | 27 | 19 | 8 |
| 2 | 28,6 | 29 | 19 | 10 |
| 2,25 | 29,3 | 29 | 19 | 10 |
| 2,5 | 29,6 | 30 | 19 | 11 |

Sieht man von den Vorzeichen der Abweichungen ab, so entspricht die Verteilung gut dem Gesetze von Gauss, wie zu erwarten ist, da diese Verteilung einer symmetrischen Kurve für $\beta_2 = 2{,}8$ entspricht; berücksichtigt man aber die Vorzeichen, so kommt die starke Asymmetrie zum Ausdruck, welche eine normale Verteilung ausschliesst. Der Irrtum des Bearbeiters dieser Zahlenreihe in den Mitteilungen des statistischen Bureaus des Kantons Bern ist demnach leicht erklärlich und auf die Nichtberücksichtigung der Asymmetrie zurückzuführen.

Beispiel 7. Als weiteres Beispiel einer Verteilung mit starker Asymmetrie sollen noch die Zahlen der von 1875 bis 1894 in den preussischen Armeekorps durch Schlag eines Pferdes Getöteten behandelt werden.

Die Zahl von $x$ Todesfällen kam nach folgender Tabelle $n$-mal vor.

| $x$ | $n$ | $x-m$ |
|---|---|---|
| 0 | 144 | $-0{,}7$ |
| 1 | 91 | $+0{,}3$ |
| 2 | 32 | $+1{,}3$ |
| 3 | 11 | $+2{,}3$ |
| 4 | 2 | $+3{,}3$ |
| | 280 | |

$$m = \frac{\sum x n}{280} = 0{,}7$$

$$\mu_2 = 0{,}76 \qquad \mu_3 = +\,0{,}8181 \qquad \mu_4 = 2{,}3990$$

$$\beta_1 = 1{,}5248 \qquad \beta_2 = 4{,}1534$$

Auf Grund der Formeln (3) findet man

$$b_0 = -\,1{,}7473 \qquad b_1 = a = -\,1{,}4706 \qquad b_2 = +\,0{,}4330$$

Die Wurzeln der Gleichung $b_2 x^2 + b_1 x + b_0 = 0$ sind

$$x_1 = -\,0{,}9328 \qquad x_2 = +\,4{,}3284$$

Der Wert für $a$ fällt auch hier ausserhalb der Grenzen $x_1$, $x_2$, also findet in diesem Intervalle kein Zeichenwechsel der Ableitung statt.

Die Gleichung der Kurve hat die Form

$$y = c\,(x - x_1)^r\,(x - x_2)^s$$

zwischen $x_1$ und $x_2$ muss $y$ positiv sein und für $x = 0$ ist $\frac{y_0'}{y_0} = \frac{-a}{b_0} < 0$, also muss $y' < 0$ sein, und $y$ nimmt beständig ab; für $x = x_1$ muss $y$ unendlich gross und positiv sein und bei wachsendem $x$ abnehmen bis $y = 0$ für $x = x_2$. Für $x = x_1$ kann $y$ nicht gleich Null sein, da sonst zwischen den beiden Wurzeln $x_1$ und $x_2$ eine Wurzel der Ableitung liegen müsste. Die Verteilung wird also durch ein Stück der Kurve mit ständig abnehmendem $y$ dargestellt. Eine Anwendung der Dispersionstheorie von Lexis auf diesen Fall (cf. G. Udny Yule: An introduction to the theory of statistics, S. 266, Ausgabe von 1927) hat selbstverständlich keinen Sinn [1]).

Beispiel 8. Aus der Publikation des statistischen Bureaus des Kantons Bern, welche schon die Zahlen für das Beispiel 6 geliefert hat, entnehmen wir die folgenden Angaben. Für die 30 Amtsbezirke des Kantons Bern sind die Verhältniszahlen der Bestände an Mutterschweinen im Jahre 1926 zu den Beständen im Jahre 1921 berechnet worden. Es handelt sich darum, festzustellen, ob die Verteilung dieser Zahlen das Gesetz von Gauss befolgt.

Das Mittel dieser Zahlen ist $m = 79{,}9$; die Abweichungen $x - m$ vom Mittel sind wachsend geordnet.

| $x-m$ | $x-m$ | $x-m$ |
|---|---|---|
| $-48{,}1$ | $-11{,}3$ | $+11{,}5$ |
| $-43{,}9$ | $-11{,}3$ | $+12{,}0$ |
| $-33{,}2$ | $-5{,}0$ | $+15{,}6$ |
| $-31{,}3$ | $-4{,}4$ | $+16{,}3$ |
| $-25{,}9$ | $+0{,}8$ | $+17{,}1$ |
| $-19{,}6$ | $+1{,}2$ | $+18{,}2$ |
| $-16{,}2$ | $+1{,}5$ | $+20{,}1$ |
| $-15{,}2$ | $+4{,}2$ | $+36{,}1$ |
| $-13{,}6$ | $+4{,}7$ | $+40{,}3$ |
| $-11{,}8$ | $+8{,}4$ | $+81{,}6$ |

$$\mu_2 = 666{,}95 \qquad \mu_3 = +\,12.802{,}41 \qquad \mu_4 = 2.041.780{,}06$$

$$\beta_1 = 0{,}5524 \qquad \beta_2 = 4{,}5901 \qquad \sqrt{\mu_2} = 25{,}825$$

$$a = -\,6{,}8493 \qquad \frac{a}{\sqrt{\mu_2}} = 0{,}265$$

Wir haben eine Verteilung mit ausgesprochener Asymmetrie, für welche $\beta_2 > 3$.

Die Verteilung in Intervallen $\pm t\sqrt{\mu_2}$ verglichen mit der normalen Verteilung ist in folgender Tabelle gegeben.

In diesen Verteilungen ist sowohl der Einfluss der Asymmetrie als der des grossen Wertes für $\beta_2$ deutlich zu erkennen.

---

[1]) Yule gelangt übrigens nur dadurch zu einem befriedigenden Ergebnisse, dass er den Ausdruck $\sqrt{np(1-p)}$ bei kleinem $p$ durch $\sqrt{np}$ ersetzt.

| $t$ | normal | beobachtet | — | + |
|---|---|---|---|---|
| 0,25 | 5,9 | 7 | 2 | 5 |
| 0,5 | 11,5 | 13 | 5 | 8 |
| 0,75 | 16,4 | 20 | 8 | 12 |
| 1 | 20,5 | 22 | 9 | 13 |
| 1,25 | 23,7 | 24 | 11 | 13 |
| 1,5 | 26,0 | 26 | 12 | 14 |
| 1,75 | 27,6 | 28 | 13 | 15 |
| 2 | 28,6 | 29 | 14 | 15 |

Die Zahlen sind dichter am arithmetischen Mittel zusammengedrängt, als dem Gesetze von Gauss entspricht, verteilen sich aber ungleichmässig beiderseits des Mittels.

## Anwendung auf die Ergebnisse von Ziehungen

Beispiel 9. Charlier hat auf folgende Art eine Poissonsche Verteilung verwirklicht: Man verfügt über 27 Kartenspiele zu je 52 Karten. Das erste Spiel enthält 52 schwarze Karten, das zweite 51 schwarze und eine rote, das dritte 50 schwarze und 2 rote usw.; das 27. Spiel enthält 26 schwarze und 26 rote Karten. Aus jedem Spiele zieht man eine Karte und verzeichnet die Zahl der roten Karten unter den 27 gezogenen (Charlier, Grundzüge der mathematischen Statistik, S. 32).

Bei 100 Versuchen wurden nach folgender Tabelle $x$ rote Karten in 11 Fällen gezogen

| $x$ | $n$ | $x$ | $n$ |
|---|---|---|---|
| 1 | 0 | 8 | 17 |
| 2 | 0 | 9 | 14 |
| 3 | 2 | 10 | 8 |
| 4 | 6 | 11 | 1 |
| 5 | 14 | 12 | 1 |
| 6 | 14 | 13 | 1 |
| 7 | 22 | 14 | 0 |

Die Berechnung der verschiedenen Konstanten gibt:

| $x-m$ | $n$ | $x-m$ | $n$ |
|---|---|---|---|
| —6,16 | 0 | +0,84 | 17 |
| —5,16 | 0 | +1,84 | 14 |
| —4,16 | 2 | +2,84 | 8 |
| —3,16 | 6 | +3,84 | 1 |
| —2,16 | 14 | +4,84 | 1 |
| —1,16 | 14 | +5,84 | 1 |
| —0,16 | 22 | +6,84 | 0 |

$$m = 7{,}16 \qquad \mu_2 = 3{,}7544 \qquad \mu_3 = +1{,}5338$$

$$\mu_4 = 41{,}4610 \qquad \sqrt{\mu_2} = 1{,}9377 \qquad \beta_1 = 0{,}04445 \qquad \beta_2 = 2{,}9414$$

$$a = -0{,}2231 \qquad \frac{|a|}{\sqrt{\mu_2}} = 0{,}115$$

Wir haben eine sehr geringe Asymmetrie, und die Verteilung entspricht dem Gesetze von Gauss, da die Konstante $\beta_2$ vom Werte 3 nur sehr wenig verschieden ist.

Beispiel 10. Dieses Beispiel behandelt folgende Versuche von Charlier (Angaben nach Fréchet et Halbwachs, le calcul des probabilités à la portée de tous, S. 288).

Aus einem Kartenspiele, welches gleich viel rote und schwarze Karten enthält, wurden im ganzen 10.000 Karten gezogen und jeweilen wieder eingelegt. Die Ergebnisse von je 10 Ziehungen wurden zusammengefasst. Solche Gruppen, welche $x$ gezogene rote Karten aufweisen, kamen nach folgender Tabelle $n$-mal vor.

| $x$ | $n$ | $x-m$ |
|---|---|---|
| 0 | 3 | $-5$ |
| 1 | 10 | $-4$ |
| 2 | 43 | $-3$ |
| 3 | 116 | $-2$ |
| 4 | 221 | $-1$ |
| 5 | 247 | 0 |
| 6 | 202 | $+1$ |
| 7 | 115 | $+2$ |
| 8 | 34 | $+3$ |
| 9 | 9 | $+4$ |
| 10 | 0 | $+5$ |

$\sum n = 1000$. $m = 4{,}933$, zur Vereinfachung der Rechnung kann man also ohne weiteres $m = 5$ nehmen, was dem theoretischen Werte entspricht.

$$\mu_2 = 2{,}419 \qquad \mu_3 = -0{,}709 \qquad \mu_4 = 17{,}095$$

$$\beta_1 = 0{,}0355 \qquad \beta_2 = 2{,}9214$$

$$a = +0{,}1389 \qquad \sqrt{\mu_2} = 1{,}5553 \qquad \frac{a}{\sqrt{\mu_2}} = 0{,}0893$$

Auch hier befolgt die Verteilung sehr gut das Gesetz von Gauss.

Beispiel 11. Westergaard hat aus einer Urne, welche gleich viele weisse und schwarze Kugeln enthielt, 10.000 Ziehungen vorgenommen, wobei jedesmal die gezogene Kugel zurückgelegt wurde. Die Ergebnisse wurden in 100 Gruppen von je 100 Ziehungen zusammengefasst. Gruppen mit $x$ weissen Kugeln kamen nach folgender Tabelle $n$-mal vor.

| $x$ | $n$ | $x-m$ | $x$ | $n$ | $x-m$ |
|---|---|---|---|---|---|
| weniger als 34 | 0 | — | 42 | 2 | $-8$ |
| 34 | 1 | $-16$ | 43 | 3 | $-7$ |
| 35 | 0 | $-15$ | 44 | 3 | $-6$ |
| 36 | 0 | $-14$ | 45 | 4 | $-5$ |
| 37 | 0 | $-13$ | 46 | 5 | $-4$ |
| 38 | 0 | $-12$ | 47 | 6 | $-3$ |
| 39 | 1 | $-11$ | 48 | 5 | $-2$ |
| 40 | 2 | $-10$ | 49 | 11 | $-1$ |
| 41 | 2 | $-9$ | | | |

| $x$ | $n$ | $x-m$ | $x$ | $n$ | $x-m$ |
|---|---|---|---|---|---|
| 50 | 9 | 0 | 58 | 4 | + 8 |
| 51 | 5 | + 1 | 59 | 0 | + 9 |
| 52 | 10 | + 2 | 60 | 0 | + 10 |
| 53 | 4 | + 3 | 61 | 1 | + 11 |
| 54 | 8 | + 4 | 62 | 1 | + 12 |
| 55 | 3 | + 5 | 63 | 1 | + 13 |
| 56 | 5 | + 6 | 64 und mehr | 0 | |
| 57 | 4 | + 7 | | | |

$m = 50, 11$. Wir nehmen $m = 50$.

$$\mu_2 = 27{,}37 \qquad \mu_3 = -17{,}83 \qquad \mu_4 = 2377{,}57$$

$$\beta_1 = 0{,}0155 \qquad \beta_2 = 3{,}1738$$

$$a = +0{,}2965 \qquad \sqrt{\mu_2} = 5{,}2316 \qquad \frac{a}{\sqrt{\mu_2}} = 0{,}0567$$

Wir haben hier eine leicht unternormale Verteilung, bei sehr kleiner Asymmetrie.

Die Abweichung vom Gaussschen Verteilungsgesetze ist ausgeprägter als in den beiden vorhergehenden Beispielen.

Ich habe versucht, anhand von verschiedenartigen Beispielen darzulegen, wie die Pearsonsche Theorie mit Vorteil zur Bestimmung der wichtigsten Merkmale von Zahlenreihen angewendet werden kann. Die eigentliche Funktion führt, selbst wenn man über die Pearsonschen Tafeln verfügt, zu sehr komplizierten Berechnungen, ausserdem ist im allgemeinen eine analytische Darstellung der Verteilung gar nicht notwendig. Ich schliesse mich in dieser Frage ganz den Ausführungen im kürzlich erschienenen, ausgezeichneten Lehrbuche von Risser und Traynard [1]) an: «Le calcul des ordonnées des courbes de Pearson est toujours laborieux et peut difficilement être simplifié par des tables; au contraire, les constantes dont elles dépendent s'obtiennent immédiatement lorsque les moments sont calculés et elles suffisent à caractériser le type de courbe correspondant. Et c'est à cela que doivent tendre les efforts de la statistique mathématique; caractériser une distribution expérimentale par des nombres convenablement choisis de façon à pouvoir comparer des distributions analogues et chercher dans la différence des caractères numériques l'explication des variations des caractères expérimentaux.»

In meinen beiden Abhandlungen sind die Anwendungsmöglichkeiten der Pearsonschen Theorie bei weitem nicht erschöpft, es wären besonders noch die Ergebnisse periodischer Erhebungen zu untersuchen, um die auftretenden Änderungen zum Ausdruck zu bringen, wie ich es bereits in meinem ersten Aufsatze, welcher vor Erscheinen des soeben erwähnten Werkes entstanden ist, angedeutet habe. Auch auf diesem Gebiete dürften sich sehr nützliche Ergebnisse erzielen lassen.

---

[1]) R. Risser et C.-E. Traynard, Les principes de la statistique mathématique, Gauthier-Villars & C[ie], Paris 1933, S. 65.