Über den Begriff «Exzess» in der mathematischen Statistik

Von *E. Michalup*, Caracas

Die formelmässige Darstellung des Exzesses nach Pearson ist

$$\beta_2 - 3 = \frac{\mu_4}{\mu_2^2} - 3$$

während Charlier den Ausdruck

$$E = 3\beta_4 = \frac{1}{8}\left(\frac{\mu_4}{\mu_2^2} - 3\right)$$

von seiner Frequenzkurve vom Typus A

$$\frac{1}{5}Y = \varphi_0 + \beta_3\varphi_0^{\mathrm{III}} + \beta_4\varphi_0^{\mathrm{IV}} + \beta_5\varphi_0^{\mathrm{V}} + \cdots$$

herleitet, wobei φ_0 die Wahrscheinlichkeitsfunktion und φ_0^i deren $i$te Ableitungen bedeuten. Hiezu schreibt Charlier [1]: «Der Exzess E beeinflusst die symmetrische Form der Frequenzkurve nicht, verändert aber die durch die Normalkurve bestimmte Verteilung der Elemente auf die verschiedenen Klassen. Ist der Exzess positiv, so ist die Anzahl von Elementen in der Nähe des Mediums grösser als bei normaler Verteilung. Die Frequenzkurve erhöht sich in der Mitte (also in der Umgebung des Mediums) über die Normalkurve (hieraus der Name *Exzess*) und die Definition des Exzesses ist so gewählt, dass diese Erhöhung gleich ist der mit E multiplizierten Höhe der Normalkurve.» Anderson [2] in seinem etwas polemisch gehaltenen, jedoch ausgezeichneten Buch bemerkt, dass «was die geraden Momente um das arithmetische Mittel anbetrifft, so ist es möglich, mit ihrer Hilfe eine gewisse Vorstellung von der *Steilheit* der Verteilungsreihe zu bekommen», ohne sich aber weiter über diesen Punkt auszubreiten. v. Mises [3] ist etwas vorsichtiger und meint, dass ein positiver Exzess bedeutet, dass in der gegebenen Verteilung grössere Abweichungen vom Mittelwert stärker, kleinere Abweichungen schwächer vertreten sind als bei der Gaußschen Verteilung. Das Umgekehrte ist Kennzeichen negativen Exzesses. Jedoch lässt seine graphische Darstellung die Vermutung aufkommen, dass er zur Ansicht Charliers hinneigt. In seinem originell geschriebenen Buch begnügt sich Jordan [4] zu bemerken, dass es sich beim Exzess um eine Art Dispersionskoeffizient höherer Ordnung handelt, der zusammen mit der Schiefheit in der mathematischen Darstellung der Pearsonschen Verteilungskurven von Bedeutung ist. Das in Nordamerika weit verbreitete Lehrbuch von Richardson [5] bringt im Prinzip die Charliersche Behauptung, dass bei positivem Exzess die Anzahl der Elemente um das arithmetische Mittel grösser ist als bei normaler Verteilung und dass bei negativem Exzess die Verteilungskurve flacher ist als bei der normalen Verteilung. Eine dieser Erklärung entsprechende graphische Darstellung ist beigegeben. Man könnte noch eine grosse Anzahl von anderen Werken über mathematische Statistik erwähnen, welche analoge Erläuterungen dieses Begriffes bringen. Von anderen Autoren hingegen wird nur der formelmässige Ausdruck angeführt, und man kann daher vermuten, dass sie die Ansicht [6] Willigens teilen, wonach es sich um einen, von der Masseinheit übrigens unabhängigen, reinen Zahlenwert handelt. Mit dieser Art der Darstellung begnügen sich z. B. Henderson [7] und Wolfenden [8] in ihren von der Actuarial Society of America und Hardy [9] in seinem vom Institut of Actuaries in England herausgegebenen Lehrbüchern. Es scheint, dass Lindeberg [10] als erster auf die unrichtige Erklärung dieses Begriffes aufmerksam gemacht hat. Er schlug die folgende Definition vor: «Es sei p die in Prozenten der gesamten Variantenzahl ausgedrückte Anzahl derjenigen Varianten, die zwischen den Grenzen $M - \frac{\sigma}{2}$ und $M + \frac{\sigma}{2}$ liegen. Weiter sei λ die Wahrscheinlichkeit, dass ein Fehler, der dem Gaußschen Gesetz folgt, absolut genommen, kleiner als der halbe Mittelfehler ausfällt. Unter dem Exzess wird die Zahl

$$E = p - 100\,\lambda$$

verstanden. Der exakte Wert von λ ist ja gleich

$$\lambda = \int\limits_{M-\frac{\sigma}{2}}^{M+\frac{\sigma}{2}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-M)^2}{2\sigma^2}} dx = \int\limits_{-\frac{1}{2}}^{+\frac{1}{2}} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx$$

in den meisten Anwendungen dürfte es aber genügen, $\lambda = 0{,}383$ zu nehmen. Als Mass des Exzesses wäre also im allgemeinen die Differenz

$$E = p - 38{,}3$$

anzusehen. Dann wendet er seine Formel auf das Charliersche Bohnenbeispiel an und findet, dass ein positiver Exzess nach seiner Definition nicht als nachgewiesen angesehen werden kann, während nach Charlier ein positiver Exzess als sicher festgestellt erscheint.

Jedenfalls hat weder seine Kritik noch seine neue Definition die ihr gebührende Beachtung gefunden, und man kann annehmen, dass die Hauptursache darin liegt, dass das von ihm gewählte Beispiel nicht genügend deutlich den Wert seiner Arbeit hervortreten liess. Es hat daher 20 Jahre gedauert, bis neuerdings auf das Unrichtige der fast allgemein als richtig angesehenen Definition hingewiesen wurde, und Kaplansky [11] zeigt nun an einigen symmetrischen und kontinuierlichen Verteilungen, dass bei positivem Exzess nach der Definition von Pearson und Charlier die Kurve beim Scheitelpunkt tiefer als die Normalkurve liegen kann und dass bei negativem Exzess die Verteilungskurve beim Scheitel einen höheren Wert annehmen kann als die Gaußsche Fehlerkurve.

Wir wollen nun ergänzend zeigen, dass die von Lindeberg gegebene Definition auch in diesen Fällen eine richtige Auskunft über den tatsächlichen Exzess gibt. Die Streuung der beiden normalisierten Verteilungen

$$P(x) = \frac{1}{3\sqrt{\pi}} \left(\frac{9}{4} + x^4\right) e^{-x^2}$$

und

$$Q(x) = \frac{3}{2\sqrt{2\pi}} e^{-\frac{x^2}{2}} - \frac{1}{6\sqrt{\pi}} \left(\frac{9}{4} + x^4\right) e^{-x^2}$$

ist die Einheit. Das vierte Moment der ersten Verteilung ist 2,75, demzufolge ist der Exzess negativ sowohl nach Pearson

$$\beta_2 - 3 = -0{,}25$$

als auch nach Charlier

$$E = -0{,}03125$$

während in Wirklichkeit die Kurve beim Nullpunkt höher liegt als die Normalkurve, denn

$$P(0) = \frac{1}{3\sqrt{\pi}} \cdot \frac{9}{4} = 0{,}423$$

und der der Normalkurve entsprechende Wert beträgt bloss

$$\frac{1}{\sqrt{2\pi}} = 0{,}399\,.$$

Nach der Lindebergschen Definition erhalten wir

$$\frac{9}{12\sqrt{\pi}} \int_{-\frac{1}{2}}^{+\frac{1}{2}} e^{-x^2}\,dx + \frac{1}{3\sqrt{\pi}} \int_{-\frac{1}{2}}^{+\frac{1}{2}} x^4 e^{-x^2}\,dx = 0{,}392$$

also einen positiven Exzess

$$E = 39{,}2 - 38{,}3 = +0{,}9$$

in völliger Übereinstimmung mit den tatsächlichen Verhältnissen.

Bei der zweiten Verteilung beträgt das vierte Moment 3,125 und

$$Q(0) = \frac{3}{2\sqrt{2\pi}} - \frac{1}{6\sqrt{\pi}}\,\frac{9}{4} = 0{,}387 < 0{,}399\,.$$

Der Exzess ist daher positiv sowohl nach Pearson

$$\beta_2 - 3 = +0{,}125$$

als auch nach Charlier

$$E = +0{,}015625$$

während in Wirklichkeit und nach Lindeberg

$$\frac{3}{2\sqrt{2\pi}} \int_{-\frac{1}{2}}^{+\frac{1}{2}} e^{-\frac{x^2}{2}} dx - \frac{1}{6\sqrt{\pi}} \int_{-\frac{1}{2}}^{+\frac{1}{2}} \left(\frac{9}{4} + x^4\right) e^{-x^2} dx = 0{,}378$$

also $$E = 37{,}8 - 38{,}3 = -0{,}5$$

ein negativer Exzess vorliegt. Aus denselben Gründen kann daher das von Milton Da Silva Rodrigues [12] angeführte Mass für den Exzess nicht als Ausdruck für den tatsächlichen Exzess angesehen werden. Ob Relationen von absoluten Momenten existieren, wie sie z. B. Friedli [13] untersucht hat, die eher geeignet sind, sich ein Bild über die Verteilungskurve und deren Form machen zu können, soll hier nicht untersucht werden. Das Problem bleibt offen, ob es Verteilungen gibt, welche positiven Exzess im Sinne der Lindebergschen Definition haben, deren Scheitelpunkt aber unterhalb der Normalkurve liegt oder umgekehrt.

Zitierte Quellen

[1] Vorlesungen über die Grundzüge der mathematischen Statistik, Lund 1920, Seite 78.

[2] Einführung in die mathematische Statistik, Wien 1935, Seite 165.

[3] Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und mathematischen Physik, Wien 1931, Seite 242.

[4] Statistique mathématique, Paris 1927, Seite 179.

[5] An Introduction to Statistical Analysis, New York 1939, Seite 102/103.

[6] «Methode zur Bestimmung der wichtigsten Merkmale einer statistischen Zahlenreihe», Zeitschrift für schweizerische Statistik und Volkswirtschaft, 68. Band, Heft 3, Seite 447.

[7] Mathematical Theory of Graduation, New York 1938, Seite 74.

[8] The Fundamental Principles of Mathematical Statistics, Toronto 1942, Seite 74.

[9] The Theory of the Construction of Tables of Mortality and of similar Statistical Tables in use by the Actuary, London 1909, Seite 41.

[10] «Über die Begriffe Schiefheit und Exzess in der mathematischen Statistik», S. A. T., Jahrgang 1925, Seite 108.

[11] «A common Error concerning Kurtosis», Journal of the American Statistical Association, Volumen XL, Nr. 230, Seite 259.

[12] «On an extension of the concept of moment with applications to measures of variability, general similarity, and overlapping». The Annals of Mathematical Statistics, Volumen XVI, Nr. 1, Seite 78.

[13] «Über eine einfache Momentenbeziehung beim Gaußschen Fehlergesetz», M. S. V. M., Band 31, Seiten 131—139.