„Die Arbeit ist nicht gedruckt. Die Handschrift ist in der Bibliothek der Universität Tartu deponiert.

Über das System der einmodigen Häufigkeitskurven im Falle Lexisscher Reihen mit Anwendungsbeispielen auf die Klimatologie

(Zusammenfassung)

A. Kärsna

Tartu 1936

Die Charakterisierung des Kollektivs durch ein Merkmal (veränderliches Merkmal) erfolgt mittels der Häufigkeitsverteilung (kurz H.-V.). Da aber diese durch Häufigkeitszahlen aller Klassen bestimmt wird (im Falle von n Klassen müssen auch n Charakteristiken sein), ist es von grossem Interesse einen analytischen Ausdruck zu finden, welcher 1) den Charakter der H.-V. mit einer kleineren Zahl von Charakteristiken darstellt und 2) die Eigenschaften der H.-V. in einer konzentrierten Form zu gebrauchen ermöglicht (in diesem Fall könnten alle Effekte, welche durch diese H.-V. bestimmt werden, mit den Koeffizienten der Gleichung in Verbindung gebracht werden).

Von den Systemen der Häufigkeitskurven sind am bekanntesten das Charliersche und das Pearsonsche System. Das Charliersche System besteht aus 2 Typen A und B und ist ein Approximationssystem. Das Pearsonsche System gibt 7 Typen einmodiger Kurven, welche verschiedene analytische Darstellungen haben, und will infolge seiner deduktiven Struktur auch die Entstehungsursachen der H.-V. ergründen.

Das Berechnen der Koeffizienten der beiden Systeme beruht auf der sog. Methode der Momente; dieselbe ermöglicht auch das Berechnen der mit der Gleichung in Verbindung stehenden Charakteristiken der H.-V. Im Falle, dass die analytische Form nicht notwendig ist, benutzt man zur Charakterisierung der H.-V. diese Charakteristiken.

Da es aber eine endliche Zahl von Entstehungsgründen der H.-V., welche eine bestimmte Grösse haben, gibt (z. B. ist die H.-V. der Temperatur durch mehrere physikalische Konstanten, spezifische Wärme, Wärmeleitung usw. bestimmt oder ist die H.-V. in den Sozialmassen beeinflusst durch bestimmte Gesetze), so ist die inhaltliche Basis des Pearsonschen Systems nicht genügend. Das auf dem erweiterten Elementarfehler beruhende System verliert seine allgemeine Gültigkeit und bei empirischen Kollektiven, von denen eine grosse Menge eine Dispersion nach der Lexisschen Reihe aufweist, hat der Grundgedanke des Systems mehr keinen Sinn [1]). Das ganze System wird zu einem Approximationssystem. Als solches ist es aber zu verwickelt und bei praktischer Anwendung zu unbequem, denn 1) das Bestimmen der Koeffizienten verlangt zeitraubende und schwere mathematische Arbeit, 2) die Koeffizienten der Gleichung sind mit den

---

[1]) In der Arbeit ist gezeigt, dass die H.-V. antropologischer Daten, die gewöhnlich mit Hilfe des Prinzips des Elementarfehlers erklärt wird und die zur Bestimmung der biologischen Typen benutzt wird, für jedes Merkmal verschieden sein können, da die Merkmale miteinander korrelativ verbunden sind. Theoretisch kann man nach der Häufigkeitskurve der Körperlänge die Häufigkeitskurve des Körpergewichts konstruieren (denn die Regressionsgleichung zwischen der Körperlänge und dem Körpergewicht ist theoretisch ableitbar), die sehr gut mit der wirklichen Kurve zusammenfällt.

typischen Eigenschaften der H.-V. nicht verbunden und die Gleichung gibt somit keine Übersicht über diese H.-V., 3) den einzelnen Typen entsprechen verschiedene Gleichungen und 4) einer und derselben H.-V. entsprechen manchmal in dem Bernoullischen Streuungsstreifen mehrere Typen.

Die auf den Momenten beruhenden Charakteristiken, welche die Grundlage beider Systeme bilden, haben bei der Charakterisierung der H.-V. viele Nachteile: 1) ihre Genauigkeiten haben ein ungleiches Gewicht, 2) die Grösse eines jeden Momentes steht in unnormaler Beziehung zur H.-V., denn das typische Gebiet des Merkmals, welches relativ genauer bestimmt ist, beeinflusst die Grösse der Momente mit einem kleineren Gewicht als das Gebiet der kleinen Häufigkeiten, welches relativ ungenauer ist (dieser unnormale Zustand wächst mit der Grösse der Momente und kann bei höheren Momenten zu kuriosen Erscheinungen führen), 3) wenn die H.-V. von der Normalkurve stark abweicht, verlieren die auf den Momenten beruhenden Charakteristiken (arithmetisches Mittel, mittlere Abweichung, Schiefheit und Exzess) ihren Sinn, welchen sie in der Nähe der Normalkurve haben. Das viel gebrauchte Streuungsmass (σ) kann schon bei symmetrischen einmodigen Kurven 57—100% aller Abweichungen enthalten und bei grossen Schiefheiten können Fälle vorkommen, in denen zu der einen Seite des arithmetischen Mittels mehr als 90% der σ-Weite ohne Elemente bleiben.

Bei den Seen Estlands (über 1600 Seen) ist das mittlere Areal ohne dem Võrtsjärv-See 14 ha, wobei bei 86% aller Seen die Grösse der Fläche unter dem Mittelwert derselben liegt. Mit dem Võrtsjärv beträgt das mittlere Areal 32 ha und 92% aller Seen haben die Fläche, die kleiner als 32 ha ist[2]. Die maximale Häufigkeit liegt bei 0,5 ha und ist 30 mal grösser als die Häufigkeit für 14 ha und 120 mal grösser als die Häufigkeit für 32 ha. Im ersten Falle ist die Dispersion $\sigma = 67$ ha und in die Grenzen $\pm\sigma$ fällt 96,6% aller Elemente; zu einer Seite des arithmetischen Mittels aber reichen die äusseren Elemente bis $0{,}21\,\sigma$. Im zweiten Falle ist $\sigma = 700$ ha; in die Grenzen $\pm\sigma$ fällt 98,8% aller Elemente, aber auf der einen Seite des arithmetischen Mittels reicht die H.-V. nur bis $0{,}04\,\sigma$.

Aus diesem Grunde geben die mittels der Momente berechneten Charakteristiken kein genügendes Bild der H.-V., denn sie ermöglichen nicht das Zeichnen der Häufigkeitskurve. Auf Grund der gegebenen Charakteristiken lassen sich viele H.-V. konstruieren, die eine unendlich kleine Wahrscheinlichkeit haben können, dass sie aus ein und demselben Kollektiv stammen.

Um von den genannten Mängeln freizukommen, ist ein neues System zur Charakterisierung der H.-V. geschaffen worden. Dieses ist ein Approximationssystem, d. h. es behandelt die H.-V. quantitativ. Das System stützt sich auf eine Gleichung, deren Grundlage die Sinuskurve ist, und welche nach zweimaliger Transformation (eine Potenztransformation $z = cx^{\frac{1}{n}}$ und eine Expo-

[2]) H. Riikoja und A. Kärsna, On the Distribution of Lakes in Estonia. Loodusuurijate Seltsi aruanded, XLII 3—4, 1935, Tartu.

nentialtransformation $z = a(e^{bx} - 1)$) die Eigenschaften des Exzesses und der Schiefheit wiedergibt. Die Gleichung ist

$$y = 1 + \cos \pi \left[\frac{\log (1 - \frac{1}{b} x)}{\log \beta}\right]^n$$

wo n die Grösse des Exzesses, β die Grösse der Schiefheit angibt und b der Abstand von der Mode eines mit der Kurve in Verbindung stehenden Punktes, des sog. Poles ist. Bei einer symmetrischen Verteilung kann die Gleichung vereinfacht werden und dann erhält man

$$y = 1 + \cos \pi \left(\frac{x}{A}\right)^n$$

wo n die frühere Bedeutung hat und A der Abstand eines Endpunktes der Kurve von der Mode ist. Weiter ist gezeigt worden, das bei einem bestimmten n-Wert ($n = 0{,}77$) die Normalkurve mit einer genügenden Genauigkeit approximiert werden kann.

Es ist durchgeführt worden, dass **die Koeffizienten der Gleichung** auch gleichzeitig **die Charakteristiken seien**, denn in diesem Fall kann die H.-V. auch durch die Koeffizienten charakterisiert werden. Als Charakteristiken der H.-V. können allgemein folgende Grössen dienen: 1) das typische Merkmal (die Mode) -- es ist die Abszisse des höchsten Punktes der Kurve (es entspricht dem arithmetischen Mittel der anderen Systeme und kann als geometrisches Mittel aufgefasst werden, wobei die Grösse des Merkmals vom frühergenannten Pole aus genommen werden kann), 2) die Variationsweite — der Abstand der beiden Endpunkte voneinander (analog der mittleren Abweichung der frühergenannten Systeme), 3) die Schiefheit — das Verhältnis der Längen beider Zweige der Kurve (die Länge des rechtsliegenden Zweiges wird durch die des linksliegegenden dividiert) und 4) der Exzess. Die Benennungen der beiden letzten Charakteristiken stimmen mit denen der Methode der Momente überein, sind jedoch mit diesen inhaltlich nicht indentisch. Die Schiefheit und der Exzess sind durch die Koeffizienten der Gleichung direkt gegeben, die Mode und Variationsweite sind aus den Koeffizienten leicht zu erhalten. Aus der Gleichung in allgemeiner Gestalt

$$y = 1 + \cos \pi \left[\frac{\log (a + cx)}{\log \beta}\right]^n$$

greift man die Gleichung

$$a + cx = 1$$

zur Bestimmung der Koordinate der Mode heraus, eine zweite

$$a + cx = \beta$$

zur Bestimmung der Koordinate des Endpunktes der Kurve und eine dritte

$$a + cx = \frac{1}{\beta}$$

zur Bestimmung des Anfangspunktes der Kurve.

Bei Berechnung der Koeffizienten ist die Genauigkeit, mit der die Häufigkeitszahlen jeder Klasse gegeben sind, in Betracht gezogen worden. Wenn die H.-V. der Bernoullischen Reihe folgt, sind die genannten Berechnungen sehr einfach. Da aber die empirischen H.-V. der Lexisschen Reihe folgen, in der eine jede Klasse ihre eigene Störungsstreuung hat, die durch Vergrössern des Kollektivumfangs nicht verkleinert werden kann, so muss man ein zweckmässiges Mass zum Schätzen der Genauigkeit der Häufigkeiten in jeder Klasse wählen. Die Einheit der relativen Häufigkeit ist nicht brauchbar, da bei praktischer Anwendung es sich erweist, dass grossen Häufigkeiten immer grössere Störungen als den kleinen entsprechen — die Störungsstreuung ist mit der relativen Häufigkeit in korrelativem Zusammenhang. Eine nähere Betrachtung zeigt, dass der von Charlier eingeführte „Störungskoeffizient" nicht annehmbar ist — bei diesem entspricht den grossen Häufigkeiten eine kleinere Störung als den kleinen Häufigkeiten — d. h. er ergibt eine Störung, die mit der relativen Häufigkeit in negativ-korrelativer Beziehung steht, was aber inhaltlich dennoch nicht begründet werden kann. Fasst man die Klassen zusammen, so gilt der Charliersche Störungskoeffizient, wenn die Störungen miteinander korrelativ verbunden sind und den Korrelationskoeffizienten $r = 1$ haben. Allgemein aber, d. h. bei nichtkorrelativem Zusammenhang oder bei einem von 1 abweichenden Korrelationskoeffizienten, gilt der Charliersche Koeffizient nicht. Fehlt der korrelative Zusammenhang, so bleibt der Lexissche Faktor konstant, was auf Grund gewisser annehmbarer Voraussetzungen auch empirisch bewiesen werden kann. Deshalb kann man eine H.-V. nach Lexisscher Reihe als eine solche nach Bernoulischer Reihe bloss mit kleinerem Kollektivumfang betrachten. Dieses ermöglicht beim Berechnen der Charakteristiken das Gewicht der Genauigkeit in Betracht zu ziehen.

Um die beste Approximation zu erhalten, wird statt der Methode der kleinsten Quadrate oder der von Pearson eingeführten Methode der Momente eine einfache sog. Flächenmethode angewandt, die darauf beruht, dass man bei der empirischen und der Approximationskurve die Häufigkeiten für gewisse Klassen gleich nimmt. Der Gebrauch dieser Methode ist sehr bequem (es sind einige Beispiele der Approximation durch lineare, exponentiale und Potenzgleichungen gegeben) und ihre Genauigkeit ist bei zweckmässigem Gebrauch ausreichend. Um dieses zu erläutern, ist die obengenannte Methode mit derjenigen der kleinsten Quadrate theoretisch verglichen worden und es hat sich erwiesen, das die Güte der beiden die gleiche ist. Zu demselben Resultat führte auch ein empirischer Vergleich beim Würfelversuch. Eine wesentliche Rolle spielt dabei die Lage der Klassen. Diese müssen genügend weit voneinander liegen; dadurch gelangen sie aber in das Bereich der kleinen Häufigkeiten, was eine Verringerung der Genauigkeit der Koeffizienten mit sich bringt. Zur Lösung des Problems ist von der graphischen und numerischen Integration Gebrauch gemacht worden. Die weitere Analyse der Resultate (die Untersuchung des Einflusses der Grösse und Lage der Klassen auf die Genauigkeit der Koeffizienten) ist nach der graphischen Methode durchgeführt und die Standardgrössen die sog. Indikatoren, welche die Eigenschaften der H.-V. sehr deutlich zum Vorschein bringen, bestimmt worden. Die Beziehungen zwischen den Indikatoren und Koeffizienten sind ebenfalls graphisch dargestellt. Zu diesem Zweck sind Nomogramme gezeichnet worden, die es ermöglichen auf Grund einiger der empirischen Kurve auf eine sehr einfache Art zu entnehmenden Daten die Koeffizienten der Gleichung zu bestimmen. Ebenso ermöglichen diese ohne besondere Mühe die der Gleichung entsprechende graphische Darstellung zu finden. Die Untersuchung einer H.-V. von 30 Klassen (Das Bestimmen der Gleichung und der ihr entsprechenden graphischen Darstellung) erfordert ungefähr 10—15 Minuten Zeit, wogegen das Pearsonsche System dazu 4—5 Stunden und noch bedeutend mehr Kenntnisse im mathematischen Rechnen benötigt.

Zusammenfassend kann gesagt werden, dass ein System der einmodigen H.-V. gegeben worden ist, welches mit geringen Zeitaufwand und Rechenkenntnissen eine Häufigkeitskurve zu bestimmen ermöglicht, deren Approximationsgüte im Falle einer Lexisschen Reihe ausreichend ist. Da die Koeffizienten der Gleichung für die H.-V. charakteristisch sind, so ist dadurch auch ein neues System der Charakteristiken der H.-V. gegeben. Bei normaler und auch bei symmetrischer Verteilung ist das auf den Momenten beruhende System ungefähr ebenso gut, wie das vorgelegte, bei schiefen Verteilungen wird es aber wenigsagend und bei sehr schiefen Verteilungen gibt es überhaupt nicht die Möglichkeit einer Vorstellung über die H.-V. Das in vorliegender Arbeit vorgeschlagene System gibt aber in jedem Falle die charakteristischen Eigenschaften der H.-V. wieder und ermöglicht das Zeichnen der Kurve.

Da den meteorologischen Elementen hauptsächlich schiefe H.-V. entsprechen, hat eine Darstellung durch die nach der Methode der Momente berechneten Charakteristiken viele Nachteile, da diese die wesentlichen klimatologischen Eigenschaften nicht zu Tage bringt. In vorliegender Arbeit ist analog dem jährlichen Gange der Temperatur nach den 65-jährigen Beobachtungen für Tartu auch der Gang der häufigsten Temperaturen untersucht und zur Kontrolle mit den 8-jährigen Daten verglichen worden. Aus diesem ist zu ersehen, dass der Gang der häufigsten Temperatur ebenso genau bestimmbar ist, wie auch der Gang der mittleren Temperatur, klimatologisch sagt aber diese häufigste Temperatur viel mehr als die mittlere. Sie veranschaulicht den wesentlichen Teil des jährlichen Ganges, wogegen das arithmetische Mittel eine künstliche Grösse ist. Weiter ist der Gang der Schiefheit und der Variationsbreite gegeben, wobei letztere eine viel grössere Beständigkeit aufweist (die Unterschiede zwischen den 8- und 65-jährigen Werten sind geringer) als der Gang der gebräuchlichen extremen Amplituden.

Als neues ist eingeführt eine klimatologisch wichtige Grösse, die typische Häufigkeit, deren Gang durch die sog. **Typhäufigkeitskurve** gegeben ist. Um das Wesen des neuen Begriffs zu erläutern, sei folgendes gesagt: die Häufigkeiten einer Klasse des Merkmals sind in bezug auf die Zeit nicht normal, sondern schief verteilt und dadurch fallen bei der H.-V. die Mode und das arithmetische Mittel nicht zusammen. Beim Konstruieren der H.-V. nimmt man die Mittel der Häufigkeiten mehrerer Jahre, mit anderen Worten, gebraucht man die mittlere Häufigkeit. Vom klimatologischen Standpunkt aus ist aber die Grösse der typischen Häufigkeit bedeutend wichtiger. Diese typischen Häufigkeiten, berechnet für jede Klasse, ergeben die Typhäufigkeitskurve Z. B. ist nach den Daten von 65 Jahren die häufigste Temperatur im Januar 1^0 bis 2^0, wobei für die Merzahl der Jahre die grösste Häufigkeit unter 0^0 liegt und nur einzelne Jahre mit ausserordentlich hohen Temperaturen die mittlere Häufigkeit für die ganze Zeitspanne zu den höheren Temperaturwerten verschieben. Die Typhäufigkeitskurve zeigt, dass die grösste Häufigkeit in das Gebiet von 0^0 bis -5^0 fällt, was für das Klima gerade charakteristisch ist.

Zum Schluss ist mit Hilfe der H.-V. und der Lexisschen Reihe das Problem der Klimaänderung untersucht und sind die Methoden erklärt, die dasselbe zu lösen ermöglichen. An dieser Stelle wird auch der Begriff der **Lexisschen Reihe zweiter Ordnung** (gibt die Streuung der Häufigkeit der Häufigkeiten) und ihre Beziehungen zu den Klimaänderungen genannt. Als Beispiel sei hier die Analyse der Temperaturdaten von Tartu (64-jährige Beobachtungen) durchgeführt. Das Beobachtungsmaterial von je 8 Jahren wird zu einem „Versuch" zusammengefasst. Da die Änderungen der Temperatur von Jahr zu Jahr der Lexisschen Reihe folgen, so entsteht bei jedem Versuch die Poissonsche Reihe. Dieses ermöglicht für alle Versuche die theoretische Streuung zu berechnen und dieselbe mit der wahren Streuung zu vergleichen. Die Berechnungen ergaben eine Streuung nach der Lexisschen Reihe (bei unverändertem Klima wäre die Bernoullische Reihe entstanden) mit dem Faktor $L = 1{,}21 \pm 0{,}10$ was auf eine Klimaänderung, wenn auch nicht sehr anschaulich, hinweist.