STATISTIK AUF DEM KREIS

nach K. E. Selkirk, University of Nottingham
Originaltitel in 'Teaching Statistics' Vol. 4 (1982)
Nr. 3: Statistics on a Circle
Übersetzung und Bearbeitung: B. Wollring

Die meisten Zufallsgrößen, für die man Mittelwerte und Standardabweichungen berechnen möchte, können durch Punkte auf der reellen Zahlengeraden dargestellt werden; man kann sich zunächst nur schwer vorstellen, daß eine andere Situation zu untersuchen ist. Es gibt jedoch Fälle in der Praxis, bei denen Werte, die dem Mittelwert oder der Standardabweichung entsprechen, für Verteilungen gefragt sind, die mit Hilfe von Winkeln gemessen werden oder aus Punkten bestehen, die längs einer geschlossenen Kurve verteilt sind.

Solche Fälle sind nicht alltäglich, treten aber in einigen Gebieten auf. In der Meteorologie zum Beispiel könnte man sich für die "Hauptwindrichtung" an einem Ort interessieren und für ein Maß dafür, inwieweit diese Richtung vorherrscht. Eine ähnliche Fragestellung entsteht, wenn man die Bewegung von Strömungen untersucht, und kann nützlich sein, wenn man bestimmen möchte, wohin schwimmende Objekte auf dem Wasser voraussichtlich driften werden. In der Biologie ist ein solches Maß besonders bei der Untersuchung von Vogelzügen nützlich, und man hat es im Zusammenhang mit Experimenten benutzt, die den Rückflug von Tauben betrafen (siehe SCHMIDT-KOENIG). Auch Geologen sind an derartigen statistischen Maßen interessiert, wenn sie verschiedene Maße für Orientierungen suchen. Im dreidimensionalen Raum ist für Astronomen die Verteilung bestimmter seltener Objekte am Himmel interessant.

Dehnt man diesen Ansatz auf periodische Funktionen aus, die von der Zeit abhängen, so kann man mit diesen Maß-zahlen auch Vorgänge untersuchen, die täglich, wöchentlich, jährlich oder in anderen Zyklen auftreten. Zum Beispiel kann die "Haupttageszeit" ermittelt werden, zu der Unfälle mit der größten Wahrscheinlichkeit passieren, oder es kann das "Haupt-Datum" ermittelt werden, zu dem man allgemeine Wahlen ansetzen könnte. Eine historisch interessante Anwendung in diesem Zusammenhang geht auf v. MISES zurück (1918), der die These prüfen wollte, daß Atomgewichte bis auf Meßfehler ganze Zahlen sind. Dies wurde geprüft, indem man nur die gebrochenen Teile der gemessenen Gewichte in ein Winkelmaß übertrug und diese im Bereich von 0° bis 360° auf Gleichverteilung hin untersuchte (siehe v. MISES).

Die meiste Literatur zu "kreisförmigen Verteilungen" ist in einer Vielzahl verschiedener Zeitschriften verstreut und nicht leicht zugänglich. Der leichtest erreichbare Text findet sich bei MARDIA (1972), der Leser sei allerdings gewarnt; denn dieses Buch wird sehr schnell sehr schwer zu verstehen.

Für Leser, die an Problemen aus der Biologie interessiert sind, bietet der Übersichtsartikel von SCHMIDT-KOENIG (1977) zum Orientierungsvermögen bei Vögeln eine gute Sammlung von Daten und Diagrammen und deren Interpretation. Dort findet man auch Zitate der ursprünglichen Arbeiten, z.B. von EMLEN und anderen.

Mittlere Richtung ("Hauptrichtung")

Die erste Aufgabe besteht darin, auf dem Kreis ein Analogon zum arithmetischen Mittel zu finden. Es mag nicht unmittelbar einsichtig sein, daß die üblichen Mittelwertbildungen für Kreisverteilungen nicht zu verwenden sind, aber ein einfaches Beispiel zeigt es: Das Mittel von 135° und 225° ist natürlich $0.5\cdot(135° + 225°) = 180°$, und analog möchte man sagen, das Mittel von 45° und 315° ist $0.5\cdot(45° + 315°) = 180°$. Kurzes Nachdenken zeigt allerdings, daß die mittlere Richtung zu 45° und 315° natürlich 0° ist, gerade entgegengesetzt zu 180°. Es

ist klar, daß man die zyklische Natur dieser Art von Daten berücksichtigen muß.

Den Schlüssel für einen natürlichen Ansatz zur Entwicklung einer passenden Statistik liegt bereits in den Beispielen, die in der Einführung genannt sind. Alle dort genannten Größen sind Vektoren, und wir können das Problem, zu einer Menge von Winkeln die mittlere Richtung zu finden, ersetzen durch das Problem, die mittlere Richtung einer Menge von Einheitsvektoren mit den entsprechenden Richtungen zu finden. Die Richtung der Resultierenden wird dann als die mittlere Richtung der Winkel definiert. Diese Definition kann man leicht auf drei oder mehr Dimensionen erweitern; und man kann auch mit Häufigkeiten rechnen, indem man die Einheitsvektoren durch Vektoren mit einer Länge ersetzt, die den Häufigkeiten entspricht.

Das Bild zeigt eine einfache Situation. Das Problem bestehe darin, die mittlere Richtung der Winkel 17°, 59°, 119° und 203° zu finden. Man kann einfach Einheitsvektoren mit diesen Richtungen zeichnen, ihre Resultierende bestimmen und deren Richtung messen, allerdings häufen sich dabei die Fehler sehr schnell, wenn man viele Vektoren zeichnet.

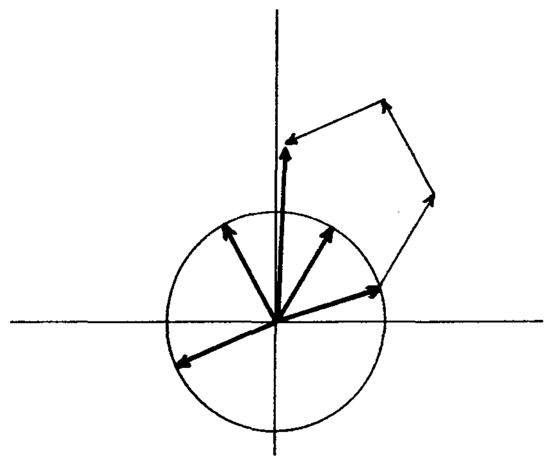

Bild 1 : zur Mittelwertbildung bei zirkularen Daten

Die entsprechende Rechnung ist ein klein wenig komplizierter, da sie die Berechnung der unabhängigen Komponenten erfordert, die den Vektor mit der mittleren Richtung festlegen. Für die obengenannten Winkel, von der x-Achse aus nach links gezählt, ist die Rechnung folgende:

$$\begin{aligned}\text{x-Komponente} &= \cos 17° + \cos 59° + \cos 119° + \cos 203°\\ &= 0.956 + 0.515 + (-0.485) + (-0.921)\\ &= 0.066\end{aligned}$$

$$\begin{aligned}\text{y-Komponente} &= \sin 17° + \sin 59° + \sin 119° + \sin 203°\\ &= 0.292 + 0.857 + 0.874 + (-0.391)\\ &= 1.633\end{aligned}$$

So ist die Richtung der Resultierenden, also die Hauptrichtung:

$$\operatorname{atan}\left(\frac{1.633}{0.066}\right) = 88°$$

In vielen Fällen wird der Winkel nicht im ersten Quadranten liegen, aber es ist nicht schwer, den richtigen Winkel anhand der Vorzeichen der beiden Komponenten zu bestimmen.

Nun können wir die allgemeine formale Definition der Hauptrichtung $\bar{x}_0$ mehrerer Winkel angeben (der Index 0 soll die zyklische Natur des Mittelwertes anzeigen): Sind die Winkel $\Theta_1, \Theta_2, -, \Theta_n$ gegeben, so definieren wir:

$$\bar{C} = \frac{1}{n}\sum \cos\Theta_\nu \qquad \text{und} \qquad \bar{S} = \frac{1}{n}\sum \sin\Theta_\nu$$

Anstatt $\bar{x}_0$ mit Hilfe dieser Größen direkt zu definieren, ist es aus Gründen, die weiter unten klar werden, zweckmäßiger, zunächst folgende Größe zu definieren:

$$\bar{R}^2 = \bar{C}^2 + \bar{S}^2 \quad (\bar{R} \geq 0)$$

Der Winkel $\bar{x}_0$, *die Hauptrichtung der* $\Theta_1, -, \Theta_n$, wird dann durch die beiden Gleichungen:

$$\bar{C} = \bar{R}\cos\bar{x}_0 \quad \text{und} \quad \bar{S} = \bar{R}\sin\bar{x}_0$$

definiert. Aus der Gleichung $\sin^2\bar{x}_0 + \cos^2\bar{x}_0 = 1$ folgt, daß diese Definition konsistent ist.

Bei dem Beispiel oben ist:

$$\bar{C} = \frac{0.066}{4} = 0.017 \quad \text{und} \quad \bar{S} = \frac{1.633}{4} = 0.408$$

Also ist $\bar{R}^2 = 0.017^2 + 0.408^2$, und damit folgt :

$$\bar{R} = 0.409$$

Der Winkel $\bar{x}_0 = 88°$ kann nun entweder aus der Sinus- oder aus der Cosinus-Gleichung oben berechnet werden, wobei man die zweite Gleichung benötigt, um sicherzustellen, daß der Winkel im richtigen Quadranten liegt.

Zirkulare Varianz

Wir stehen nun vor dem Problem, in der "Kreissituation" ein statistisches Analogon zur Varianz zu finden. Auf den ersten Blick sieht man nicht gleich, wie eine solche Größe zu definieren ist, und der Leser mag einen Moment unterbrechen, um das Problem zu bedenken.

Die Antwort liegt darin, wie in der Zeichnung auf die Resultierende zurückzugreifen. Bisher haben wir nur ihre Richtung benutzt, daher mag man fragen, was ihr Betrag bedeutet. Natürlich hängt die Länge der Resultierenden von der Zahl der Einheitsvektoren ab, aus denen sie berechnet wurde; je mehr Einheitsvektoren auftreten, desto größer wird sie. Eine sinnvollere Größe könnte also die Länge der Resultierenden, dividiert durch n, sein. Das allerdings ist gerade die oben definierte Größe $\bar{R}$.

Wenn die Winkel die geringste Streuung zeigen, sind sie alle gleich, und die n Einheitsvektoren haben alle dieselbe Richtung. In diesem Fall ist $\bar{R} = 1$. Die Winkel zeigen die größte Streuung, wenn sie ein symmetrisches Muster bilden und die Einheitsvektoren ein regelmäßiges n-Eck aufspannen. Der Wert von $\bar{R}$ ist dann 0. Die Streuung ist also maximal für $\bar{R} = 0$ und minimal für $\bar{R} = 1$. Das ist gerade das Gegenteil von dem, was wir wünschen; denn gefragt ist eine Größe, die bei maximaler Streuung 0 und bei minimaler Streuung 1 wird. Daher definieren wir die *zirkulare Varianz* S_0 durch:

$$S_0 = 1 - \bar{R}$$

Nun wird auch deutlich, weshalb man zunächst $\bar{R}$ und dann erst $\bar{x}_0$ berechnen sollte; denn die zirkulare Varianz ergibt sich bei unserem Beispiel unmittelbar als

$$S_0 = 1 - 0.409 = 0.591$$

Während also der Wert $\bar{x}_0$ komplizierter als das arithmetische Mittel zu berechnen ist, ergibt sich das überraschende und ausgleichende Ergebnis, daß die zirkulare Varianz wesentlich leichter als die übliche Varianz zu berechnen ist. Es ist sorgfältig darauf zu achten, daß S_0 als zirkulare Varianz definiert ist und nicht als zirkulare Standardabweichung, obwohl bei S_0 kein Quadrat auftritt.

Viele Eigenschaften der üblichen Varianz treten auch bei der zirkularen Varianz auf: Aus der Geometrie der Figur ist leicht zu sehen, daß $\bar{R}$ und daher auch S_0 von der Basisrichtung, von der aus die Winkel gemessen werden, unabhängig ist, ebenso wie bei einer linearen Skala die Varianz nicht von der Wahl des Nullpunktes abhängt. Die zirkulare Varianz besitzt außerdem eine Varianzanalyse-Eigenschaft, aus der heraus mehrere interessante Tests entwickelt werden können (siehe MARDIA, 1972). Wir schließen mit einem einfachen praktischen Beispiel.

Beispiel

SCHMIDT-KOENIG hat einige Experimente mit heimfliegenden Tauben durchgeführt, deren Daten bei WATSON (1962) und MARDIA (1972) genannt sind. Bei einem Experiment waren die Richtungen, in denen fünfzehn abfliegende Vögel außer Sicht kamen, gemessen auf 5° genau folgende:

i	cos i	sin i
70	0.3420	0.9397
155	-0.9063	0.4226
190	-0.9848	-0.1736
195	-0.9659	-0.2588
215	-0.8192	-0.5736
235	-0.5736	-0.8192
235	-0.5736	-0.8192
240	-0.5000	-0.8660
255	-0.2588	-0.9659
260	-0.1736	-0.9848
290	0.3420	-0.9387
300	0.5000	-0.8660
300	0.5000	-0.8660
300	0.5000	-0.8660
	-3.5718	-7.6366

Daraus berechnen wir :

$$\bar{C} = - \frac{3.5718}{14} = -0.2551$$

$$\bar{S} = - \frac{7.6366}{14} = -0.5455$$

$$\text{atan} \left(\frac{0.5455}{0.2551}\right) = 64.93°$$

$$\bar{R} = \sqrt{(0.2551^2 + 0.5455^2)} = 0.6022$$

Die Hauptrichtung ist somit etwa 180° + 65° = 245°, und die zirkulare Varianz ist $1 - \bar{R}$, das ist etwa 0.4 (siehe Bild 2).

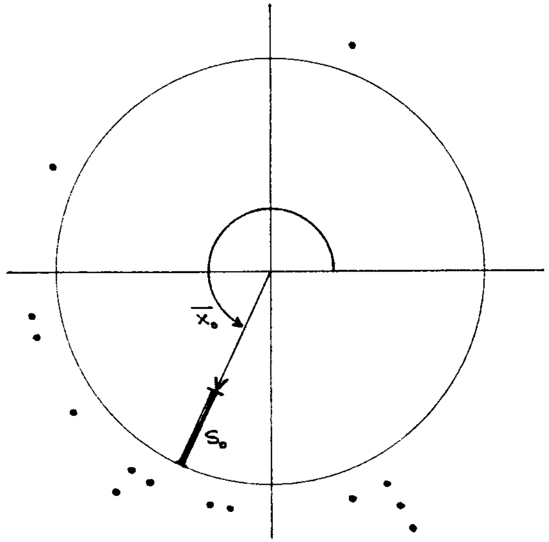

Bild 2 : Ein Histogramm zu den Daten der Tabelle

Folgerungen

Hauptrichtung und zirkulare Varianz spielen eine interessante aber begrenzte Rolle in der Statistik. Leser, die bei ihren Problemen Daten benutzen, die nicht auf der reellen Zahlengeraden darzustellen sind, mögen sich angeregt fühlen, über weitere Maße analog zu Mittelwert und Varianz für Punkte nachzudenken, die in einer Ebene verteilt sind. Die in diesem Beitrag dargestellten Ideen können auf solche Tests ausgedehnt werden, die in der allgemeineren linearen Situation angewendet werden, obgleich die Dinge dabei etwas durch die Tatsache kompliziert werden, daß keine Verteilungsfunktion zyklischer Daten die Rolle spielt, die der Normalverteilung bei eindimensionalen linearen Daten zukommt (siehe MARDIA, 1972, Seite 68 - 69).

Wie immer, so gibt es auch hier Gefahren beim gedankenlosen Anwenden solcher Statistik. Das ist insbesondere dann der Fall, wenn die Winkel so verteilt sind, daß zwei einander gegenüberliegende "Schwerpunkte" auftreten.

Einer der Zwecke der Statistik ist es ja, Daten zu vereinfachen; dies wäre ein Fall, wo man zu sehr vereinfacht. Es ist wohl ungeschickt, die Hauptrichtung zu benutzen, wenn die zirkulare Varianz sehr hoch ist, sie ist sicher eine bedeutendere statistische Größe, wenn die zirkulare Varianz gering ist.

Literatur

MARDIA, K. V.: Statistics of Directional Data. - London: Academic Press 1972

MARDIA, K. V. et al.: Multivariate Analysis. - London: Academic Press 1979

V. MISES, R.: Über die 'Ganzzahligkeit' der Atomgewichte und verwandte Fragen. - Physikalische Z. 19 (1918), Seite 490 - 500.
Reprinted in: Selected Papers of Richard v. Mises, Vol. II, Seite 15 - 34; Rhode Island: AMS 1964

SCHMIDT-KOENIG, K.: Migration and Homing in Animals. - Reihe Zoophysiology and Ecology, Band 6, Heidelberg: Springer 1975

SCHMIDT-KOENIG, K.: Orientierung und Navigation bei Vögeln. - Mannheimer Forum 1977/78, Seite 123 - 167

WATSON, G. S.: Goodness-of-fit tests on a circle, II. - Biometrika 49 (1962), 57 - 63