BESTIMMUNG DES ZUSAMMENHANGS ZWISCHEN ZUFALLSGRÖSSEN

L. SEBESTYÉN—S. SZABÓ

Lehrstuhl für Mathematik, Fakultät für Bauingenieurwesen, TU Budapest, H-1521

(Eingegangen am 12. Juni 1981)

Vorgelegt von Prof. Dr. J. Reimann

Summary

DETERMINATION OF THE RELATIONSHIP BETWEEN RANDOM VARIABLES — Theorems will be presented on the examination or determination of relationships between random variables, showing the relationship between random variables ξ and η may be expressed by means of their distribution functions $F(x)$ and $G(y)$. Under certain conditions also the relationship between random variables ξ, η, ζ, can be given in terms of the respective distribution functions $F(x)$, $G(y)$, $L(z)$

Zusammenfassung

In der Arbeit werden Sätze für die Prüfung bzw. Bestimmung von Zusammenhängen zwischen Zufallsgrößen beschrieben. Es wird gezeigt, daß im Falle der Erfüllung gewisser Bedingungen der Zusammenhang zwischen den Zufallsgrößen ξ und η mit Hilfe ihren Verteilungsfunktionen $F(x)$ bzw. $G(y)$ angegeben werden kann.

In ähnlicher Weise wird gezeigt, daß (sind die entsprechenden Bedingungen erfüllt) auch der Zusammenhang zwischen den Zufallsgrößen ξ, η, ς mit den zu diesen gehörenden Verteilungsfunktionen $F(x)$, $G(y)$, $L(z)$ angegeben werden kann.

1. Um verschiedene technische und andere praktische Probleme zu lösen, müssen oft Zusammenhänge zwischen Variablen untersucht werden. Im weiteren sollen Sätze dargelegt werden, die bei der Prüfung von Beziehungen zwischen Variablen bzw. bei deren Bestimmung benutzt werden können. Die zu untersuchenden Variablen weisen im allgemeinen zufallsbestimmte Schwankungen auf, daher können sie bei der mathematischen Analyse als Zufallsgrößen betrachtet werden. So führt das Problem zur Prüfung und Bestimmung eines Zusammenhanges zwischen Zufallsgrößen.

Um sich einen Überblick zu verschaffen werden folgende Arten von Zusammenhängen unterschieden:

1. Zwischen den Zufallsgrößen ξ und η besteht ein funktioneller Zusammenhang, der in der Form $\eta = \eta(\xi)$ angeschrieben werden kann.
2. Zwischen den Zufallsgrößen ξ und η besteht ein stochastischer Zusammenhang.

Der Unterschied zwischen den beiden Arten von Zusammenhängen zwischen Zufallsgrößen wird an zwei Beispielen dargestellt.

Nehmen wird an, daß wir in Besitz eines Kreisplattenvorrats bestehend aus Elementen der Anzahl n sind. Es seien ξ der Halbmesser einer zufällig gewählten Kreisplatte, η die Fläche derselben Platte. Dann sind ξ und η Zufallsgrößen; zwischen ξ und η besteht eine gut bestimmbare Beziehung, im vorliegenden Falle der funktionelle Zusammenhang $\eta = \pi\xi^2$.

Im Beispiel für die zweite Art seien durch ξ die Körperhöhe, durch η das Körpergewicht eines zufallsbedingt ausgewählten Mannes bezeichnet. ξ und η sind Zufallsgrößen und es liegt auf der Hand, daß im Großteil der Fälle zu einem hohen Wert von ξ ein höher Wert von η gehört. Es besteht also ein brauchbarer Zusammenhang zwischen ξ und η, dieser ist jedoch kein funktioneller Zusammenhang.

Im weiteren soll die erstere Art in folgenden Spezialfällen behandelt werden:

1.1. Die analytische Form der Beziehung $\eta = \eta(\xi)$ zwischen ξ und η ist bekannt, die Werte der darin vorkommenden Parameter sind aber unbekannt.

2.2. Die analytische Form der Beziehung zwischen den Variablen ξ und η ist nicht bekannt, durch irgendeine Überlegung ergab sich aber, daß zwischen ξ und η eine Beziehung besteht und diese durch eine monotone Funktion beschrieben werden kann.

Die Beziehung zwischen ξ und η soll im allgemeinen aus empirischen Daten bestimmt werden. Der Versuch mit den Zufallsgrößen ξ und η wird unter denselben Verhältnissen, voneinander unabhängig n-mal wiederholt, und nach den so erhaltenen empirischen Daten $(\xi_1, \eta_1), \ldots, (\xi_n, \eta_n)$ soll die Beziehung zwischen ξ und η bestimmt werden.

Mit Fall 1.1 hängt die Methode der kleinsten Quadrate zusammen. Besteht nämlich zwischen ξ und η eine lineare Beziehung, d. h. gilt $\eta = a\xi + b$, wird die Methode der kleinsten Quadrate aufgrund der empirischen Daten $(\xi_1, \eta_1), \ldots, (\xi_n, \eta_n)$ die »beste« Schätzung der Größen a und b ergeben. Ist die analytische Form des Zusammenhanges zwischen den Variablen ξ und η unbekannt, fragt sich vor allem, was für ein Kurventyp an den Punkthaufen angepaßt werden kann. Die Bestimmung der Art der gesuchten Kurve läßt sich in gewissen Fällen auf die wohlausgearbeiteten Passungs- bzw. Homogenitäts-Prüfungen der mathematischen Statistik zurückführen.

2. Um den Zusammenhang zwischen den Zufallsgrößen zu bestimmen, wird der Satz von *Reimann* benutzt.

Satz 1: Sind die Verteilungsfunktionen der Zufallsgrößen ξ und η $F(x)$ bzw. $G(y)$, so besteht zwischen ξ und η eine Beziehung, die sich durch eine monoton wachsende Funktion der Form $\eta = \eta(\xi)$ beschreiben läßt:

$$\eta(x) = G^{-1}\big(F(x)\big)$$

vorausgesetzt, daß die inverse Funktion $G^{-1}(z)$ der Funktion $G(y)$ im Zahlenintervall $0 \leq z \leq 1$ existiert.

Nimmt die unbekannte Funktion $\eta = \eta(x)$ monoton ab, so gilt unter den entsprechenden Bedingungen:

$$\eta(x) = G^{-1}\big(1 - F(x)\big).$$

Beweis:

$$F(x) = P(\xi < x) = P\big(\eta(\xi) < \eta(x)\big) =$$
$$= P\big(\eta < \eta(x)\big) = G(y) = G\big(\eta(x)\big).$$

Die Anwendbarkeit dieses Satzes wird an zwei Beispielen gezeigt. Es seien η die monoton wachsende Funktion von ξ und beide Zufallsgrößen exponentialer Verteilung, dann sind ihre Verteilungsfunktionen:

$$F(x) = \begin{cases} 1 - e^{-\alpha x} & \text{wenn} \quad x \geq 0, \\ 0 & \text{wenn} \quad x < 0, \end{cases}$$

$$G(y) = \begin{cases} 1 - e^{-\beta x} & \text{wenn} \quad y \geq 0, \\ 0 & \text{wenn} \quad y < 0. \end{cases}$$

Dann erhält man:

$$f(x) = G^{-1}\big(F(x)\big) = -\frac{1}{\beta}\log\big(1 - (1 - e^{-\alpha x})\big) = \frac{\alpha}{\beta}x, \qquad x \geq 0.$$

In diesem Falle kann also für die Prüfung des Zusammenhanges zwischen den Variablen ξ und η mit vollem Recht die Methode der kleinsten Quadrate für lineare Regression angewandt werden.

Im zweiten Beispiel sei die Funktion $\eta = \eta(\xi)$ monoton wachsend, ξ und η seien normal verteilt, d.h. ihre Verteilungsfunktionen lauten:

$$F(x) = \Phi\left(\frac{x - M(\xi)}{D(\xi)}\right),$$

$$G(y) = \Phi\left(\frac{y - M(\eta)}{D(\eta)}\right).$$

Dann lautet die Funktion $\eta(x)$, welche die gesuchte Beziehung beschreibt:

$$\eta(x) = G^{-1}\big(F(x)\big) = D(\eta)\Phi^{-1}\left(\Phi\left(\frac{x - M(\xi)}{D(\xi)}\right)\right) + M(\eta) =$$

$$= D(\eta)\,\frac{x - M(\xi)}{D(\xi)} + M(\eta) = \frac{D(\eta)}{D(\xi)}x - \frac{D(\eta)}{D(\xi)}M(\xi) + M(\eta).$$

Die Methode der kleinsten Quadrate wird also auch in diesem Falle mit Recht angewandt.

Die Verteilungen von ξ und η können durch eine *Passungsprüfung* bestimmt werden.

3. Im folgenden wird der Zusammenhang zwischen Zufallsgrößen ohne die strenge Monotonitätsbedingung bezüglich der Verteilungsfunktion $G(y)$ behandelt.

Es ist bekannt, daß die gemeinsame Verteilungsfunktion zweier Zufallsgrößen bzw. die Verteilungsfunktion der Vektorvariablen (ξ, η) alle Informationen über ξ und η enthält.

Der Definition gemäß ist nämlich die gemeinsame Verteilungsfunktion der Zufallsgrößen ξ und η die Funktion

$$H(x, y) = P(\xi < x, \eta < y), \quad (-\infty < x < \infty, -\infty < y < \infty).$$

Daraus ergeben sich die Verteilungsfunktionen von ξ und η zu:

$$F(x) = P(\xi < x) = \lim_{y \to \infty} H(x, y),$$

$$G(y) = P(\eta < y) = \lim_{x \to \infty} H(x, y).$$

Sind ξ und η unabhängig, läßt sich ihre gemeinsame Verteilungsfunktion mit Hilfe der Funktionen $F(x)$ und $G(y)$ angeben.

Nämlich:

$$H(x, y) = P(\xi < x, \eta < y) = P(\xi < x(P(\eta < y) = F(x)G(y).$$

Das bedeutet, daß falls ξ und η unabhängige Zufallsgrößen sind, die Verteilungsfunktionen $P(\xi < x)$ und $P(\eta < y)$ alle Informationen über ξ und η enthalten. In Abb. 1 sind die Funktionen $F(x)$, $G(y)$ und $H(x, y)$ für den Fall dargestellt, wenn sowohl ξ als auch η im Zahlenintervall $0 < x < a$ gleichmäßig verteilt sind.

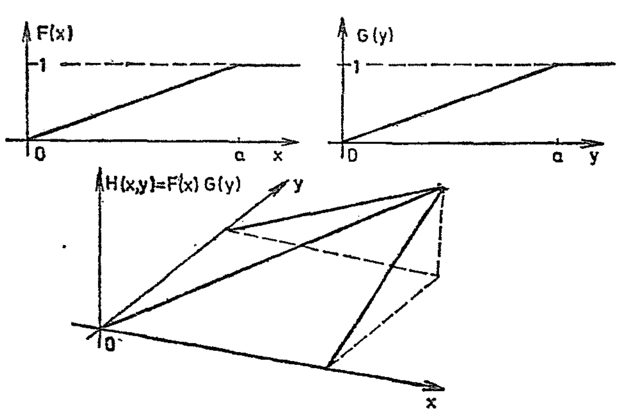

Abb. 1

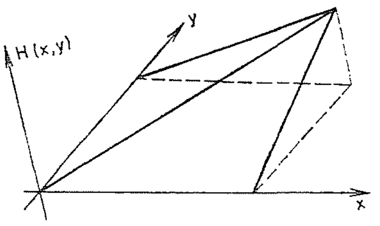

Abb. 2

Besteht zwischen ξ und η ein monotoner Zusammenhang der Form $\eta = \eta(\xi)$, dann ist

$$H(x, y) = P(\xi < x, \eta < y) = P(\xi < x, \eta(\xi) < y),$$

und da in diesem Falle

$$\eta(\xi) < \eta(x)$$

ist

$$H(x, y) = \begin{cases} P(\xi < x) = F(x) & \text{wenn} \quad \eta(x) < y, \\ P(\eta < y) = G(y) & \text{wenn} \quad \eta(x) > y. \end{cases}$$

Diese Funktion $H(x, y)$ ist für den Fall von Zufallsgrößen gleichmäßiger Verteilung im Intervall $0 < x < a$ in Abb. 2 graphisch dargestellt.

Im vorigen war zu sehen, daß besteht zwischen ξ und η eine Beziehung, diese durch eine monotone Funktion beschrieben werden kann, und sind die Verteilungsfunktionen $P(\xi < x)$, $P(\eta < y)$ bekannt, dann auch die Verteilungsfunktion $H(x, y) = P(\xi < x, \eta < y)$ bekannt ist und aufgrund von Satz 1 auch der Charakter der Beziehung zwischen den Variablen bestimmt werden kann. Nun möchten wir ohne die strenge Monotonitätsbedingung die Funktion $G(y)$ bestimmen. Für diesen Zweck werden die zwischen ξ und η bestehenden Zusammenhänge $\eta = \eta(\xi)$ bzw. $\varphi(\xi, \eta) = 0$ als Quantilkurven der Verteilungsfunktionen $F(x)$ und $G(y)$ interpretiert.

Ist die Verteilungsfunktion $F(x)$ streng monoton wachsend und ist α im Zahlenintervall (0,1) eine beliebige reele Zahl, dann hat die Gleichung $F(x) = \alpha$ eine einzige Lösung.

Ist die Verteilungsfunktion $F(x)$ nicht streng monoton wachsend, so ist die Lösungsmenge der Gleichung $F(x) = \alpha$ die leere Menge oder ein einziger Punkt, oder eine Zahlenmenge, deren Elemente in einem Intervall liegen.

Die Lösung der Gleichung $F(x) = \alpha$ wird als α-Quantil der Zufallsgröße ξ bekannt, wenn die Lösungsmenge ein einziger Punkt ist; ist die Lösung eine Strecke, wird der Halbierungspunkt der Lösung α-Quantil der Zufallsgröße bezeichnet, und ist die Lösung eine leere Menge, so sagt man, daß es kein α-Quantil gebe. So erhält man die Funktion $\xi = Q^F(\alpha)$ der Zufallsgröße ξ mit der Verteilungsfunktion $F(x)$. In ähnlicher Weise erhält man die Funktion $\eta = Q^G(\alpha)$ der Zufallsgröße mit der Verteilungsfunktion $G(y)$. Dadurch wurde aber das Parametergleichungssystem der Zusammenhänge $\eta = \eta(\xi)$ bzw. $\varphi(\xi, \eta) = 0$ angegeben.

$$\xi = Q^F(\alpha), \quad \eta = Q^G(\alpha), \quad 0 < \alpha < 1.$$

Es ist evident, daß kann die Funktion $G^{-1}(F(x))$ gebildet werden, dann diese Kurve gleich der mit dem vorigen Parametergleichungssystem angegebenen Kurve sein wird.

Nach dem Gesagten lautet Satz 1 in etwas verallgemeinerter Form:

Satz 2. Sind $F(x)$ und $G(y)$ die Verteilungsfunktionen der Zufallsgrößen ξ, bzw. η, die in Beziehung stehen, so sind die monotone Kurven $\eta = \eta(\xi)$ bzw. $\varphi(\xi, \eta) = 0$ gleich den Kurven

$$\xi = Q^F(\alpha), \quad \eta = Q^G(\alpha), \quad (0 < \alpha < 1)$$

das heißt:

$$Q^G(\alpha) = \eta(Q^F(\alpha))$$

bzw.

$$\varphi(Q^F(\alpha), Q^G(\alpha)) = 0.$$

Beweis

Aus der Gleichheit $F(x) = G(\eta(x))$ ergeben sich:

$$F(Q^F(\alpha)) = G(\eta(Q^F(\alpha))),$$
$$\alpha = G(\eta(Q^F(\alpha))).$$

Unter Berücksichtigung der vorigen Ausführungen kann bezüglich der Bestimmung des Zusammenhanges zwischen drei Zufallsgrößen folgendes gesagt werden.

Satz 3. Ist die Verteilungsfunktion der Zufallsgrößen ξ, η, ζ

$$P(\xi < x) = F(x); \; P(\eta < y) = G(y); \; P(\xi < z) = L(z),$$

die gemeinsame Verteilungsfunktion von ξ und η

$$P(\xi < x, \eta < y) = H(x, y)$$

und die Verteilungsfunktion $L(z)$ überall $(-\infty < z < +\infty)$ streng monoton, dann gilt für die Funktion $z = \zeta(x, y)$, die die Beziehung zwischen den Variablen ξ, η und ζ beschreibt — sofern diese existiert und eine streng monoton wachsende Funktion ihrer Variablen ist —, daß

$$z = \zeta(x, y) = L^{-1}(H(x, y)).$$

Beweis

Sind die Bedingungen erfüllt, so ist:

$$L(z) = L(\zeta(x, y)) = P(\zeta < \zeta(x, y)) =$$
$$= P(\zeta(\xi, \eta) < \zeta(x, y)) = P(\xi < x, \eta < y).$$

Ähnlich kann auch verfahren werden, wenn der Zusammenhang zwischen mehreren Zufallsgrößen zu prüfen ist.

Literatur

1. PRÉKOPA, A.: Wahrscheinlichkeitstheorie mit technischen Anwendungen.* Műszaki Könyvkiadó, Budapest, 1978.
2. Statistische Qualitätskontrolle,* Herausgeber: István Vincze. Budapesti Közgazdasági és Jogi Kiadó, 1958
3. VINCZE, I.: Mathematische Statistik mit industriellen Anwendungen.* Műszaki Könyvkiadó, Budapest, 1970
4. TAKÁCS, I.-L., ZIERMANN, M.: Wahrscheinlichkeitsrechnung.* Tankönyvkiadó, Budapest, 1972
5. REIMANN, J.: Wahrscheinlichkeitsrechnung und mathematische Statistik,* Lehrstoffheft, Manuskript, Tankönyvkiadó, Budapest, 1979
6. RÉNYI, A.: Wahrscheinlichkeitsrechnung,* Tankönyvkiadó, Budapest, 1968

Adj. Dr. Lukács SEBESTYÉN
Adj. Dr. Sándor SZABÓ
H-1521, Budapest

* In ungarischer Sprache