Über eine Ungleichung der Wahrscheinlichkeitsrechnung.

Heinrich Tietze zu seinem 70. Geburtstag gewidmet.

Von Georg Aumann in Würzburg (Deutschland).

1. 1. Hat man ein Kollektiv von n einander ausschließenden und ergänzenden Merkmalen $M_1, M_2, \ldots, M_n$ irgend welcher Art, mit den Wahrscheinlichkeiten $w_1, w_2, \ldots, w_n, \sum\limits_k w_k = 1$, so ergibt sich eine zweckmäßige „Arithmetisierung" der Merkmale M_k, wenn man M_k den k-ten Grundvektor $\mathfrak{e}_k = (\mathfrak{e}_{ki})$, $\mathfrak{e}_{ki} = 0$ oder 1, je nachdem $i \neq k$ oder $i = k$, eines n-dimensionalen Raumes R der Vektoren $\mathfrak{x} = (x_1, x_2, \ldots, x_n)$ zuordnet. Als Erwartungswert erhält man dann $\bar{\mathfrak{x}} = \sum\limits_k w_k \mathfrak{e}_k = (w_1, w_2, \ldots, w_n)$ und als Streuung $\overline{(\mathfrak{x} - \bar{\mathfrak{x}})^2} = \sigma^2 = \overline{\mathfrak{x}^2} - \bar{\mathfrak{x}}^2 = \sum\limits_k w_k - \sum\limits_k w_k^2 = 1 - p$, wo p die sogenannte Paschwahrscheinlichkeit bedeutet. Es gilt also die Beziehung

$$\sigma^2 + p = 1, \tag{1}$$

welche dahin gedeutet werden kann, daß bei einem Wahrscheinlichkeitsproblem mit Merkmalen, für welche sich nicht in natürlicher Weise eine zahlenmäßige Ordnung darbietet, bei Verzicht auf die obige Arithmetisierung die Paschwahrscheinlichkeit die Rolle der Streuung zu übernehmen hat. Es sei hier nur nebenbei bemerkt, daß dieser Umstand einen Aufbau der Korrelationstheorie ermöglicht, bei welchem der Begriff der Paschwahrscheinlichkeit die Grundlage bildet.

1. 2. In der kontinuierlichen Wahrscheinlichkeitsrechnung etwa einer zufälligen Variablen x, wo eine Verteilung in der Gestalt der Wahrscheinlichkeitsdichte $w(x)$, $-\infty < x < +\infty$, gegeben ist, besteht kein Bedürfnis, die Streuung durch eine andere Maßzahl zu ersetzen. Es gibt aber auch hier eine der Gleichung (1) verwandte Beziehung, allerdings nur noch in Form einer Ungleichung. Setzen wir voraus, daß $w(x)$ in $(-\infty, +\infty)$ integrierbar ist mit $\int\limits_{-\infty}^{+\infty} w(x)\,dx = 1$, ferner, daß $\int\limits_{-\infty}^{+\infty} x\,w(x)\,dx = \bar{x}$ und $\int\limits_{-\infty}^{+\infty} (x-\bar{x})^2 w(x)\,dx = \sigma^2$, die Streuung, und

$\int\limits_{-\infty}^{+\infty} w^2(x)\,dx = \beta$, die Paschdichte[1]), existieren, so gilt die Ungleichung

$$\beta\sigma \geqq \frac{3}{5\sqrt{5}} = 0{,}268. \tag{2}$$

Sie ist mit dem Gleichheitszeichen erfüllt für die Verteilung

$$w(x) = \frac{3}{4}(1-x^2) \text{ für } |x| \leqq 1, \quad w(x) = 0 \text{ sonst}$$

Eine Beschränkung von $\beta\sigma$ nach oben besteht nicht, wie schon einfache monotone Verteilungen lehren. Immerhin hält sich für die gebräuchlichsten Verteilungen $\beta\sigma$ in beachtlicher Nähe des obigen Minimalwertes, was die Verwandtschaft von (2) mit der Gleichung (1) unterstreicht. Z. B. ist für Gleichverteilung (Rechteckverteilung) $\beta\sigma = 0{,}285$, für die symmetrische Dreiecksverteilung $\beta\sigma = 0{,}272$, und für die Gaußsche Normalverteilung $\beta\sigma = 0{,}282$.

1. 3. In Verallgemeinerung von (2) gilt für eine Verteilung $w(x_1, \ldots, x_n)$ im n-dimensionalen Raum R mit $\int\limits_R w\,dm = 1$, (in erlaubter Vereinfachung) $\int\limits_R \mathfrak{x}\, w\, dm = 0$, $\beta = \int\limits_R w^2\, dm$ und $\sigma^2 = \int\limits_R \mathfrak{x}^2\, w\, dm$ die Ungleichung

$$\beta^{\frac{1}{n}}\sigma \geqq \frac{1}{\sqrt{1+\frac{4}{n}}}\left(\frac{n(4+2n)}{c_{n-1}(4+n)}\right)^{\frac{1}{n}}, \tag{3}$$

worin c_{n-1} den $(n-1)$-dimensionalen Inhalt der $(n-1)$-dimensionalen Sphäre vom Radius 1 bezeichnet ($c_0 = 2$, $c_1 = 2\pi$, $c_2 = 4\pi, \ldots$). Das Gleichheitszeichen wird für eine ähnlich einfache Verteilung, wie im Falle $n = 1$, angenommen.

1. 4. Die vorausgehenden Ungleichungen sind im wesentlichen Spezialfälle einer allgemeinen *Minimumsaufgabe:* Es sei $\lambda > 0$ vorgegeben. Es ist für alle nicht negativen, samt ihrem Quadrat für $x > 0$ integrierbaren Funktionen $W(x)$ mit den Normierungen

$$\text{(4)} \quad \int\limits_0^{+\infty} W(x)\,dx = 1, \qquad \text{(5)} \quad \int\limits_0^{+\infty} W^2(x)\,dx = 1,$$

das Infimum P_0 der Zahlen $P_W = \int\limits_0^{+\infty} x^\lambda W(x)\,dx$ zu bestimmen. Hier lautet die Lösung: $P_0 = (2\lambda+2)^\lambda (2\lambda+1)^{-(\lambda+1)} = P_{W_0}$ mit

[1]) Auch als „Nachbarschaftswahrscheinlichkeitsdichte" zu bezeichnen. Die Wahrscheinlichkeit dafür, daß bei kleinem positiven r zwei Ereignisse x_1 und x_2 einen Abstand kleiner als r haben, ist βr.

$$W_0(x) = \frac{1}{2\lambda P_0}(x_0^\lambda - x^\lambda) \text{ für } 0 < x \leqq x_0, \quad W_0(x) = 0 \text{ für } x > x_0,$$

wobei $x_0 = \frac{2\lambda+2}{2\lambda+1}$.

2. 1. Wir zeigen zunächst, daß sich die mit den in 1.2 und 1.3 genannten Ungleichung verbundenen Minimumsaufgaben auf 1.4 zurückführen lassen. Betrachten wir den Fall 1.3. Gehen wir mit einem $\varrho > 0$ von $w(\mathfrak{x})$ über zu $w'(\mathfrak{x}) = \varrho^n w(\varrho \mathfrak{x})$, so ist wieder $\int\limits_R w' dm = 1$, dagegen $\beta' = \int\limits_R w'^2 dm = \varrho^n \int\limits_R w^2 dm = \varrho^n \beta$, ferner $\sigma'^2 = \int\limits_R \mathfrak{x}^2 w' dm = \varrho^{-2}\sigma^2$, sodaß $\beta'^{\frac{1}{n}}\sigma' = \beta^{\frac{1}{n}}\sigma$. Diese Invarianzeigenschaft erlaubt neben der Bedingung $\int\limits_R w dm = 1$ auch noch $\int\limits_R w^2 dm = 1$ einzuführen. Die Aufgabe ist dann, für in solcher Weise „normierte" Verteilungen das Infimum von σ zu ermitteln.

2. 2. Ferner dürfen wir uns auf die Betrachtung symmetrischer Verteilungen $w(\mathfrak{x})$ (d. h. solcher mit $w(\mathfrak{x}) = w(\mathfrak{y})$ für $|\mathfrak{x}| = |\mathfrak{y}|$) beschränken. Um dies zu zeigen setzen wir

$$w_S(\mathfrak{x}) = \begin{cases} \frac{1}{2}(w(\mathfrak{x}) + w(-\mathfrak{x})) & \text{für } n = 1, \\ \frac{1}{c_{n-1}} \int\limits_{|\mathfrak{y}|=|\mathfrak{x}|} w(\mathfrak{y})\, df & \text{für } n > 1,\text{[2]} \end{cases}$$

wobei df das $(n-1)$-dimensionale Volumelement auf der $(n-1)$-dimensionalen Sphäre vom Radius 1 bezeichnet und für $w(\mathfrak{y})$ die entsprechenden Werte auf der Sphäre $|\mathfrak{y}| = |\mathfrak{x}|$ vom Radius $|\mathfrak{x}|$ zu nehmen sind. Bei dieser Symmetrisierung ist $\int\limits_R w_S dm = 1$, $\sigma_S^2 = \sigma^2$. Dagegen folgt nach der Schwarzschen Ungleichung $\left(\int w df\right)^2 \leqq \left(\int w^2 df\right)\left(\int df\right)$ die Beziehung $c_{n-1} w_S^2(r) \leqq \int\limits_{|\mathfrak{x}|=r} w^2 df$, sodaß

$$\beta_S = \int\limits_R w_S^2 dm = \int\limits_0^{+\infty} r^{n-1} \Big(\int\limits_{|\mathfrak{x}|=r} w_S^2 df \Big) dr = \int\limits_0^{+\infty} r^{n-1} c_{n-1} w_S^2(r)\, dr \leqq$$

$$\leqq \int\limits_0^{+\infty} r^{n-1} \Big(\int\limits_{|\mathfrak{x}|=r} w^2 df \Big) dr = \beta,$$

also $\beta_S^{\frac{1}{n}} \sigma_S \leqq \beta^{\frac{1}{n}} \sigma$.

---

[2]) $w_S(\mathfrak{x})$ ist nach bekannten Sätzen für Lebesguesche Integrale in mehreren Dimensionen fast überall vorhanden.

2. 3. Es sei $r>0$. Für die normierte und symmetrische Verteilung $w(\mathfrak{x})$ setzen wir $q(r)=w(\mathfrak{x})$ mit $|\mathfrak{x}|=r$. Dann ist

$$1=c_{n-1}\int_0^{+\infty} r^{n-1}q(r)\,dr,\quad 1=c_{n-1}\int_0^{+\infty} r^{n-1}q^2(r)\,dr,\quad \sigma^2=c_{n-1}\int_0^{+\infty} r^{n+1}q(r)\,dr.$$

Mit der Transformation $\frac{c_{n-1}r^n}{n}=x$ und $q\left(\left(\frac{nx}{c_{n-1}}\right)^{\frac{1}{n}}\right)=W(x)$ ergibt sich

$$\int_0^{+\infty} W(x)\,dx=1,\quad \int_0^{+\infty} W^2(x)\,dx=1,\quad \text{und}\quad \left(\frac{c_{n-1}}{n}\right)^{\lambda}\sigma^2=\int_0^{+\infty} x^{\lambda}W(x)\,dx\quad \text{mit}$$

$\lambda=\frac{2}{n}$. Damit ist der gewünschte Zusammenhang von 1.4 mit 1.3 (und damit auch 1.2) hergestellt.

3. 1. Bei der Behandlung des Problems 1.4 dürfen wir uns auf monotone (nicht steigende) Funktionen $W(x)$ beschränken. In der Tat, kann man zu jedem $W(x)$ die „monotone Umordnung" $\overline{W}(x)$ bilden, erklärt als die Umkehrung der Funktion $x=f(y)$, wo $f(y)$ das Lebesgüesche Maß aller positiven ξ mit $W(\xi)\geqq y$ bedeutet. Es ist augenscheinlich, aber es ist auch nicht schwer zu beweisen, daß $\int_0^{+\infty}\overline{W}(x)\,dx=\int_0^{+\infty}(\overline{W}(x))^2\,dx=1$, hingegen $P_{\overline{W}}\leqq P_W$, weil nämlich x^{λ} eine wachsende Funktion ist.

3. 2. Bei der unten vorzunehmenden Variation von $W(x)$ ist folgender Normierungsprozeß von Bedeutung: Hat man eine Funktion $V(x)$ mit $\int_0^{+\infty} V\,dx=1$, $\int_0^{+\infty} V^2\,dx=\beta$ und $\int_0^{+\infty} x^{\lambda}V\,dx=P_V$, so gelten für $W(x)=\frac{1}{\beta}V\left(\frac{x}{\beta}\right)$ die Gleichungen $\int_0^{+\infty} W\,dx=\int_0^{+\infty} W^2\,dx=1$ und $P_W=\beta^{\lambda}P_V$. Dies ergibt sich analog wie in 2.1.

3. 3. Für die spezielle Funktion $W(x)=1$ für $0<x\leqq 1$, $W(x)=0$ für $x>1$ ist $P_W=\frac{1}{\lambda+1}$. Man braucht also bei der Bestimmung von P_0 nur solche W zu untersuchen, für die $P_W\leqq\frac{1}{\lambda+1}$. Dies hat aber bei monotonem W zur Folge, daß, $x^{\lambda+1}=(\lambda+1)\xi$ gesetzt,

$$\int_0^{+\infty} W\left([(\lambda+1)\xi]^{\frac{1}{\lambda+1}}\right)d\xi\leqq\frac{1}{\lambda+1},$$

also gewiß $W\left([(\lambda+1)\xi]^{\frac{1}{\lambda+1}}\right)\xi\leqq\frac{1}{\lambda+1}$, sodaß $W(x)\leqq\frac{1}{x^{\lambda+1}}$. Ganz analog ergibt sich aus der Gleichung (5) die Ungleichung $W(x) \leq \frac{1}{\sqrt{x}}$. Damit haben wir $W(x) \leq \mathrm{Min}\left(\frac{1}{\sqrt{x}}, \frac{1}{x^{\lambda+1}}\right)$, sodaß die noch zur Konkurrenz zugelassenen $W(x)$ alle eine gemeinsame im Bereich $x > 0$ integrierbare Majorante haben. Es gibt nun eine minimalisierende Folge $W_1(x), W_2(x), \ldots$, welche die Gleichungen (4) und (5) erfüllt, und für welche $\lim_n P_{W_n} = P_0$ ist. Für jedes $x > 0$ sind die $W_n(x)$ beschränkt. Wir können mit Hilfe des Diagonalverfahrens und mit Benutzung der Monotonie der W_n eine konvergente Teilfolge ausgreifen[8]). Sei der Einfachheit halber bereits die ursprüngliche Folge konvergent zum Limes $W^*(x)$. Da die W_n eine integrierbare Majorante gemeinsam haben, ist gliedweise Integration erlaubt:

$$1 = \int_0^{+\infty} W_n dx \to \int_0^{+\infty} W^* dx,$$

sodaß $\int_0^{+\infty} W^* dx = 1$. Da wir aber für $W_n^2(x)$ und $x^\lambda W_n(x)$ keine solchen Majoranten zur Verfügung haben, beweisen wir das Bestehen von (5) für W^* und die Gleichung $P_{W^*} = P_0$ folgendermaßen: Jedenfalls ist

$$(*) \qquad \beta^* = \int W^{*2} dx \leq 1, \quad P^* = \int x^\lambda W^*(x)\, dx \leq P_0.$$

Wäre etwa $\beta^* > 1$, so für passende positive a and b auch $\int_a^b W^{*2} dx > 1$, was wegen der beschränkten Konvergenz im Intervall $a \leq x \leq b$ mit der Normierung der W_n in Widerspruch steht. Analog ergibt sich die zweite Ungleichung. Für $W^{**}(x) = W^*(x/\beta^*)/\beta^*$ ist dann

$$(**) \qquad P_{W^{**}} = \beta^{*\lambda} P^* \leq P_0.$$

Wäre eine der Ungleichungen (*) mit dem Kleinerzeichen erfüllt, so auch (**), was der Minimaleigenschaft von P_0 widerspricht. Daher ist $\beta^* = 1$ und $P^* = P_0$.

4. 1. Nun haben wir die Aufgabe, die Minimalfunktion $W^*(x)$, die wir jetzt einfach mit $W(x)$ bezeichnen, durch ein geeignetes Variationsverfahren zu bestimmen und den zugehörigen Wert P zu ermitteln.

Wir betrachten zwei verschiedene Stellen x_1 und x_2, in deren Umgebungen $W(x)$ positiv ist. Wir nehmen mit $W(x)$ folgende Variation vor: Sei $\varepsilon > 0$; im Intervall $x_1 < x < x_1 + \varepsilon$ setzen wir $V(x) = W(x) + t$,

[8]) Siehe Haupt—Aumann—Pauc, *Differential- und Integralrechnung*, Bd. I. (Berlin, 1948), S. 150, Satz 2.

in $x_2 < x < x_2+\varepsilon$ setzen wir $V(x) = W(x) - t$, außerhalb beider Intervalle sei $V(x) = W(x)$. ε und t seien dem Betrag nach so klein, daß die obigen Intervalle fremd und daß $V(x) \geqq 0$. Es ist $\int\limits_0^{+\infty} V(x)\,dx = 1$. Weiter wird $\beta = \int\limits_0^{+\infty} V^2(x)\,dx = 1+u$ mit

$$u = \int\limits_{x_1}^{x_1+\varepsilon} (2tW(x)+t^2)\,dx + \int\limits_{x_2}^{x_2+\varepsilon} (-2tW(x)+t^2)\,dx = At + Bt^2.$$

Ferner ergibt sich $P_V = P + v$ mit $v = \int\limits_{x_1}^{x_1+\varepsilon} x^\lambda t\,dx - \int\limits_{x_2}^{x_2+\varepsilon} x^\lambda t\,dx = Ct$. Gehen wir nun gemäß 3.2 über zu $W_1(x) = V(x/\beta)/\beta$, sodaß W_1 den Normierungen (4) und (5) genügt, so wird $P_{W_1} = (1+u)^\lambda (P+v)$, also

$$P_{W_1} - P = v + \lambda u P + \ldots = (C + \lambda AP)\,t + \ldots,$$

wo die letzten Punkte Glieder höherer Ordnung in t bezeichnen. Da t im Rahmen der obigen Bemerkung frei ist, so muß wegen der Minimaleigenschaft von P die Gleichung $C + \lambda AP = 0$ bestehen:

$$\int\limits_0^\varepsilon [(x_1+\xi)^\lambda - (x_2+\xi)^\lambda]\,d\xi + 2\lambda P \int\limits_0^\varepsilon [W(x_1+\xi) - W(x_2+\xi)]\,d\xi = 0.$$

Seien nun x_1 und x_2 Stetigkeitspunkte von $W(x)$. Dann können wir die letzte Gleichung entwickeln:

$$[x_1^\lambda - x_2^\lambda + 2\lambda P(W(x_1) - W(x_2))]\,\varepsilon + o(\varepsilon) = 0,$$

wobei $o(\varepsilon)/\varepsilon \to 0$ für $\varepsilon \to 0$. Dies führt auf

$$x_1^\lambda + 2\lambda P W(x_1) = x_2^\lambda + 2\lambda P W(x_2).$$

Für alle Stetigkeitspunkte x, wo $W(x) > 0$, hat $x^\lambda + 2\lambda P W(x)$ einen konstanten Wert a, sodaß für diese Stellen

$$W(x) = a - \frac{x^\lambda}{2\lambda P}.$$

Da diese Stellen (als Stetigkeitsstellen einer monotonen Funktion) überall dicht liegen, gilt nach dem bekannten Erweiterungssatz für monotone Funktionen die letzte Gleichung sogar für alle x, wo $W(x) > 0$. Dies führt zur Darstellung

$$W(x) = \frac{1}{2\lambda P}(x_0^\lambda - x^\lambda) \text{ für } 0 \leqq x \leqq x_0, \quad W(x) = 0 \text{ für } x > x_0.$$

Die Normierungsgleichungen (4) und (5) führen zu den in 1.4 angegebenen Werten für x_0 und $P = P_0$; die Gleichung $P_0 = \int\limits_0^{+\infty} x^\lambda W\,dx$ ist dann von selbst erfüllt.

(Eingegangen am 2. Januar 1950.)