Mittelwerte als Minima

Alfred Schreiber

25. Februar 2019

Mittelwerte sind von fundamentaler Bedeutung in zahlreichen Anwendungsgebieten wie Physik, Ökonometrie, Stochastik und Datenanalyse; auch innermathematisch spielen sie eine wesentliche Rolle in Geometrie und Analysis, nicht zuletzt aufgrund ihrer engen Beziehung zur Theorie der Ungleichungen. Das Kompendium von Bullen [2] gibt einen Überblick über eine Fülle verschiedenartigster Ausformungen des Mittelwertbegriffs. Die Situation verlangt nach einer Strukturierung. Schon früh (um 1930) wurde die axiomatische Methode mit Erfolg eingesetzt, um spezielle Mittelwertklassen zu charakterisieren, am bekanntesten die quasi-arithmetischen Mittelwerte (Kolmogorov [7], Nagumo [10]). Im Folgenden soll, alternativ dazu, der Idee nachgegangen werden, Mittelwerte durch Minimalitätsforderungen zu kennzeichnen. Dabei kommt – in einer allgemeinen Form – ein Verfahren ins Spiel, das bereits um 1800 von Gauß und Legendre ersonnen und in der Ausgleichsrechnung unter der Bezeichnung *Methode der kleinsten Quadrate* praktiziert wurde.

1 Die Methode der kürzesten Abstände

Die Anwendung der Methode der kleinsten Quadrate auf eine Stichprobe $\underline{x} = (x_1, x_2, \ldots, x_N)$ reeller Messwerte x_j läuft darauf hinaus, einen Wert $u \in \mathbb{R}$ so zu bestimmen, dass die Summe der Abweichungsquadrate

$$f(u) := (x_1 - u)^2 + (x_2 - u)^2 + \cdots + (x_N - u)^2 \tag{1}$$

ein Minimum wird. Wertet man in (1) die einzelnen Summanden aus, so erweist sich $f(u)$ als eine nach oben geöffnete Parabel der Form $a + bu + Nu^2$. Deren einziges relatives Minimum liegt im Scheitel, ist also auch absolut, und eine kurze Rechnung bestätigt: Die Abszisse des Scheitels ist gerade das *arithmetische Mittel*

$$u_{\min} = \mu = A(x_1, x_2, \ldots, x_N) = \frac{x_1 + x_2 + \cdots + x_N}{N}.$$

Das Verfahren lässt sich im $\mathbb{R}^N$ auf naheliegende Weise geometrisch deuten. Es ist ja $f(u)^{\frac{1}{2}}$ der euklidische Abstand zwischen $(x_1, \ldots, x_N)$ und dem Ausgleichspunkt $(u, \ldots, u)$, der für $u = A(x_1, x_2, \ldots, x_N)$ am kleinsten wird. Schon das allein ist ein triftiger Grund, die historischen "kleinsten Quadrate" terminologisch durch "kürzeste Abstände" abzulösen (Freudenthal und Steiner [5], S. 175).

Um diese Idee allgemeiner zu fassen, betrachten wir *Distanzen* (*Distanzfunktionen*) auf einer Teilmenge $T \subseteq \mathbb{R}^N$, d. h. symmetrische und positiv-definite Abbildungen

$d : T \times T \to [0, \infty)$ mit $d(\underline{x}, \underline{y}) = d(\underline{y}, \underline{x})$ und $d(\underline{x}, \underline{y}) = 0 \iff \underline{x} = \underline{y}$ für alle $\underline{x}, \underline{y} \in T$. Gilt darüberhinaus die Dreiecksungleichung $d(\underline{x}, \underline{y}) \leq d(\underline{x}, \underline{z}) + d(\underline{z}, \underline{y})$ $(\underline{x}, \underline{y}, \underline{z} \in T)$, so liegt eine *Metrik* d vor. Als *Ausgleichsmenge* $\mathbb{E} \subset \mathbb{R}^N$ fungiere die Gesamtheit aller Punkte $\overline{u} := (u, \ldots, u)$ mit $u \in \mathbb{R}_+$ (Menge der positiven reellen Zahlen).

Die Methode der kürzesten Abstände besteht nun darin, mittels einer Distanz oder Metrik d den Abstand eines Stichprobenpunkts $\underline{x} = (x_1, \ldots, x_N) \in \mathbb{R}_+^N$ von $\mathbb{E}$ zu bestimmen, genauer: $d(\underline{x}, \overline{u})$ mit $\overline{u} \in \mathbb{E}$ zu minimieren. Wenn ein eindeutiger Wert $u = u_{\min} = \mu$ dieser Art existiert, so soll μ der *d-Mittelwert von* $\underline{x}$ genannt und mit $M_d(\underline{x})$ bezeichnet werden.

Die geläufigsten Metriken auf dem $\mathbb{R}^N$ haben die Form $d_p(\underline{x}, \underline{y}) := \|\underline{x} - \underline{y}\|_p$ mit der für reelles $p \geq 1$ definierten L^p-Norm $\|\underline{x}\|_p := (|x_1|^p + \cdots + |x_N|^p)^{1/p}$. – Betrachten wir zunächst die Fälle $p = 2$, $p = 1$ und $p = \infty$:

$$d_2(\underline{x}, \underline{y}) = \Big(\sum_{j=1}^{N} (x_j - y_j)^2\Big)^{\frac{1}{2}} \qquad \text{(Euklidische Metrik)}$$

$$d_1(\underline{x}, \underline{y}) = \sum_{j=1}^{N} |x_j - y_j| \qquad \text{(Summen-Metrik)}$$

$$d_\infty(\underline{x}, \underline{y}) = \operatorname*{Max}_{1 \leq j \leq N} |x_j - y_j| \qquad \text{(Maximum-Metrik)}$$

Welche Mittelwerte lassen sich aus diesen Metriken mit der Methode der kürzesten Abstände gewinnen?

(1): Im Fall der euklidischen Metrik kennen wir das Ergebnis bereits aus der eingangs angestellten Überlegung. Die Situation soll aber noch einmal veranschaulicht werden, vorab für $N = 2$. Hier ist der Abstand des Stichprobenpunkts S (mit den Koordinaten x_1, x_2) von der Winkelhalbierenden $\mathbb{E}$ im ersten Quadranten zu bestimmen (Abb. 1).

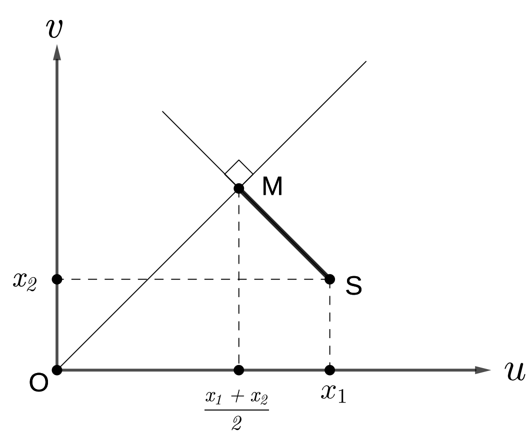

Abb. 1

Von allen Strecken, die S mit einem Punkt auf $\mathbb{E}$ verbinden, ist das Lot SM $\perp$ $\mathbb{E}$ die kürzeste. Die Gerade OM hat die Gleichung $v = u$, die Gleichung der dazu senkrechten Geraden SM lautet: $v = -u + (x_1 + x_2)$. Infolgedessen ergibt sich für den Schnittpunkt M beider Geraden: $u = -u + x_1 + x_2$, mithin $u = A(x_1, x_2)$.

Diese einfache Argumentation lässt sich unmittelbar auf den $\mathbb{R}^N$ erweitern. Der dem Stichprobenpunkt $\underline{x}$ nächstgelegene Punkt $\overline{u} \in \mathbb{E}$ liegt zusammen mit $\underline{x}$ im Orthogonalraum

$\mathbb{E}^\perp$, hier: in der zu $\mathbb{E}$ senkrechten Hyperebene durch $\underline{x}$. Diese genügt einer Gleichung

$$u_1 + \cdots + u_N = c,$$

wobei (wegen $\underline{x} \in \mathbb{E}^\perp$) gilt: $c = x_1 + \cdots + x_N$. Für den einpunktigen Schnitt $\mathbb{E} \cap \mathbb{E}^\perp = \{\overline{\mu}\}$ ergibt sich dann:

$$\mu + \cdots + \mu = x_1 + \cdots + x_n \tag{2}$$

und damit das arithmetische Mittel $\mu = A(\underline{x}) = M_{d_2}(\underline{x})$.

(2) liefert im Übrigen das Vorbild, nach dem im Niveauflächen-Ansatz Mittelwerte ("level surface means" bei Bullen [2], S. 420) *ad hoc* als eindeutige Lösungen von Gleichungen des Typs $F(\mu, \ldots, \mu) = F(x_1, \ldots, x_N)$ gewonnen werden.

(2): Bei Verwendung der Summen-Metrik d_1 entsteht eine stückweise lineare Funktion $g(u) := d_1(\underline{x}, \overline{u}) = |x_1 - u| + \cdots + |x_N - u|$, die ein absolutes Minimum annimmt. Um dieses zu bestimmen, denken wir uns die Stichprobenwerte angeordnet: $x_1 \leq x_2 \leq \ldots \leq x_N$. Bei ungeradem $N = 2m-1$ besitzt $g(u)$ ein eindeutiges Minimum für $u = \mu := x_m$. Bei geradem $N = 2m$ wird $g(u)$ für jeden Wert u aus dem abgeschlossenen Intervall $[x_m, x_{m+1}]$ minimal. Um auch im nicht-ausgearteten Fall $x_m < x_{m+1}$ Eindeutigkeit zu erzielen, wird üblicherweise der Mittelpunkt $\mu := (x_m + x_{m+1})/2$ ausgezeichnet. Der so festgelegte Mittelwert $M_{d_1}(\underline{x}) := \mu$ ist der *Median* (*Zentralwert*) der Stichprobe $\underline{x}$.

Die Minimum-Eigenschaft des Zentralwerts möge folgendes Beispiel illustrieren: Ein Unternehmen muss jeden Monat an sieben Orte $\mathrm{S}_1, \ldots, \mathrm{S}_7$, die in dieser Reihenfolge auf einer Fahrtroute liegen, beziehentlich $6, 18, 2, 7, 7, 5, 6$ Wagenladungen eines bestimmten Gutes liefern. Nun soll ein Versorgungslager an einem Ort U der Route eingerichtet werden, um den Gesamtweg zu minimieren. – Sei O ein fester, etwa vor S_1 gelegener Punkt, x_j die Längen der Strecken OS_j und u die Länge von OU. Die Gesamtweglänge (einschließlich Rückfahrten) beläuft sich dann auf das Doppelte der Summe $6|x_1 - u| + 18|x_2 - u| + \cdots + 6|x_7 - u|$. Diese wird minimal für den an 26-ter Stelle in der angeordneten Folge x_1 (6-mal) $\leq x_2$ (18-mal) $\leq \ldots \leq x_7$ (6-mal) stehenden Zentralwert $u = x_3$; dieser steht für $\mathrm{U} = \mathrm{S}_3$ als Standort des Versorgungslagers. Man überzeugt sich leicht, dass keine weitere Verbesserung erzielt wird, wenn man U zwischen zwei Orte platziert.

(3): Für $N = 2$ nimmt der Abstand $d_\infty(\underline{x}, \overline{u}) = \mathrm{Max}(|x_1 - u|, |x_2 - u|)$ auf $\mathbb{R}^2$ genau dann ein absolutes Minimum an, wenn $u = A(x_1, x_2)$. Denn setzen wir ohne Einschränkung $x_1 \leq x_2$ voraus, so zeigt eine einfache Fallunterscheidung: $d_\infty(\underline{x}, \overline{u}) = x_2 - u$ für $u \leq A(x_1, x_2)$ und $d_\infty(\underline{x}, \overline{u}) = u - x_1$ für $u \geq A(x_1, x_2)$. – Hieraus darf allerdings nicht geschlossen werden, der zu d_∞ gehörige Mittelwert sei im Allgemeinen das arithmetische Mittel. Schon im Fall $N = 3$ scheitert diese Vermutung. Vielmehr erweist sich der fragliche Mittelwert als das sog. *Bereichsmittel* (engl. *midrange*), das arithmetische Mittel aus dem kleinsten und größten Wert der Stichprobe: $\mu = A(\mathrm{Min}(\underline{x}), \mathrm{Max}(\underline{x})) = M_{d_\infty}(\underline{x})$. Speziell gilt bei $N = 2$ Elementen: $M_{d_\infty}(x_1, x_2) = A(x_1, x_2)$. – Die Einzelheiten der einfachen Überlegungen können dem Leser überlassen bleiben.

Das Bereichsmittel eignet sich vor allem für mittig tendierende Messwerte; es minimiert die Variation unter allen Auswahlen eines Mittelpunkts. Gelegentlich wird es dazu verwendet, die durchschnittliche Tagestemperatur an einem bestimmten Ort anzugeben. Schiehlen und Seifried [11] berichten über mögliche Anwendungen auf Probleme aus der Strukturdynamik von Balken.

Betrachten wir schließlich noch den allgemeinen Fall der L_p-Metrik $d_p(\underline{x}, \underline{y}) = \|\underline{x} - \underline{y}\|_p$ für $p \geq 1$. Tatsächlich nimmt zu beliebigem, aber fest gewähltem $\underline{x} \in \mathbb{R}_+^N$ die Funktion $u \mapsto d_p(\underline{x}, \overline{u})$ ein eindeutiges Minimum $\mu = M_{d_p}(\underline{x})$ an. Allem Anschein nach lässt sich dieses jedoch nicht in expliziter geschlossener Form darstellen (vgl. dazu Hajja [6]). Für wachsendes p nähert sich $M_{d_p}(\underline{x})$ dem Bereichsmittel; ob von oben oder von unten, hängt von der Stichprobe $\underline{x}$ ab. Hajja vermutet, dass keine zwei d_p-Mittelwerte vergleichbar sind.

2 Eine Metrik für konjugierte Mittelwerte

In diesem Abschnitt fragen wir umgekehrt: Besteht ein spezifischer Zusammenhang zwischen bestimmten Klassen von Mittelwerten und den sie erzeugenden Metriken (Distanzen)? Aus welchen Metriken d lassen sich (wenn überhaupt) beispielsweise das harmonische (H), das geometrische (G) und das quadratische (Q) Mittel gewinnen?

$$H(\underline{x}) = \frac{N}{\sum_{j=1}^N \frac{1}{x_j}}, \qquad G(\underline{x}) = (\prod_{j=1}^N x_j)^{\frac{1}{N}}, \qquad Q(\underline{x}) = \sqrt{\frac{1}{N}\sum_{j=1}^N x_j^2}.$$

H, G, Q gehören (wie A selbst) zu den *quasi-arithmetischen* Mittelwerten, die im Folgenden genauer definiert werden sollen. Tatsächlich lässt sich zeigen, dass alle quasi-arithmetischen Mittel in natürlicher Weise als Minima bezüglich einer geeigneten Familie von Metriken darstellbar sind.

Wir beginnen mit einigen Definitionen und vorbereitenden Überlegungen.

Mittelwerte werden einer Stichprobe von N Elementen (hier: N-Tupel positiver reeller Zahlen, $N \geq 2$ ganz) zugeordnet, sie sind somit Werte einer Funktion (noch näher zu beschreibenden Typs) $F : \mathbb{R}_+^N \to \mathbb{R}_+$. Die Bezeichnung "mittelnde Funktion" wäre dafür durchaus angebracht, ebenso "Mittelwert-Funktion"; hingegen bezeichnen "Mittelwert" oder "Mittel" eigentlich nicht die Funktion F selbst, sondern die von ihr gelieferten Werte. Als Bezeichnung für F sind sie, da eingebürgert, zumindest *cum grano salis* zu verwenden.

Es ist sinnvoll, von einer Mittelwert-Funktion F zu verlangen, dass sie *reflexiv* ist, d. h. die Identität $F(u, \dots, u) = u$ erfüllt. Ferner sollte sie *isoton* sein, d. h. monoton wachsend in jedem ihrer Argumente. Eine reflexive und isotone Funktion F ist stets auch *internal* in dem Sinn, dass sie ihre Werte zwischen dem kleinsten und dem größten Wert einer Stichprobe annimmt: $\mathrm{Min}(\underline{x}) \leq F(\underline{x}) \leq \mathrm{Max}(\underline{x})$.

Um die engere Klasse der quasi-arithmetischen Mittelwerte definieren und kennzeichnen zu können, bedarf es noch weiterer Begriffe.

Zu gegebenen offenen Intervallen $I, J \subseteq \mathbb{R}_+$ betrachten wir die Menge $\mathcal{S}(I, J)$ aller stetigen, streng monoton wachsenden Abbildungen φ von I auf J. Ein solches φ ist bijektiv, mithin ein Homöomorphismus $I \to J$. Damit definieren wir die *φ-Konjugierte* F^φ von F wie folgt:

$$F^\varphi(x_1, \dots, x_N) := \varphi^{-1}(F(\varphi(x_1), \dots, \varphi(x_N))).$$

Bemerkung. – Die in dieser Beziehung (eine Äquivalenzrelation) stehenden Funktionen F und F^φ bezeichnet Bullen als "corresponding means" und führt sie auf Andreoli [1] zurück, während der Daróczy [4] entnommene Begriff des "conjugate mean" vom hier erklärten Begriff der φ-Konjugierten deutlich abweicht (vgl. [2], S. 422 und S. 320).

Beim Übergang von F zu einer Konjugierten F^φ bleiben viele Eigenschaften der ursprünglichen Funktion erhalten, z. B. Reflexivität, Isotonie (und damit die Eigenschaft, Mittelwert-Funktion zu sein), ferner Stetigkeit und Symmetrie (sofern sie bereits bei F vorliegen). Auch die *Assoziativität* bleibt erhalten. Definitionsgemäß liegt diese bei einer Funktion F vor, wenn sich für beliebiges ganzes p mit $1 \leq p \leq N$ die ersten p Elemente jeder Stichprobe $(x_1, \ldots, x_N)$ durch $u = F(x_1, \ldots, x_p)$ ersetzen lassen, ohne dass sich dadurch der Mittelwert der Stichprobe ändert:

$$F(u, \ldots, u, x_{p+1}, \ldots, x_N) = F(x_1, \ldots, x_N).$$

Eine Mittelwert-Funktion F heißt nun *quasi-arithmetisch*, wenn $F = A^\varphi$ für ein geeignetes $\varphi \in \mathcal{S}$ gilt (hier der Einfachheit halber ohne Angabe der Definitionsintervalle notiert). Wie Kolmogorov [7] gezeigt hat, ist dies genau dann der Fall, wenn F eine stetige, symmetrische und assoziative Mittelwert-Funktion ist.

Ohne große Mühe ist zu erkennen: $A^\varphi = Q$ für $\varphi(x) = x^2$, $A^\varphi = H$ für $\varphi(x) = \frac{1}{x}$ und $A^\varphi = G$ für $\varphi(x) = \log x$.

Auch die Potenz-Mittelwert-Funktionen P_r (mit reellem Parameter $r \neq 0$), nach Hölder durch

$$P_r(x_1, \ldots, x_N) := \left(\frac{1}{N} \sum_{j=1}^{N} x_j^r \right)^{\frac{1}{r}}$$

definiert, sind quasi-arithmetisch; man hat ja lediglich $\varphi(x) = x^r$ zu wählen. Offensichtlich ist $H = P_{-1}, A = P_1$ und $Q = P_2$. Auch der geometrische Mittelwert $G = P_0$, verstanden als Grenzfall von P_r für $r \to 0$, reiht sich unter die Potenz-Mittelwerte ein (vgl. etwa [12]).

Es soll nun gezeigt werden, dass sich jeder quasi-arithmetische Mittelwert mit der Methode der kürzesten Abstände als d-Mittelwert mit einer geeigneten Metrik d auf $\mathbb{R}^N$ gewinnen lässt. Allgemeiner gilt sogar der

Satz 1. *Existiert bezüglich der Metrik d die zugehörige Mittelwert-Funktion M_d, so lässt sich jede Konjugierte $(M_d)^\varphi$ ($\varphi \in \mathcal{S}$) als Mittelwert-Funktion bezüglich einer geeigneten Metrik $d\varphi$ darstellen, für die gilt: $(M_d)^\varphi = M_{d\varphi}$.*

Beweis. 1. Wir betrachten der Einfachheit halber Homöomorphismen des Typs $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$. Diese dehnen wir auf den Stichprobenraum $\mathbb{R}_+^N$ aus vermöge der Definition:

$$\overline{\varphi}(x_1, \ldots, x_N) := (\varphi(x_1), \ldots, \varphi(x_N)).$$

Zu gegebener Metrik d auf $\mathbb{R}_+^N$ werde nun $d\varphi : \mathbb{R}_+^N \times \mathbb{R}_+^N \to \mathbb{R}$ definiert durch

$$d\varphi(\underline{x}, \underline{y}) := d(\overline{\varphi}(\underline{x}), \overline{\varphi}(\underline{y})) \qquad (\underline{x}, \underline{y} \in \mathbb{R}_+^N).$$

Man verifiziert unmittelbar, dass $d\varphi$ eine Metrik auf $\mathbb{R}_+^N$ ist.

2. Zu gegebener Stichprobe $\underline{x} \in \mathbb{R}_+^N$ sei nun μ ein eindeutig bestimmtes Abstandsminimum bezüglich der Metrik d. Es gilt dann für alle $u \in \mathbb{R}_+$:

$$d(\underline{x}, \overline{\mu}) \leq d(\underline{x}, \overline{u}). \tag{3}$$

Da φ bijektiv ist, haben wir $\overline{\varphi}(\underline{y}) = \underline{x}$ mit eindeutig bestimmtem $\underline{y} \in \mathbb{R}_+^N$ sowie $\mu = \varphi(\mu_0)$ und $u = \varphi(u_0)$ mit eindeutig bestimmten $\mu_0, u_0 \in \mathbb{R}_+$. Damit ergibt sich aus (3)

$$\begin{aligned} d(\overline{\varphi}(\underline{y}), \overline{\varphi(\mu_0)}) &\leq d(\overline{\varphi}(\underline{y}), \overline{\varphi(u_0)}), \quad \text{und folglich} \\ d\varphi(\underline{y}, \overline{\mu_0}) &\leq d\varphi(\underline{y}, \overline{u_0}). \end{aligned} \tag{4}$$

Mit u durchläuft auch u_0 ganz $\mathbb{R}_+$, es liefert daher μ_0 einen Punkt kleinsten Abstands zu $\underline{y}$ bezüglich $d\varphi$. Für jedes $\underline{y} \in \mathbb{R}_+^N$ existiert eindeutig ein solches Minimum.

3. Da mit $\mu = M_d(\underline{x})$ auch μ_0 eindeutig ist, können wir schreiben: $\mu_0 = M_{d\varphi}(\underline{y})$. Aus der Ungleichung (4) folgt $\varphi(\mu_0) = M_d(\overline{\varphi}(\underline{y}))$, also $\mu_0 = (\varphi^{-1} \circ M_d \circ \overline{\varphi})(\underline{y})$, und wir erhalten für $M_{d\varphi}$ und die φ-Konjugierte von M_d die Beziehung

$$M_{d\varphi} = (M_d)^{\varphi}. \tag{5}$$

Mithin lässt sich die konjugierte Funktion $(M_d)^{\varphi}$ durch Minimierung bezüglich der homöomorph "verformten" Metrik $d\varphi$ gewinnen. □

Speziell für $d = d_2$ (euklidische Metrik) geht die rechte Seite von (5) in A^{φ} über, d. h. jeder quasi-arithmetische Mittelwert kann als $d_2\varphi$-Mittelwert aufgefasst werden.

Beispiel: Für die harmonische Mittelwert-Funktion H gilt $H = A^{\varphi}$ mit $\varphi(x) = 1/x$. Dann ist $d_2\varphi$ eine Metrik auf $\mathbb{R}_+^N$, und der Abstand

$$d_2\varphi(\underline{x}, \overline{u}) = \sqrt{\sum_{j=1}^{N} \Big(\frac{1}{x_j} - \frac{1}{u}\Big)^2}$$

wird genau dann minimal, wenn $u = H(\underline{x})$. Es gilt demnach $H = M_{d_2\varphi}$.

3 Eine Verallgemeinerung der Metrik d_p

Im Folgenden soll die Minimierung mit Distanzen folgender Bauart erörtert werden:

$$d(\underline{x}, \underline{y}) = \varphi^{-1}\Big(\sum_{j=1}^{N} \varphi(|x_j - y_j|)\Big). \tag{6}$$

Hierbei ist φ als eine auf $\mathbb{R}_+ \cup \{0\}$ definierte stetige und streng monoton wachsende Funktion mit $\varphi(0) = 0$ vorausgesetzt. Das in (6) erklärte d ist dann symmetrisch und positiv-definit, also eine Distanz. Ob auch die Dreiecksungleichung gilt (und damit eine Metrik vorliegt), hängt von gewissen Eigenschaften der Funktion φ ab. Im speziellen Fall $\varphi(u) = u^p$, $p \geq 1$, folgt die Dreiecksungleichung aus der Minkowski-Ungleichung für endliche Summen (vgl. etwa Bullen [2], S. 189).

Suchen wir (im allgemeinen Fall) zu vorgegebener Stichprobe $\underline{x} = (x_1, \ldots, x_N)$ deren d-Abstand zu $\mathbb{E}$, so genügt es aufgrund der Monotonie von φ (und damit auch von φ^{-1}), die Werte u mit kleinster Summe $\sum_{j=1}^{N} \varphi(|x_j - u|)$ zu bestimmen. Nach einem Satz von Nagumo [10] besitzt diese Summe ein eindeutiges absolutes Minimum genau dann, wenn φ eine auf $\mathbb{R}_+$ strikt konvexe Funktion ist. In dem Fall erweist sich die resultierende Funktion M_d zudem als eine in allen Argumenten $x_1, \ldots, x_N$ stetige Abbildung.

Die Konvexität von φ allein garantiert allerdings noch nicht die Gültigkeit der Dreiecksungleichung. Schreibt man diese für d einmal gemäß der rechten Seite von (6) aus, so wird ersichtlich, dass erst die folgende Verallgemeinerung der Minkowski-Ungleichung zum Ziel führt:

$$\Sigma^{\varphi}(\underline{x}+\underline{y}) \leq \Sigma^{\varphi}(\underline{x}) + \Sigma^{\varphi}(\underline{y}),$$

wo $\Sigma(\underline{x}) := x_1 + \cdots + x_N$. Mit anderen Worten: Die Subadditivität der φ-Konjugierten des Summenoperators Σ impliziert die Dreiecksungleichung für die durch (6) gegebene Distanz. Mulholland [8] hat nachgewiesen: Σ^{φ} ist subadditiv, wenn neben $\varphi \in \mathcal{S}$ auch $\varphi^{\exp}$ ($= \log \circ \varphi \circ \exp$) konvex ist.

Wir fassen die vorangegangenen Überlegungen zusammen:

Satz 2. *Sei $\varphi : \mathbb{R}_+ \cup \{0\} \to \mathbb{R}_+ \cup \{0\}$ ein strikt konvexer Homöomorphismus mit $\varphi(0) = 0$; ferner sei $\varphi^{\exp}$ konvex. Dann ist die durch* (6) *definierte Distanz d eine Metrik.*

Je nach Wahl von φ kann der durch (6) induzierte Mittelwert quasi-arithmetisch sein oder nicht. Der zweite Fall soll hier einmal durch das Beispiel

$$\varphi(u) := \cosh(u) - 1 \tag{7}$$

illustriert werden. Man überzeugt sich unschwer davon, dass φ die Voraussetzungen von Satz 2 erfüllt. Auswertung von (6) ergibt

$$d_h(\underline{x}, \underline{y}) := \operatorname{arcosh}(1 - N + \sum_{j=1}^{N} \cosh(x_j - y_j)),$$

wodurch aufgrund von Satz 2 eine Metrik d_h auf $\mathbb{R}_+^N$ gegeben ist (h für “hyperbolisch”). Wie im allgemeinen Fall genügt es hier, den Ausdruck $\sum_{j=1}^{N} \cosh(x_j - u)$ zu minimieren. Mittels Nullsetzen der ersten Ableitung nach u und Vorzeichenprüfung der zweiten Ableitung erhält man auf übliche Weise das eindeutige (sogar absolute) Minimum:

$$u_{\min} = \frac{1}{2} \log \frac{e^{x_1} + \cdots + e^{x_N}}{e^{-x_1} + \cdots + e^{-x_N}} =: M_{\text{hyp}}(x_1, \ldots, x_N).$$

Die so gewonnene Abbildung $M_{\text{hyp}} : \mathbb{R}_+^N \to \mathbb{R}_+$ ist, wie sich einfach verifizieren lässt, eine symmetrische, in jedem Argument stetige und isotone Mittelwert-Funktion. Ferner ist M_{hyp} nicht-assoziativ und daher auch nicht quasi-arithmetisch (außer im 2-dimensionalen Fall, wo $M_{\text{hyp}}(x_1, x_2) = A(x_1, x_2)$).

Immerhin lässt sich M_{hyp} einer umfangreichen Klasse von Mittelwert-Funktionen einordnen, welche (unter vielen anderen) auch die quasi-arithmetischen Mittelwerte enthält. Es handelt sich um die sogenannten Abweichungsmittelwerte, die folgendermaßen definiert sind: Sei $I \subseteq \mathbb{R}_+$ ein nicht-ausgeartetes Intervall. Eine Funktion $E : I \times I \to \mathbb{R}$ heißt dann *Abweichung(sfunktion)* (engl. *deviation*), wenn $E(x, x) = 0$ und $E(x, \cdot)$ im zweiten Argument stetig und streng monoton wachsend ist für alle $x \in I$. Für die kumulierte Abweichung

$$K(\underline{x}, u) := E(x_1, u) + \cdots + E(x_N, u)$$

hat man dann: $K(\underline{x}, \operatorname{Min}(\underline{x})) \leq 0 \leq K(\underline{x}, \operatorname{Max}(\underline{x}))$, und es existiert stets genau ein zwischen $\operatorname{Min}(\underline{x})$ und $\operatorname{Max}(\underline{x})$ gelegenes $\mu \in I$ mit $K(\underline{x}, \mu) = 0$. Es ist $\mu =: D_E(x_1, \ldots, x_N)$ der von Daróczy [3] eingeführte *Abweichungsmittelwert* (engl. *deviation mean*) der Stichprobe $\underline{x} = (x_1, \ldots, x_N)$.

Wählen wir zu gegebenem φ speziell $E(x,u) = \varphi(u) - \varphi(x)$, so ergibt sich

$$K(\underline{x}, u) = N \cdot \varphi(u) - \sum_{j=1}^{N} \varphi(x_j)$$

und somit $D_E(\underline{x}) = A^{\varphi}(\underline{x})$. Die quasi-arithmetischen Mittelwerte erweisen sich daher als Unterklasse der Abweichungsmittelwerte. Auch führt von hier aus (zumindest in gewissen Fällen) ein Weg zu einer Metrik, welche das jeweilige quasi-arithmetische Mittel erzeugt. Dazu berechnet man für die einzelnen Stichprobenwerte x_j das Integral von $E(x_j, t)$ über dem Intervall $[x_j, u]$ und summiert anschließend über alle $j = 1, \ldots, N$. Beispielsweise ergibt sich so für $\varphi(u) = u$:

$$\sum_{j=1}^{N} \int_{x_j}^{u} E(x_j, t)\, dt = \sum_{j=1}^{N} \int_{x_j}^{u} (t - x_j)\, dt = \frac{1}{2} \sum_{j=1}^{N} (x_j - u)^2 = \frac{1}{2} f(u),$$

wo f die Funktion aus Gleichung (1) ist, deren Minimierung zum d_2-Mittelwert A führt.

Tatsächlich ist auch der hyperbolische Mittelwert M_{hyp} ein Abweichungsmittelwert. Das erkennt man, wenn $E(x,u) = \sinh(u - x)$ als Abweichungsfunktion gewählt wird. In diesem Fall liefert die Summe der Integrale

$$\int_{x_j}^{u} \sinh(u - t)\, dt = \cosh(x_j - u) - 1 \quad (j = 1, \ldots, N)$$

das Argument von φ^{-1} in (6) (mit dem φ aus (7)), mithin den Ausdruck, der für die Ermittlung des kleinsten (mit d_h gemessenen) Abstands zu minimieren ist.

Literatur

[1] Andreoli, G.: Aspetto gruppale e funzionale delle medie, *Giorn. Mat. Battaglini* 5/85 (1957), 12–30.

[2] Bullen, P. S.: *Handbook of Means and Their Inequalities.* Kluwer Academic Publishers, Dordrecht 2003.

[3] Daróczy, Z.: Über eine Klasse von Mittelwerten, *Publ. Math. Debrecen* 19 (1972), 211–217.

[4] Daróczy, Z.: On a class of means of two variables, *Publ. Math. Debrecen* 55 (1999), 177–197.

[5] Freudenthal, H., Steiner, H.-G.: Aus der Geschichte der Wahrscheinlichkeitsrechnung und der mathematischen Statistik. In: Behnke, H., Bertram, G., Sauer, R. (Hrsg.), *Grundzüge der Mathematik, Band 4*, Vandenhoeck & Ruprecht, Göttingen 1966.

[6] Hajja, M.: Some elementary aspects of means. *Int. Journ. of Mathematics and Mathematical Sciences* (2013), Article ID 689560, DOI.org/10.1155/689560.

[7] Kolmogorov, A. N.: Sur la notion de la moyenne. *Atti Accad. Naz. Lincei. Rend.* 12/9 (1930), 388–391.

[8] Mulholland, H. P.: On generalizations of Minkowski's inequality in the form of a triangle inequality. *Proc. London Math. Soc.* 51/2 (1950), 294–307.

[9] Nagumo, M.: Über eine Klasse der Mittelwerte. *Japan. J. Math* 7 (1930), 71–79.

[10] Nagumo, M.: Über den Mittelwert, der durch die kleinste Abweichung definiert wird, *Japan. J. Math.* 10 (1933), 53–56.

[11] Schiehlen, W., Seifried, R.: Impacts on beams. Uncertainty in experiments and numerical simulation. Chapter 9 in: Papadrakakis, M., et al. (eds.): *Computational Structural Dynamics and Earthquake Engineering*, CRC Press, 2008.

[12] Schreiber, A.: Vergleichbarkeit quasi-arithmetischer Mittel am Beispiel von Potenz- und Exponential-Mittelwerten. *Math. Semesterber.* 65/2 (2018),171-182. DOI.org/10.1007/s00591-017-0207-2.

Alfred Schreiber
Abteilung für Mathematik und ihre Didaktik
Europa-Universität Flensburg
Auf dem Campus 1
D-24943 Flensburg
e-mail: `info@alfred-schreiber.de`