Behandlung einiger Extremalprobleme für Wahrscheinlichkeitsverteilungen

Vom Fachbereich Mathematik
der Universität Hannover
zur Erlangung des Grades
Doktor der Naturwissenschaften
Dr. rer. nat.
genehmigte Dissertation
von

Dipl.-Math. LUTZ MATTNER
geboren am 30. 11. 1961 in Hannover

1990

Referent: PD Dr. N. Henze
Korreferent: Prof. Dr. D. Plachky
Tag der Promotion: 20.06.1990

Lutz Mattner

Behandlung einiger Extremalprobleme für Wahrscheinlichkeitsverteilungen

Inhaltsverzeichnis

1 Einleitung

1.1 Ein Satz von Plackett

Den Ausgangspunkt dieser Arbeit bildet der

Satz 1 (Plackett) *Sind X und Y unabhängige und identisch verteilte reellwertige Zufallsvariablen mit*

$$E[X] = 0, \qquad Var(X) = 1, \tag{1}$$

so ist

$$E[|X - Y|] \leq \frac{2}{\sqrt{3}}. \tag{2}$$

Gleichheit tritt genau dann auf, wenn X im Intervall $[-\sqrt{3}, \sqrt{3}]$ gleichverteilt ist.

Es handelt sich dabei um einen Spezialfall allgemeinerer Resultate von Plackett [21], Moriguti [19], Gumbel [14] und Hartley und David [15], die zusammen mit ihren Beweismethoden kurz erläutert seien.

Plackett betrachtet unabhängige und identisch verteilte Zufallsvariablen $X_1, \ldots, X_n$ mit Werten in einem Intervall $[-a, a]$ und fragt nach dem Maximum des Verhältnisses von mittlerer Spannweite zur Standardabweichung

$$d_n = \frac{E[X_{(n)} - X_{(1)}]}{\sqrt{Var(X_1)}} \quad ,$$

wobei $X_{(1)} = \min(X_1, \ldots, X_n)$ und $X_{(n)} = \max(X_1, \ldots, X_n)$ sei. Er nimmt die Existenz einer extremalen Verteilung F_0 an und setzt in der Darstellung

$$d_n = \frac{\int_{-a}^{a} 1 - F^n(x) - (1 - F(x))^n \, dx}{\left(\int_{-a}^{a} x^2 \, dF(x) - \left(\int_{-a}^{a} x \, dF(x)\right)^2\right)^{\frac{1}{2}}} \tag{3}$$

(mit $F =$ Verteilungsfunktion von X_1) $F(x) = F_0(x) + tv(x)$, wobei v eine beliebige stetige Funktion mit $v(-a) = v(a) = 0$ und t eine reelle Zahl bedeutet. Es wird dann behauptet, daß F für hinreichend kleine t eine Verteilungsfunktion ist (das setzt offenbar voraus, daß F_0 in $[-a, a]$ streng wächst) und die Ableitung von d_n an der Stelle 0 gleich 0 gesetzt. Aus der so erhaltenen Bedingung wird F_0 bestimmt und mit dem Hinweis, daß d_n kein Minimum besitzt, geschlossen, daß bei F_0 ein Maximum vorliegen muß. Für $n = 2$ ist F_0 Verteilungsfunktion der Gleichverteilung auf $[-a, a]$ und es ergibt sich $d_n \leq \frac{2}{\sqrt{3}}$. Durch Grenzübergang wird gefolgert, daß diese Ungleichung (und damit die Implikation (1) $\Rightarrow$ (2)) für Verteilungen mit beliebigem Träger gilt.

In der Arbeit von Gumbel sollte $E[X_{(n)}]$, der Erwartungswert des Maximums von n unabhängigen und standardisierten Zufallsvariablen $X_1, \ldots, X_n$, maximiert werden. Für $n = 2$ ist wegen

$$E[X_{(2)} - X_{(1)}] = E[2X_{(2)} - X_1 - X_2] = 2E[X_{(2)}] \tag{4}$$

dieses Problem mit dem Plackettschen äquivalent. Gumbel schreibt

$$E[X_{(n)}] = \int_0^1 nF^{-1}(u)u^{n-1}\,du, \tag{5}$$

$$E[X] = \int_0^1 F^{-1}(u)\,du, \qquad E[X^2] = \int_0^1 \left(F^{-1}(u)\right)^2 du, \tag{6}$$

wendet auf das zu behandelnde Extremalproblem („Maximiere (5) als Funktional von F^{-1} unter den genannten Nebenbedingungen an (6)") formal die Lagrangesche Multiplikatorenregel an und erhält im Fall $n = 2$ die Behauptung des Satzes 1. Über die Existenz einer Lösung sowie über die eventuell zu berücksichtigende Nebenbedingung, daß F^{-1} wachsend sein muß, verliert er kein Wort.

Den ersten strengen Beweis des Satzes 1 findet man wohl in der Arbeit von Moriguti (Fußnote 5 auf Seite 534) oder, expliziter, bei Hartley und David. Unter Verzicht auf unnötige Regularitätsvoraussetzungen läßt er sich wie folgt formulieren:

Beweis des Satzes 1 Wir schreiben X_1, X_2 an Stelle von X, Y und bezeichnen mit F die Verteilungsfunktion von X. Sei $F^{-1} : [0,1] \to \mathbb{R}$, $F^{-1}(u) := \inf\{x \in \mathbb{R} : F(x) \geq u\}$, ihre verallgemeinerte Umkehrfunktion (Quantilfunktion). Dann gilt, wenn U_1, U_2 unabhängig und jeweils gleichverteilt im Intervall $[0,1]$ sind, die Verteilungsgleichheit

$$\left(F^{-1}(U_1), F^{-1}(U_2)\right) \stackrel{\mathcal{D}}{=} (X_1, X_2) \tag{7}$$

und daher auch

$$F^{-1}(U_{(2)}) \stackrel{\mathcal{D}}{=} X_{(2)}. \tag{8}$$

Aus (1) und (7) folgt

$$\int_0^1 F^{-1}(u)\,du = 0, \quad \int_0^1 \left(F^{-1}(u)\right)^2 du = 1 \tag{9}$$

und (4) und (8) ergeben, zusammen mit der Dichte $f_{U_{(2)}}(u) = 2u$,

$$\begin{aligned} E[|X_1 - X_2|] &= 2E[F^{-1}(U_{(2)})] \\ &= 2\int_0^1 F^{-1}(u)2u\,du \\ &= 2\int_0^1 F^{-1}(u)\,(2u-1)\,du \\ &\leq 2\left(\int_0^1 \left(F^{-1}(u)\right)^2 du\right)^{\frac{1}{2}} \left(\int_0^1 (2u-1)^2\,du\right)^{\frac{1}{2}} \\ &= \frac{2}{\sqrt{3}}. \end{aligned}$$

Beim dritten und beim letzten Gleichheitszeichen wurde (9) benutzt. Nun tritt das Gleichheitszeichen in der Schwarzschen Ungleichung nur auf, wenn

$$F^{-1}(u) = m(2u-1) \qquad (u \in [0,1])$$

für ein positives m gilt. Das aber führt auf die behauptete Gleichverteilung. ∎

Ein neuer Beweis wurde von Baringhaus und Henze [4] gegeben. Er basiert auf einem Satz von Terrell [30] (später verallgemeinert in [29]), welcher mit der (laut David [8] von Sugiura in [27,28] eingeführten) Methode der Entwicklung der inversen Verteilungsfunktion nach orthogonalen Funktionen bewiesen wurde.

Baringhaus und Henze zeigten weiter, wie man mit Hilfe von Satz 1 einen Test für das statistische Testproblem „Bei Vorliegen von n unabhängigen und identisch verteilten Zufallsvariablen $(X_1, \ldots, X_n)$ teste man die Hypothese, daß die X_i in einem (unbekannten) Intervall gleichverteilt sind, gegen die allgemeine Alternative " aufstellen kann. Der Test basiert auf der naheliegenden Schätzung für d_2 und der Berechnung ihrer asymptotischen Verteilung mittels geeigneter Grenzwertsätze für U–Statistiken.

1.2 Fragestellungen dieser Arbeit

Die oben referierten Ergebnisse und Methoden legen nun verschiedene Fragen nahe. Zunächst kann man nach „eindimensionalen" Verallgemeinerungen des Satzes 1 fragen:

Frage 1 *Kann man andere eindimensionale Verteilungen als die Gleichverteilung als Lösungen ähnlicher Extremalprobleme charakterisieren? Existiert etwa zu gegebener Verteilungsfunktion F eine Funktion $K : \mathbb{R}^2 \to \mathbb{R}$ derart, daß $E[K(X,Y)]$ für unabhängige und identisch verteilte Zufallsvariablen X, Y unter geeigneten Nebenbedingungen an ihre Momente genau dann maximal wird, wenn sie nach F verteilt sind?*

Diese Frage mag auf den ersten Blick künstlich erscheinen. Könnte man jedoch ein K wie oben stets finden, so könnte man versuchen, die Vorgehensweise der Arbeit von Baringhaus und Henze auf das Testen eines beliebigen vorgegebenen Typs zu übertragen.

Weiter ist es interessant, nach „mehrdimensionalen" Analoga des Satzes 1 zu fragen. Faßt man ihn als Charakterisierung der Gleichverteilung auf, so stellt sich etwa

Frage 2 *Gibt es eine Funktion $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, derart, daß für unabhängige und identisch verteilte Zufallsvektoren X, Y mit Werten im $\mathbb{R}^d$ der Erwartungswert $E[K(X,Y)]$ unter den Nebenbedingungen*

$$E[X] = 0, \qquad E\left[|X - E[X]|^2\right] = 1 \tag{10}$$

genau dann maximal wird, wenn X und Y gleichverteilt in einer Kugel um den Ursprung sind?

Faßt man dagegen Satz 1 als Bestimmung des Maximums eines interessanten Funktionals auf, so gelangt man zur

Frage 3 *Wie groß kann $E[|X - Y|]$, der Erwartungswert des euklidischen Abstandes zweier unabhängiger und identisch verteilter Zufallspunkte, unter der Nebenbedingung (10) höchstens werden und bei welcher Verteilung wird er maximal?*

Nicht zuletzt jedoch erscheint es lohnend, sich methodisch mit Satz 1 (oder ähnlichen Problemen) auseinanderzusetzen, da alle in der Einleitung erwähnten Beweismethoden ihre Nachteile haben:

- Die Vorgehensweisen von Plackett und Gumbel werden wohl in der Regel nicht zu falschen Ergebnissen führen, können aber einen strengen Beweis nicht ersetzen.
- Ein Beweis wie bei Moriguti bzw. Hartley und David setzt die Kenntnis einer mutmaßlichen Lösung voraus sowie das Glück und Geschick, eine Standard-Ungleichung anwenden zu können.
- Die Methode der Entwicklung der inversen Verteilungsfunktion nach orthogonalen Funktionen setzt die Wahl des „richtigen" Systems voraus (in [30] die Legendre-Polynome, in [29] gewisse Jacobi-Polynome).
- Allen drei Methoden ist gemeinsam, daß sie stark von der Eindimensionalität des jeweiligen Problems Gebrauch machen.

In dieser Arbeit wird gezeigt, daß man Extremalprobleme der geschilderten Art, insbesondere die Fragen 1–3, mit Methoden der allgemeinen Theorie der Extremalaufgaben (für die wir auf Kapitel 1 des Buches von Tichomirov [31] verweisen) erfolgreich behandeln kann. Dabei wird eine im nächsten Kapitel anzugebende Multiplikatorenregel für glatte Extremalaufgaben in Kegeln mit Nebenbedingungen benutzt sowie, falls möglich und hilfreich, die Konvexität des jeweiligen Problems ausgenutzt.

1.3 Ergebnisse dieser Arbeit

In Kapitel 2 wird zunächst eine abstrakte Multiplikatorenregel (Satz 2) hergeleitet, welche eine naheliegende Verallgemeinerung eines Satzes aus [31] ist. Die durch Spezialisierung gewonnene Multiplikatorenregel (Satz 3) ist auf ziemlich allgemeine Extremalprobleme für Wahrscheinlichkeitsverteilungen anwendbar und stellt einen Fortschritt gegenüber den in [23] behandelten Methoden dar. Für die in dieser Arbeit vorgesehenen Anwendungen wird sie in Satz 4 gesondert formuliert. Zur Illustration wird für den Satz von Plackett mit Hilfe des Satzes 4 und einer sich aus dem Lemma von Helly ergebenden Existenzaussage (Lemma 1) ein neuer Beweis gegeben.

Satz 5 in Kapitel 3 ist die „konvexe Variante" des Satzes 4, welche sich unabhängig von Kapitel 2 direkt aus dem Satz von Kuhn–Tucker ergibt, wenn das betrachtete Extremalproblem konvex ist. Ein erneuter Beweis des Satzes von Plackett illustriert die Vorzüge des Satzes 5 gegenüber Satz 4 im Fall konvexer Extremalprobleme. Satz 6 gibt, in Verfeinerung eines Satzes aus der Theorie der negativ definiten Funktionen, eine Klasse streng konvexer Funktionale von Wahrscheinlichkeitsverteilungen an.

In Kapitel 4 wird die Frage 1 dieser Einleitung auf zwei verschiedene Weisen beantwortet: Satz 7 enthält die überraschende Aussage, daß man im allgemeinen nicht $K(x,y) = u(x-y)$ mit einer geeigneten Funktion u setzen kann, während Satz 8 zu jeder hinreichend regulären Verteilungsfunktion einen Kern der Form $K(x,y) = |h(x) - h(y)|$ mit einer gewissen Funktion h angibt.

Kapitel 5 gibt eine Antwort auf Frage 2: Die Gültigkeit des Satzes von Plackett beruht darauf, daß die Betragsfunktion bis auf einen konstanten Faktor Grundlösung des eindimensionalen Laplaceschen Differentialoperators $\frac{d^2}{dx^2}$ ist. Entsprechend erhält man Charakterisierungen von Gleichverteilungen in höheren Dimensionen mit den Kernen $K(x,y) = \log|x-y|$ für $d = 2$ und $K(x,y) = -|x-y|^{2-d}$ für $d \geq 3$ (Satz 9).

In Kapitel 6 wird Frage 3 beantwortet. Für $d \geq 3$ ergibt sich als Extremalverteilung eine ausgeartete Verteilung, nämlich die Gleichverteilung auf der Oberfläche der Einheitskugel (Satz 10), während für $d = 2$ eine aus der 3-dimensionalen Potentialtheorie bekannte absolut stetige Verteilung extremal ist (Satz 11).

In Kapitel 7 wird eine im eindimensionalen Fall wohlbekannte Charakterisierung der Gleichverteilung durch maximale Varianz innerhalb einer Klasse unimodaler Verteilungen auf höhere Dimensionen übertragen (Satz 12). Methodisch gesehen ist dieses Kapitel unabhängig von den übrigen der vorliegenden Arbeit.

Kapitel 8 schließlich enthält einen eher didaktischen Beitrag, nämlich eine nach Ansicht des Autors besonders natürliche Herleitung der in der Kapitelüberschrift genannten Sätze. Die Verbindung zu den anderen Kapiteln dieser Arbeit besteht darin, daß der erstgenannte Satz von Bernstein beim Beweis des Satzes 6 aus Kapitel 3 benötigt wird.

1.4 Bezeichnungen und Vereinbarungen

$\mathbb{N} = \{1, 2, \ldots\}$, $\mathbb{Q}$, $\mathbb{R}$ und $\mathbb{C}$ bezeichnen die Mengen der natürlichen, rationalen, reellen und komplexen Zahlen. Für $x \in \mathbb{R}$ wird $x_+ := \max\{0, x\}$ und $x_- := \max\{0, -x\}$ gesetzt.

Für $x, y \in \mathbb{R}^d$, dem euklidischen Raum der Dimension d, bezeichnet $x'y$ das übliche Skalarprodukt und $|x| := \sqrt{x'x}$ die euklidische Norm. Stetigkeit von Funktionen mit Werten in $[-\infty, \infty] = \mathbb{R} \cup \{-\infty, +\infty\}$ ist bezüglich der üblichen Topologie des Wertebereiches zu verstehen und zieht somit nicht die Endlichkeit nach sich.

Wenn keine Mißverständnisse zu befürchten sind, wird gelegentlich von der Funktion $f(x)$

anstatt von der Funktion f gesprochen.

Ist M eine Menge und $A(x)$ für jedes $x \in M$ eine Aussage, so bedeutet „$A(x)$ $(x \in M)$", daß $A(x)$ für jedes $x \in M$ wahr ist, und $I(A(x))$ steht für 1, wenn $A(x)$ wahr ist und für 0, wenn $A(x)$ falsch ist.

Es werden allgemein übliche Sprech- und Schreibweisen der Stochastik verwendet (siehe etwa [6], [11] und [13]). Mit Maßen auf $\mathbb{R}^d$ (oder Teilmengen des $\mathbb{R}^d$), meist mit μ oder ν bezeichnet, sind stets Borel-Maße gemeint. δ_a bezeichnet das Dirac-Maß im Punkt a und speziell $\delta := \delta_0$ das im Ursprung. Den Träger eines Maßes μ bezeichnen wir mit supp μ. Unter einem signierten Maß, meist σ genannt, verstehen wir in dieser Arbeit eine Mengenfunktion, welche sich als Differenz zweier endlicher Maße schreiben läßt. Mit $\int \ldots dx$ ist ein Integral bezüglich des Lebesgue-Maßes gemeint. Dichten sind immer als Dichten bezüglich des Lebesgue-Maßes zu verstehen.

Das Ende eines Beweises wird durch das Zeichen $\blacksquare$ markiert.

Die Sätze 1 und 9–11 sind der Einfachheit halber „normalisiert" geschrieben. Auf die einfache Übersetzung in eine „invariante" Form (im Fall des Satzes 1 ergibt sich diese durch Weglassen der Nebenbedingung (1) und Ersetzung der Behauptung (2) durch $E[|X-Y|] \leq \frac{2}{\sqrt{3}}\sqrt{Var(X)}$ bei entsprechender Modifikation der Diskussion des Gleichheitszeichens) wird verzichtet.

1.5 Dank

Diese Arbeit entstand in der anregenden Atmosphäre des Instituts für Mathematische Stochastik der Universität Hannover. Besonders danke ich Herrn Privatdozent Dr. N. Henze und Herrn Privatdozent Dr. L. Baringhaus für Rat und Unterstützung.

2 Eine Multiplikatorenregel für glatte Extremalaufgaben in Kegeln mit Nebenbedingungen

2.1 Motivation

Wir betrachten das Extremalproblem, welches zu Satz 1 führt, in der folgenden Form. Es soll

$$\int_{\mathbb{R}} \int_{\mathbb{R}} |x - y| \, d\mu(x) \, d\mu(y) \tag{11}$$

maximiert werden unter allen Wahrscheinlickeitsmaßen μ, die

$$\int_{\mathbb{R}} x \, d\mu(x) = 0, \qquad \int_{\mathbb{R}} x^2 \, d\mu(x) = 1 \tag{12}$$

genügen. Es handelt sich also um ein Maximierungsproblem unter Nebenbedingungen. Läßt man in (11) signierte Maße μ zu, so erhält man ein Extremalproblem in einem Vektorraum und kann mit Hilfe der Differentialrechnung versuchen, Bedingungen für die Lösungen aufzustellen. Man hat dann aber die Nebenbedingungen

$$\mu(\mathbb{R}) = 1 \tag{13}$$

$$\mu \geq 0 \tag{14}$$

hinzuzufügen, die gerade besagen, daß μ ein normiertes, nichtnegatives Maß ist. Während (12) und (13) Nebenbedingungen vom Gleichungstyp sind, handelt es sich bei (14) um eine Nebenbedingung vom Ungleichungstyp, und zwar um eine dem Wesen nach unendlichdimensionale[1]. Gleichwohl gibt es eine Form der Lagrangeschen Multiplikatorenregel, welche hier anwendbar ist. Für das folgende verweisen wir auf Kapitel 1 von [31] und übernehmen dessen Bezeichnungen. Dort wird in Satz A' (Seite 57) ein allgemeines Extremalproblem (A') behandelt[2], in welchem endlich viele Nebenbedingungen vom Ungleichungstyp

$$\varphi_j(z, w) \leq 0, \quad j = 1, \ldots, m \tag{15}$$

mit reellwertigen Funktionen φ_j vorkommen dürfen. Schreibt man (15) in der Form

$$\varphi(z, w) = (\varphi_1(z, w), \ldots, \varphi_m(z, w)) \in -C \tag{16}$$

mit $C := [0, \infty)^m$ = positiver Kegel in $\mathbb{R}^m$, so drängt sich die im folgenden beschriebene Verallgemeinerung der Aufgabe (A') auf.

[1] Was nicht heißen soll, daß man sie nicht durch endlich viele Ungleichungen $\varphi_i(\mu) \leq 0$ mit reellwertigen φ_i ausdrücken könnte. Zum Beispiel ist ja $\varphi_1(\mu) := \mu_-(\mathbb{R}) \leq 0$, wo μ_- den negativen Teil von μ bezeichnet, äquivalent zu den unendlich vielen Ungleichungen $\mu(B) \geq 0$, B Borelmenge, die man üblicherweise als Definition von (14) auffassen würde. φ_1 hat jedoch den Nachteil, bezüglich keiner vernünftigen Nomierung des Raumes der signierten Maße differenzierbar zu sein (weil andernfalls die reelle Funktion $t \to t_- = (t\delta(\mathbb{R}))_-$ auch differenzierbar sein müßte).

[2] Es sei auf Druckfehler hingewiesen: Auf Seite 57 in Zeile 2 muß es heißen: ,,Es seien Z und Y ... ". Weiter fehlen auf manchen Buchstaben Dächer: In der Voraussetzung d) des Satzes A' muß es Im $\Phi_z(\hat{z}, \hat{w})$ heißen, in der Behauptung 4) desselben Satzes $\hat{\lambda}_j$.

2.2 Eine allgemeine Multiplikatorenregel

Es seien Z und Y normierte lineare Räume über dem Körper $\mathbb{R}$ (Verwechselungen mit Zufallsvariablen, die auch Y heißen können, sind aufgrund des jeweiligen Kontexts nicht zu befürchten). V eine Umgebung in Z, W eine beliebige Menge. Weiter sei E ein normierter linearer Raum und C ein konvexer Kegel in E. Schließlich seien

$$\begin{aligned} \varphi_0 &: V \times W \to \mathbb{R} \\ \varphi &: V \times W \to E \\ \Phi &: V \times W \to Y \end{aligned}$$

Funktionen. Es soll die Aufgabe (A")

$$\varphi_0(z, w) \to inf;$$

$$\Phi(z, w) = 0, \quad \varphi(z, w) \in -C, \quad (z, w) \in V \times W \tag{17}$$

untersucht werden.

Definition *Ein Paar (z, w) heißt* zulässig *für die Aufgabe (A"), wenn es (17) erfüllt. Ein Paar $(\hat{z}, \hat{w})$ liefert ein* starkes lokales Minimum *für (A"), wenn es zulässig für (A") ist und es ein $\epsilon > 0$ gibt, so daß für jedes zulässige Paar (z, w) gilt:*

$$||z - \hat{z}|| < \epsilon \Rightarrow \varphi_0(z, w) \geq \varphi_0(\hat{z}, \hat{w}). \tag{18}$$

Die Funktion

$$\mathcal{L} : (V \times W) \times Y^* \times E^* \times \mathbb{R} \to \mathbb{R}$$

$$\mathcal{L}((z, w), y^*, \lambda, \lambda_0) := \lambda_0 \varphi_0(z, w) + \langle \lambda, \varphi(z, w) \rangle + \langle y^*, \Phi(z, w) \rangle$$

heißt Lagrange-Funktion *der Aufgabe (A"). Dabei bezeichnen Y^* und E^* die zu Y und E dualen Räume.*

In Verallgemeinerung des Satzes A' aus [31] gilt dann der

Satz 2 (Satz A") *In der Aufgabe (A") sei $(\hat{z}, \hat{w})$ zulässig und es seien die folgenden Bedingungen erfüllt:*

a) Y, Z und E sind Banachräume;

b) Die Funktionen $z \to \varphi_0(z, w)$, $z \to \varphi(z, w)$ und $z \to \Phi(z, w)$ sind für jedes $w \in W$ im Punkt $\hat{z}$ streng differenzierbar[3]*;*

c) Für jedes $z \in V$ und beliebige $w_1, w_2 \in W$ und $p \in [0, 1]$ existiert ein $w_p = w_p(z, w_1, w_2)$ mit

$$\Phi(z, w_p) = p\Phi(z, w_1) + (1 - p)\Phi(z, w_2)$$

[3] Siehe [31, Seite 33]. Jede in einer Umgebung eines Punktes stetig Fréchet-differenzierbare Funktion ist in diesem Punkt streng differenzierbar.

$$\varphi_0(z,w_p) \le p\varphi_0(z,w_1) + (1-p)\varphi_0(z,w_2) \tag{19}$$
$$\varphi(z,w) - p\varphi(z,w_1) - (1-p)\varphi(z,w_2) \in -C;$$

d) Das Bild von[4] $(\varphi_z(\hat{z},\hat{w}), \Phi_z(\hat{z},\hat{w}))$ *ist in* $E \times Y$ *ein abgeschlossener Teilraum endlicher Kodimension.*

Liefert dann $(\hat{z},\hat{w})$ *ein starkes lokales Minimum der Aufgabe (A"), so existieren* $\hat{\lambda}_0 \in \mathbb{R}$, $\hat{\lambda} \in E^*$ *und* $\hat{y}^* \in Y^*$ *mit*

0) $\hat{\lambda}_0$, $\hat{\lambda}$ *und* $\hat{y}^*$ *sind nicht zugleich* 0;

1) Stationaritätsbedingung bezüglich z:

$$\mathcal{L}_z\left((\hat{z},\hat{w}),\hat{y}^*,\hat{\lambda},\hat{\lambda_0}\right) = 0;$$

2) Minimumprinzip bezüglich w:

$$\min_{w\in W} \mathcal{L}\left((\hat{z},w),\hat{y}^*,\hat{\lambda},\hat{\lambda_0}\right) = \mathcal{L}\left((\hat{z},\hat{w}),\hat{y}^*,\hat{\lambda},\hat{\lambda_0}\right);$$

3) Nichtnegativitätsbedingungen:

$$\lambda_0 \ge 0, \qquad \langle\hat{\lambda}, c\rangle \ge 0 \quad \textit{für } c \in C$$

4) Komplementäre Schlupfbedingung:

$$\langle\hat{\lambda}, \varphi(\hat{z},\hat{w})\rangle = 0.$$

Beweis Wie in [31] führen wir diesen Satz auf den Satz A (Seite 49 in [31]) zurück, indem wir letzteren auf eine geignete Aufgabe (A) anwenden.

1. Definition der Aufgabe (A): Sei

$$X := Z \times \mathbb{R} \tag{20}$$
$$\tilde{V} := V \times \mathbb{R} \tag{21}$$
$$\tilde{Y} := Y \times \mathbb{R} \times E. \tag{22}$$

Dabei denken wir uns das kartesische Produkt normierter linearer Räume jeweils etwa durch die Maximumsnorm normiert. Die Elemente von X bezeichnen wir mit

$$x = (z,\xi), \qquad z \in Z, \quad \xi \in \mathbb{R}.$$

Weiter sei

$$U := W \times [0,\infty) \times C.$$

[4] φ_z und Φ_z stehen jeweils für die partiellen Fréchet-Ableitungen nach der ersten Variablen.

Die Elemente von U bezeichnen wir mit

$$u = (w, \alpha_0, \alpha), \qquad w \in W, \quad \alpha_0 \in [0, \infty), \quad \alpha \in C.$$

Wir definieren

$$f_0 : \tilde{V} \to \mathbb{R};$$
$$f_0(x) = f_0(z, \xi) := \xi$$

sowie

$$F : \tilde{V} \times U \to \tilde{Y};$$

$$\begin{aligned} F(x, u) &= F((z, \xi), w, \alpha_0, \alpha) \\ &:= (\Phi(z, w), \varphi_0(z, w) - \xi + \alpha_0, \varphi(z, w) + \alpha) \end{aligned}$$

und betrachten die Aufgabe (A)

$$f_0(x) \to \inf;$$
$$F(x, u) = 0, \quad x \in \tilde{V}, \quad u \in U,$$

die also ausgeschrieben lautet:

$$\begin{gathered} \xi \to \inf; \\ \Phi(z, w) = 0, \quad \varphi_0(z, w) - \xi + \alpha_0 = 0, \quad \varphi(z, w) + \alpha = 0, \\ z \in V, \quad \xi \in \mathbb{R}, \quad w \in W, \quad \alpha_0 \geq 0, \quad \alpha \in C. \end{gathered} \tag{23}$$

2. Angabe eines starken lokalen Minimums $(\hat{x}, \hat{u})$ der Aufgabe (A):

Sei $(\hat{z}, \hat{w})$ das uns gegebene starke lokale Minimum der Aufgabe (A"). Wir setzen

$$\hat{\xi} := \varphi_0(\hat{z}, \hat{w}) \tag{24}$$

$$\hat{x} := (\hat{z}, \hat{\xi}) = (\hat{z}, \varphi_0(\hat{z}, \hat{w}))$$
$$\hat{u} := (\hat{w}, 0, -\varphi(\hat{z}, \hat{w})).$$

Aus der Zulässigkeit von $(\hat{z}, \hat{w})$ für (A") folgt offenbar (man vergleiche (23) mit (17) unter Beachtung von (24)) die Zulässigkeit von $(\hat{x}, \hat{u})$ für (A).

Um zu zeigen, daß $(\hat{x}, \hat{u})$ ein starkes lokales Minimum für (A) liefert, sei $\epsilon > 0$ gemäß Definition 1 gewählt. Ist dann $(x, u) = ((z, \xi), w, \alpha_0, \alpha)$ zulässig für (A) mit

$$||x - \hat{x}|| < \epsilon, \tag{25}$$

so ist erst recht

$$||z - \hat{z}|| < \epsilon. \tag{26}$$

Weiter ist das Paar (z, w) wegen

$$\Phi(z, w) = 0$$
$$\varphi(z, w) = -\alpha \in -C$$

$$(z, w) \in V \times W$$

zulässig für (A"). Daraus folgt mit (26)

$$\varphi_0(z, w) \geq \varphi_0(\hat{z}, \hat{w})$$

und folglich

$$f_0(x) = \xi = \varphi_0(z, w) + \alpha_0 \geq \varphi_0(z, w) \geq \varphi_0(\hat{z}, \hat{w}) = f_0(\hat{x}).$$

Wir haben also gezeigt, daß es ein $\epsilon > 0$ gibt, so daß die Zulässigkeit von (x, u) zusammen mit (25) das Bestehen von

$$f_0(x) \geq f_0(\hat{x})$$

zur Folge hat, d.h. $(\hat{x}, \hat{u})$ ist starkes lokales Minimum von (A).

3. Nachprüfen der Voraussetzungen des Satzes A:

a) X und $\tilde{Y}$ sind Banachräume, da Produkte von solchen.

b) f_0 ist trivialerweise streng differenzierbar. Aufgrund der Voraussetzungen an φ_0, φ und Φ ist auch $x \rightarrow F(x, u)$ für jedes $u \in U$ an der Stelle $\hat{x}$ streng differenzierbar.

c) Es seien $x \in \tilde{V}$, $u_1, u_2 \in U$ und $p \in [0,1]$ gegeben. Dann ist $x = (z, \xi)$ mit $z \in V$ und $u_i = (w_i, \alpha_{0i}, \alpha_i)$, $i = 1, 2$ sowie $w_i \in W$. Nach Voraussetzung existiert ein $w_p = w_p(z, w_1, w_2) \in W$, so daß (19) erfüllt ist. Setze

$$u_p := (w_p, \alpha_{0p}, \alpha_p)$$

mit

$$\alpha_{0p} := p\alpha_{01} + (1-p)\alpha_{02} + p\varphi_0(z, w_1) + (1-p)\varphi_0(z, w_2) - \varphi_0(z, w_p) \geq 0$$

und

$$\alpha_p := p\varphi(z, w_1) + (1-p)\varphi(z, w_2) - \varphi(z, w_p) + p\alpha_1 + (1-p)\alpha_2 \in C.$$

Damit ist $u_p \in U$ und

$$\begin{aligned} pF(x, u_1) + (1-p)F(x, u_2) &= p\,(\Phi(z, w_1), \varphi_0(z, w_1) - \xi + \alpha_{01}, \varphi(z, w_1) + \alpha_1) \\ &\quad +(1-p)\,(\Phi(z, w_2), \varphi_0(z, w_2) - \xi + \alpha_{02}, \varphi(z, w_2) + \alpha_2) \\ &= (\Phi(z, w_p), \varphi_0(z, w_p) - \xi + \alpha_{0p}, \varphi(z, w_p) + \alpha_p) \\ &= F(x, u_p). \end{aligned}$$

Das zeigt, daß $F(x, U)$ für jedes $x \in \tilde{V}$ eine konvexe Menge ist.

d) Es ist

$$\langle F_x(\hat{x}, u), x \rangle = (\langle \Phi_z(\hat{z}, w), z \rangle, \langle \varphi_{0_z}(\hat{z}, w), z \rangle - \xi, \langle \varphi_z(\hat{z}, w), z \rangle)\,.$$

Nach Voraussetzung liefert das Paar bestehend aus der ersten und der letzten Komponente dieses Vektors einen abgeschlossenen Teilraum endlicher Kodimension in $Y \times E$, wenn z den Raum Z durchläuft. Da in der mittleren Koordinate ξ unabhängig von z die reellen Zahlen durchläuft, folgt, daß das Bild des Operators $F_x(\hat{x}, u)$ in $\tilde{Y}$ ein abgeschlossener Teilraum endlicher Kodimension ist.

Damit sind alle Voraussetzungen des Satzes A aus [31] erfüllt.

4. Konsequenzen aus Satz A

Die Lagrangefunktion der Aufgabe (A) ist

$$\tilde{\mathcal{L}} = \mu_0 f_0(x) + \langle \tilde{y}^*, F(x, u) \rangle$$

mit $\mu_0 \in \mathbb{R}$ und $\tilde{y}^* \in \tilde{Y}^*$. Unter Benutzung von

$$\tilde{Y}^* = (Y \times \mathbb{R} \times E)^* = Y^* \times \mathbb{R} \times E^*$$

schreibt sich $\tilde{\mathcal{L}}$ in der Form

$$\tilde{\mathcal{L}} = \mu_0 f_0(x) + \langle y^*, \Phi(z, w) \rangle + \lambda_0 (\varphi_0(z, w) - \xi + \alpha_0) + \langle \lambda, \varphi(z, w) + \alpha \rangle$$

mit μ_0, $\lambda_0 \in \mathbb{R}$, $y^* \in Y^*$, $\lambda \in E^*$. Nach Satz A existieren $\hat{\mu}_0$, $\hat{\lambda}_0$, $\hat{y}^*$ und λ, mit $\hat{\mu}_0 \geq 0$, die nicht gleichzeitig verschwinden, so daß für die mit ihnen gebildetete Lagrangefunktion gilt:

$$\tilde{\mathcal{L}}_x = 0 \tag{27}$$

und

$$\min_{u \in U} \tilde{\mathcal{L}}(\hat{x}, u) = \tilde{\mathcal{L}}(\hat{x}, \hat{u}). \tag{28}$$

Ausgeschrieben lautet (27)

$$\hat{\mu}_0 \xi + \langle \hat{y}^*, \Phi_z(z, w) \rangle + \hat{\lambda}_0 (\varphi_{0_z}(z, w) - \xi) + \langle \hat{\lambda}, \varphi_z(z, w) \rangle = 0$$

für jedes $\xi \in \mathbb{R}$ und jedes $z \in Z$. Dies impliziert sofort $\hat{\mu}_0 = \hat{\lambda}_0$. Daraus folgt die Behauptung 0) unseres Satzes sowie $\hat{\lambda}_0 \geq 0$ und die Stationaritätsbedingung 1).

In (28) setzen wir $w = \hat{w}$ und erhalten unter Beachtung von $\Phi(\hat{z}, \hat{w}) = 0$, $\hat{\mu}_0 = \hat{\lambda}_0$,

$$\min_{\alpha_0 \geq 0, \alpha \in C} \hat{\lambda}_0 (\varphi_0(\hat{z}, \hat{w}) + \alpha_0) + \langle \hat{\lambda}, \varphi(\hat{z}, \hat{w}) + \alpha \rangle = \hat{\lambda}_0 \varphi_0(\hat{z}, \hat{w}),$$

also, da $\min_{\alpha_0 \geq 0} \hat{\lambda}_0 \alpha_0 = 0$ ist,

$$\min_{\alpha \in C} \langle \hat{\lambda}, \alpha \rangle = -\langle \hat{\lambda}, \varphi(\hat{z}, \hat{w}) \rangle. \tag{29}$$

Daraus folgt die Nichtnegativitätsaussage 3) für $\hat{\lambda}$, denn wäre $\langle \hat{\lambda}, c \rangle < 0$ für ein $c \in C$, so könnte man in (29) $\alpha = tc$ mit $t > 0$ setzen (hier geht ein, daß C ein Kegel ist) und würde für das Minimum den absurden Wert $-\infty$ herausbekommen.

Folglich ist die linke Seite von (29) ≥ 0, also offenbar $= 0$. Das liefert die komplementäre Schlupfbedingung 4).

Das noch zu zeigende Minimumprinzip 2) erhält man, indem man in (28) $\alpha_0 = 0$ und $\alpha = 0$ setzt, was wegen der bereits bewiesenen Nichtnegativitätsbedingungen erlaubt ist. ■

2.3 Spezialisierung

Man macht sich leicht klar, daß unser Satz A" den Satz A' aus [31] impliziert. Wir führen daß hier aber nicht aus, sondern formulieren nur die Form des Satzes A", die wir später anwenden werden.

Satz 3 *Es sei Z ein Banachraum, m, n natürliche Zahlen und*

$$\begin{aligned} \varphi_0 &: Z \to \mathbb{R} & & \\ \varphi_i &: Z \to \mathbb{R}, & i &= 1, \ldots, m \\ \psi_j &: Z \to \mathbb{R}, & j &= 1, \ldots, n \end{aligned}$$

stetig Fréchet-differenzierbare Funktionen. Weiter sei C ein konvexer Kegel in Z.

Ist dann $\hat{z}$ eine Lösung der Aufgabe

$$\varphi_0(z) \to \inf;$$

$$\varphi_i(z) \le 0, \quad i = 1, \ldots, m, \quad \psi_j(z) = 0, \quad j = 1, \ldots, n, \quad z \in C,$$

so existieren $\lambda_0, \lambda_1, \ldots, \lambda_m, \alpha_1, \ldots, \alpha_n \in \mathbb{R}$ mit den Eigenschaften

i) Die λ_i, α_j sind nicht alle $= 0$.

ii) $\lambda_0 \ge 0, \ldots, \lambda_m \ge 0$.

iii) Für die „verkürzte" Lagrangefunktion

$$\mathcal{L} = \lambda_0 \varphi_0(z) + \sum_{i=1}^{m} \lambda_i \varphi_i(z) + \sum_{j=1}^{n} \alpha_j \psi_j(z)$$

gilt

$$\langle \mathcal{L}_z(\hat{z}), z \rangle \ge 0 \quad \forall z \in C$$

und

iv) $\langle \mathcal{L}_z(\hat{z}), \hat{z} \rangle = 0$

Beweis Wir wenden den Satz (A") auf die mit

$$\begin{aligned} \tilde{Z} &:= Z \\ \tilde{V} &:= Z \\ \tilde{Y} &:= \mathbb{R}^n \\ \tilde{E} &:= Z \times \mathbb{R}^m \\ \tilde{C} &:= (-C) \times [0, \infty)^m \\ \tilde{W} &:= \text{beliebige einelementige Menge} \\ \tilde{\varphi}_0(z, w) &:= \varphi_0(z) \\ \tilde{\varphi}(z, w) &:= (z, \varphi_1(z), \ldots, \varphi_m(z)) \\ \tilde{\Phi}(z, w) &:= (\psi_1(z), \ldots, \psi_n(z)) \end{aligned}$$

gebildete Aufgabe (A") an, die mit der Aufgabe des Satzes 3 identisch ist.

Die Voraussetzungen a) bis c) des Satzes 2 sind trivialerweise erfüllt. d) gilt, weil wegen

$$\left(\tilde{\varphi}_z(\hat{z})z, \tilde{\Phi}_z(\hat{z})z\right) = (z, \langle\varphi_{1_z}(\hat{z}), z\rangle, \ldots, \langle\varphi_{m_z}(\hat{z}), z\rangle, \langle\psi_{1_z}(\hat{z}), z\rangle, \ldots, \langle\psi_{n_z}(\hat{z}), z\rangle)$$

das Bild des linearen Operators $(\tilde{\varphi}_z(\hat{z}), \tilde{\Phi}_z(\hat{z}))$ gleich dem Graphen der stetigen Abbildung

$$z \to (\langle\varphi_{1_z}(\hat{z}), z\rangle, \ldots \langle\varphi_{m_z}(\hat{z}), z\rangle, \langle\psi_{1_z}(\hat{z}), z\rangle, \ldots, \langle\psi_{n_z}(\hat{z}), z\rangle)$$

und damit abgeschlossen ist und die Bedingung der endlichen Kodimension mit Hilfe des Lemmas 0.5.5 aus [31] eingesehen werden kann.

Es gibt also

$$\begin{aligned}
\tilde{\lambda}_0 &=: \lambda_0 \in \mathbb{R}, \\
\tilde{\lambda} &=: (\lambda_\infty, \lambda_1, \ldots, \lambda_m) \in \tilde{E}^* = Z^* \times \mathbb{R}^m, \\
\tilde{y}^* &=: (\alpha_1, \ldots, \alpha_n) \in \tilde{Y}^* = \mathbb{R}^n
\end{aligned}$$

mit

$$\lambda_0\langle\varphi_{0_z}(\hat{z}), z\rangle + \langle\lambda_\infty, z\rangle + \sum_{i=1}^{m}\lambda_i\langle\varphi_{i_z}(\hat{z}), z\rangle + \sum_{j=1}^{n}\alpha_j\langle\psi_{j_z}(\hat{z}), z\rangle = 0$$

für $z \in Z$.

Es ist somit

$$\mathcal{L} = \lambda_0\varphi_{0_z}(\hat{z}) + \sum_{i=1}^{m}\lambda_i\varphi_{i_z}(\hat{z}) + \sum_{j=1}^{n}\alpha_j\psi_{j_z}(\hat{z}) = -\lambda_\infty. \tag{30}$$

Wären nun $\lambda_0, \ldots, \lambda_m, \alpha_1, \ldots, \alpha_n$ alle $= 0$, so wäre auch $\lambda_\infty = 0$, was nach Satz A" nicht der Fall sein kann. Folglich gilt i).

Weiter ist nach Satz A" $\langle\tilde{\lambda}, \tilde{c}\rangle \geq 0$ für $\tilde{c} \in \tilde{C}$. Das impliziert zum einen $\lambda_1 \geq 0, \ldots, \lambda_m \geq 0$, also, zusammen mit $\tilde{\lambda}_0 \geq 0$, die Behauptung ii).

Zum anderen folgt $\langle\lambda_\infty, z\rangle \geq 0$ für $z \in -C$, also

$$\langle-\lambda_\infty, z\rangle \geq 0 \qquad (z \in C), \tag{31}$$

was wegen (30) die Behauptung iii) ist.

iv) schließlich folgt aus

$$\begin{aligned}
0 &= \langle\tilde{\lambda}, (\hat{z}, \varphi_1(\hat{z}), \ldots, \varphi_m(\hat{z}))\rangle && \text{(Aussage 3) Satz A")} \\
&= \langle\lambda_\infty, \hat{z}\rangle + \textstyle\sum_{i=1}^{m}\lambda_i\varphi_i(\hat{z}) \\
&\leq \langle\lambda_\infty, \hat{z}\rangle && (\text{wegen } \varphi_i(\hat{z}) \leq 0,\ \lambda_i \geq 0) \\
&\leq 0 && (\text{wegen } \hat{z} \in C \text{ und } (31))
\end{aligned}$$

und (30). ∎

2.4 Konkretisierung: Definition der Aufgabe (E)

Zur Illustration des Satzes aus 2.3 wollen wir den Satz von Plackett erneut beweisen. Um spätere Wiederholungen zu vermeiden, betrachten wir jedoch gleich eine etwas allgemeinere Extremalaufgabe.

Definition (Aufgabe (E)) *Es seien*

$$K : \mathbb{R}^d \times \mathbb{R}^d \to [-\infty, \infty]$$

und

$$g_i : \mathbb{R}^d \to [-\infty, \infty], \quad i = 1, \ldots, n$$

Borel-meßbare Funktionen und a_i, $i = 1, \ldots, n$, reelle Zahlen. K sei symmetrisch in seinen Argumenten:

$$K(x, y) = K(y, x)$$

Dann bezeichnen wir Aufgaben der Form

„Minimiere

$$E[K(X, Y)],$$

wobei X und Y unabhängige und identisch verteilte Zufallsvektoren mit $E[|K(X,Y)|] < \infty$, $E[|g_i(X)|] < \infty$, $i = 1, \ldots, n$, seien, die den Nebenbedingungen

$$E[g_i(X)] = a_i, \quad i = 1, \ldots, n \tag{32}$$

genügen sollen." mit (E). Dabei dürfen alle oder einzelne der Nebenbedingungen auch vom Ungleichungstyp sein:

$$E[g_i(X)] \le a_i \quad . \tag{33}$$

Es liegt nahe, sich zunächst mittels geeigneter Kompaktheitsschlüsse (Auswahlsatz von Helly) der Existenz einer Lösung der Aufgabe (E) zu vergewissern, zumal der oben angegebene Satz 3 nur eine notwendige Bedingung für das Vorliegen eines Minimums liefert. Wir beweisen daher zur Orientierung das folgende Lemma (es wird in dieser Arbeit nur in Kapitel 6 beim Beweis des Satzes 11 wirklich benutzt). Am Ende des Abschnitts 4.2 wird eine harmlos aussehende Extremalaufgabe angegeben, welche keine Lösung besitzt.

Lemma 1 *In der Aufgabe (E) sei $n = 1$ und die Nebenbedingungen* (32), (33) *seien durch*

$$E[|X|^2] \le a \tag{34}$$

gegeben. Weiter sei

$$K : \mathbb{R}^d \times \mathbb{R}^d \to (-\infty, \infty]$$

stetig. Es gebe ein $\epsilon > 0$ und ein $c \in \mathbb{R}$ so daß

$$K_-(x, y) \le c(1 + |x|^{2-\epsilon})(1 + |y|^{2-\epsilon}) \qquad (x, y \in \mathbb{R}^d)$$

gilt. Die Menge der zulässigen Verteilungen von (X,Y) sei nicht leer. Dann hat (E) eine Lösung.

Ebenso hat (E) eine Lösung, wenn die Nebenbedingung

$$E[X] = b \tag{35}$$

mit einem $b \in \mathbb{R}^d$ hinzugefügt wird.

Beweis Es sei

$$D := \{\mu : \mu \text{ Wahrscheinlichkeitsmaß im } \mathbb{R}^d, \quad \int_{\mathbb{R}^d} |x|^2 \, d\mu(x) \le a\}$$

und $\varphi : D \to (-\infty, \infty]$, $\varphi(\mu) := \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} K(x,y)\, d\mu(x) d\mu(y)$. Setze $\gamma = \inf \varphi(D)$ und wähle eine Folge (μ_n) mit $\varphi(\mu_n) \to \gamma$. Hat X_n die Verteilung μ_n, so gilt $P(|X_n| \ge \alpha) \le \frac{1}{\alpha^2} E[|X_n|^2] \le \frac{a}{\alpha^2}$ für $\alpha > 0$. (μ_n) ist folglich straff. Wir können daher annehmen, daß $\mu_n \xrightarrow{\mathcal{D}} \mu$ mit einem Wahrscheinlichkeitsmaß μ gilt. Sind $X, X_1, X_2, \ldots$ Zufallsvariablen mit den Verteilungen $\mu, \mu_1, \mu_2, \ldots$, so folgt mit dem Lemma von Fatou [6, Satz 25.11, Seite 347] $a \ge \liminf_{n\to\infty} E[|X_n|^2] \ge E[|X|^2]$. Folglich ist $\mu \in D$.

Nun seien $Y, Y_1, Y_2, \ldots$ verteilt wie $X, X_1, X_2, \ldots$, aber unabhängig von diesen. Wegen der Stetigkeit von K gilt dann $K(X_n, Y_n) \xrightarrow{\mathcal{D}} K(X,Y)$. Die Folge $(|X_n|^{2-\epsilon})$ ist gleichgradig integrierbar (siehe [6, Seite 348]). Mit dem unten angegebenen Lemma 2 ergibt sich die gleichgradige Integrierbarkeit der Folge $((1 + |X_n|^{2-\epsilon})(1 + |Y_n|^{2-\epsilon}))$ und somit auch der Folge $(K_-(X_n, Y_n))$. Man erhält mit [6, Satz 25.12, Seite 348] $\lim_{n\to\infty} E[K_-(X_n, Y_n)] = E[K_-(X,Y)]$. Zusammen mit dem Lemma von Fatou folgt daraus $\gamma = \lim E[K(X_n, Y_n)] = -E[K_-(X,Y)] + \lim E[K_+(X_n, Y_n)] \ge -E[(K_-(X,Y)] + E[K_+(X,Y)] = E[K(X,Y)]$, wegen $\mu \in D$ also $\varphi(\mu) = \gamma$.

Damit ist das Lemma für den Fall bewiesen, daß die Nebenbedingung (35) nicht berücksichtigt werden soll. Den noch zu beweisenden Fall behandelt man aber wie oben. Man hat lediglich die Nebenbedingung (35) in die Definition von D mit aufzunehmen und sich mit Hilfe der gleichgradigen Integrierbarkeit der Folge (X_n) (koordinatenweise zu verstehen) und des Satzes 25.12 aus [6] davon zu überzeugen daß das konstruierte μ sie ebenfalls erfüllt. ■

Lemma 2 *Sind die Folgen von reellen Zufallsvariablen (X_n), (Y_n) gleichgradig integrierbar und sind X_n und Y_n für jedes n unabhängig, so ist auch die Folge (X_nY_n) gleichgradig integrierbar.*

Beweis Die Behauptung ergibt sich aus

$$\int_{\{|X_nY_n| \ge \alpha\}} |X_nY_n| \, dP \le E[|X_n|] \int_{\{|Y_n| \ge \sqrt{\alpha}\}} |Y_n| \, dP + E[|Y_n|] \int_{\{|X_n| \ge \sqrt{\alpha}\}} |X_n| \, dP.$$

■

Setzt man in Lemma 1 $d = 1$, $K(x,y) = -|x-y|$ und $a = 0$, so erhält man wegen $(-|x-y|)_- = |x-y| \leq (1+|x|)(1+|y|)$ die Existenz einer Lösung der zu Beginn von 2.1 geschilderten Aufgabe. Man muß sich nur klarmachen, daß in diesem Fall die Aufgabe (E) nicht durch eine Verteilung μ mit Varianz < 1 gelöst werden kann, was aber aus Homogenitätsgründen offensichtlich ist.

Nun folgern wir, unter gewissen Zusatzbedingungen, aus Satz 3 eine Multiplikatorenregel für die Aufgabe (E).

Satz 4 *In der Aufgabe (E) gebe es ein $c \in \mathbb{R}$ mit*

$$|K(x,y)| \leq cw(x)w(y) \tag{36}$$

mit

$$w(x) := 1 + \sum_{i=1}^{n} |g_i(x)|.$$

Ist dann die Wahrscheinlichkeitsverteilung μ eine Lösung von (E), so gibt es ein $\lambda = (\lambda_0, \ldots, \lambda_{n+1}) \in \mathbb{R}^{n+2}$ mit den Eigenschaften

$$\lambda \neq 0,$$

$$\lambda_0 \geq 0,$$

so daß für die Funktion

$$l(x) := \lambda_0 \int_{\mathbb{R}^d} K(x,y)\, d\mu(y) + \lambda_{n+1} + \sum_{i=1}^{n} \lambda_i g_i(x)$$

gilt:

$$l(x) \begin{cases} \geq 0, & x \in \mathbb{R}^d \\ = 0, & \mu\text{-fast überall.} \end{cases} \tag{37}$$

Liegen einzelne Nebenbedingungen in Ungleichungsform vor, so sind die zu ihnen gehörenden λ_i nichtnegativ.

Sind zusätzlich zu den oben gemachten Voraussetzungen die Funktionen $K, g_1, \ldots, g_n$ stetig, so ist auch l stetig und verschwindet im Träger von μ.

Beweis Für signierte Maße σ im $\mathbb{R}^d$ setzen wir

$$\|\sigma\| := \int_{\mathbb{R}^d} w(x)\, |d\sigma(x)|, \tag{38}$$

wobei $|d\sigma(x)|$ für die Integration bezüglich der totalen Variation von σ steht. Dann ist der Vektorraum

$$Z := \{\sigma : \sigma \text{ signiertes Maß}, \|\sigma\| < \infty\}$$

ein Banachraum, wenn man ihn mit der Norm (38) versieht (die Konvergenz jeder Cauchy-Folge (σ_n) erhält man aus dem wohlbekannten Spezialfall $w(x) \equiv 1$ durch Betrachtung von $(w\sigma_n)$).

Wir betrachten zunächst den Fall, daß alle Nebenbedingungen in Gleichungsform (32) vorliegen und definieren Funktionen $\varphi_0, \ldots, \varphi_n : Z \to \mathbb{R}$ durch

$$\begin{aligned}
\varphi_0(\sigma) &:= \textstyle\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(x,y)\, d\sigma(x) d\sigma(y) \\
\varphi_1(\sigma) &:= \textstyle\int_{\mathbb{R}^d} g_1(x)\, d\sigma(x) - a_1 \\
&\;\;\vdots \\
\varphi_n(\sigma) &:= \textstyle\int_{\mathbb{R}^d} g_n(x)\, d\sigma(x) - a_n \\
\varphi_{n+1}(\sigma) &:= \sigma(\mathbb{R}^d)
\end{aligned}$$

für $\sigma \in Z$. Diese Funktionen sind stetig Fréchet–differenzierbar, denn $\varphi_1, \ldots, \varphi_{n+1}$ sind nach Wahl der Norm auf Z stetige Linearformen, während φ_0 wegen der für alle $\sigma, \tau \in Z$ gültigen Abschätzung

$$\begin{aligned}
\left| \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(x,y)\, d\sigma(x) d\tau(y) \right| &\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |K(x,y)|\, |d\sigma(x)|\, |d\tau(y)| \\
&\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} c w(x) w(y)\, |d\sigma(x)|\, |d\tau(y)| \\
&= c\|\sigma\|\|\tau\|
\end{aligned}$$

die zu einer stetigen und symmetrischen Bilinearform auf $Z \times Z$ gehörige quadratische Form ist.

Weiter ist

$$C := \{\sigma \in Z : \sigma \text{ nichtnegatives Maß } \}$$

ein konvexer Kegel in Z. Damit sind alle Voraussetzungen des Satzes 3 erfüllt und es folgt, wenn μ eine Lösung der Aufgabe (E) ist, die Existenz von nicht gleichzeitig verschwindenden Zahlen $\lambda_0, \ldots, \lambda_{n+1} \in \mathbb{R}$ mit $\lambda_0 \geq 0$, so daß mit

$$\mathcal{L} = \frac{\lambda_0}{2} \varphi_0(\sigma) + \sum_{i=1}^{n+1} \lambda_i \varphi_i(\sigma)$$

die Aussage (iii) des Satzes 3 gilt. Für alle $\nu \in C$ gilt somit

$$\begin{aligned}
0 &\leq \langle \mathcal{L}_\sigma(\mu), \nu \rangle \\
&= \lambda_0 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(x,y)\, d\mu(y) d\nu(x) + \lambda_{n+1} \nu(\mathbb{R}^d) + \sum_{i=1}^{n} \lambda_i \int_{\mathbb{R}^d} g_i\, d\nu \\
&= \int_{\mathbb{R}^d} l(x)\, d\nu(x).
\end{aligned}$$

Setzt man für ν speziell jedes Dirac–Maß ein, so ergibt sich die Ungleichung $l(x) \geq 0$ für alle $x \in \mathbb{R}^d$. Setzt man weiter $\mu = \nu$, so folgt daraus mit der Aussage (iv) des Satzes 3 die Behauptung (37).

Sind nun zusätzlich die Funktionen $K, g_1, \ldots, g_n$ stetig, so ist auch l stetig (dazu weist man die Stetigkeit von $x \longrightarrow \int_{\mathbb{R}^d} K(x,y)\, d\mu(y)$ mit Hilfe des Satzes über die majorisierte Konvergenz unter Verwendung einer Majorante der Form $\tilde{c}(x_0)w(y)$ nach). Damit kann (37) nur gelten, wenn l im Träger von μ verschwindet.

Im Fall, daß einzelne Nebenbedingungen in Ungleichungsform vorliegen, verfährt man wie oben und erhält aus Satz 3, daß die zu ihnen gehörigen Multiplikatoren nichtnegativ sind.

■

2.5 Ein Beispiel: Neuer Beweis des Satzes von Plackett

Wir betrachten die Aufgabe (E) mit

$$K(x,y) = -|x-y|$$

und den Nebenbedingungen

$$E[X] = 0$$
$$E[X^2] = 1,$$

wobei wir die Bedingung an den Erwartungswert von X aber auch weglassen könnten.

Die Existenz einer Lösung wurde bereits im Anschluss an Lemma 1 festgestellt.

Sei nun μ eine Lösung von (E). Die Voraussetzungen des Satzes 4 sind erfüllt. Es gibt also nicht gleichzeitig verschwindende Multiplikatoren $\lambda_0, \ldots, \lambda_3 \in \mathbb{R}$ mit $\lambda_0 \geq 0$, so daß gilt:

$$\begin{aligned} l(x) &= -\lambda_0 \int_{\mathbb{R}} |x-y|\, d\mu(y) + \lambda_3 + \lambda_1 x + \lambda_2 x^2 \\ &=: -\lambda_0 \int_{\mathbb{R}} |x-y|\, d\mu(y) + p(x) \begin{cases} \geq 0 & (x \in \mathbb{R}) \\ = 0 & (x \in \operatorname{supp} \mu). \end{cases} \end{aligned}$$

Wäre $\lambda_0 = 0$, so wäre $l = p$, ein Polynom zweiten Grades. Dieses kann nicht identisch verschwinden, da dann alle Multiplikatoren $= 0$ wären. Es kann nicht genau zwei Nullstellen haben, da dann l auch negative Werte annehmen müßte. Es kann weiter nicht genau eine Nullstelle besitzen, da dann der Träger von μ einelementig und damit $E[K(X,Y)] = 0$, also sicherlich nicht minimal wäre. Schließlich kann p nicht nullstellenfrei sein, da dann der Träger von μ leer wäre. Folglich muß $\lambda_0 > 0$ sein und wir können annehmen, daß $\lambda_0 = 1$ ist.

Bezeichnet nun F die rechtsseitig stetige Verteilungsfunktion von μ, so folgt mit partieller Integration oder Rückführung auf die bekannte Flächenformel für Erwartungswerte[5]

$$l(x) = -\int_x^\infty 1 - F(y)\, dy - \int_{-\infty}^x F(y)\, dy + p(x).$$

[5] Lemma 4 auf Seite 33 enthält eine geringfügig allgemeinere Fassung dieses Sachverhaltes.

Man erkennt, daß l links- und rechtsseitig differenzierbar ist mit

$$l'(x-) = 1 - 2F(x-) + p'(x),$$

$$l'(x+) = 1 - 2F(x) + p'(x).$$

Ist $x \in \operatorname{supp} \mu$, so ist x Minimalstelle von l und es folgt $l'(x-) \leq 0$, $l'(x+) \geq 0$, also

$$0 \leq 2\,(F(x) - F(x-)) = l'(x-) - l'(x+) \leq 0.$$

Folglich ist F stetig in ganz $\mathbb{R}$ und $l'(x) = 0$ für $x \in \operatorname{supp} \mu$ und damit $F(x) = \frac{1}{2}(1 + p'(x)) = a + bx$ mit $a, b \in \mathbb{R}$. F ist demnach Verteilungsfunktion einer Gleichverteilung auf einem Intervall, welches durch die Nebenbedingungen wie im Satz von Plackett behauptet festgelegt wird.

3 Konvexität

3.1 Anwendung des Satzes von Kuhn-Tucker

Für eine konvexe differenzierbare Funktion $\varphi_0 : \mathbb{R}^d \to \mathbb{R}$ ist das Verschwinden der Ableitung nicht nur notwendig, sondern auch hinreichend für das Vorliegen eines Minimums, was gegebenenfalls die Beweisführung bei einem Extremalproblem sehr erleichtern kann. Ähnliches gilt nun für die sogenannten konvexen Optimierungsprobleme mit Nebenbedingungen (Satz von Kuhn-Tucker, [31, Seite 77]).

Betrachten wir nun nochmals das beim Plackettschen Satz zu minimierende Funktional

$$\varphi_0(\mu) = -\int_{\mathbb{R}}\int_{\mathbb{R}} |x-y|\, d\mu(x)\, d\mu(y).$$

Ist μ ein Wahrscheinlichkeitsmaß und F seine Verteilungsfunktion, so ist

$$\varphi_0(\mu) = \varphi_0(F) = -2\int_{-\infty}^{\infty} F(x)(1-F(x))\, dx \tag{39}$$

(dies entspricht dem Zähler der Darstellung (3)). Man schließt nun aus der Konvexität der Funktion $t \to -t(1-t)$ und der Linearität des Übergangs $\mu \to F$, daß das Funktional φ_0 jedenfalls auf der Menge der Wahrscheinlichkeitsmaße konvex ist (für andere (signierte) Maße gilt die obige Darstellung in der Regel nicht). Stellt man die Nebenbedingung an das zweite Moment in der Form

$$\varphi_1(\mu) = \int_{\mathbb{R}} x^2\, d\mu(x) \le 1,$$

so liegt, weil φ_1 linear ist, ein konvexes Optimierungsproblem vor, und der Satz von Kuhn-Tucker ist anwendbar.

Da diese Situation für die in dieser Arbeit betrachteten Probleme typisch ist, formulieren wir den

Satz 5 *In der Aufgabe (E) mögen nur Nebenbedingungen vom Ungleichungstyp (33) vorkommen. Das Funktional*

$$\mu \to \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} K(x,y)\, d\mu(x)\, d\mu(y) \tag{40}$$

sei auf der Menge der Wahrscheinlichkeitsmaße mit $\int_{\mathbb{R}^d} |g_i(x)|\, d\mu(x) < \infty$ konvex. Weiter seien alle Voraussetzungen des Satzes 4 erfüllt, einschließlich der Stetigkeit der vorkommenden Funktionen. Es gebe ein Wahrscheinlichkeitsmaß, für welches die Nebenbedingungen mit „$<$" an Stelle von „$\le$" erfüllt seien („Slater-Bedingung").

Dann ist ein Wahrscheinlichkeitsmaß μ genau dann Lösung von (E), wenn es reelle Zahlen $\alpha, \lambda_1, \ldots, \lambda_n$ gibt, für die

$$\lambda_i \ge 0, \quad i = 1, \ldots, n, \tag{41}$$

$$l(x) := \int_{\mathbb{R}^d} K(x,y)\, d\mu(y) + \alpha + \sum_{i=1}^{n} \lambda_i g_i(x) \begin{cases} \geq 0 & (x \in \mathbb{R}^d) \\ = 0 & (x \in \operatorname{supp} \mu) \end{cases} \tag{42}$$

und

$$\lambda_i \left(\int_{\mathbb{R}^d} g_i(x)\, d\mu(x) - a_i \right) = 0, \quad i = 1, \ldots, n, \tag{43}$$

gilt.

Ist (40) *sogar streng konvex, so hat (E) höchstens eine Lösung.*

Beweis Es sei W die Menge der Wahrscheinlichkeitsmaße, die der angegebenen Integrierbarkeitsbedingung genügen. Dann ist W eine konvexe Teilmenge des Vektorraumes der signierten Borelmaße und wir können den Satz von Kuhn–Tucker zusammen mit der Slater-Bedingung anwenden ([31], Seite 77). Das liefert die Aussage, daß $\hat{\mu}$ genau dann Lösung von (E) ist, wenn es reelle Zahlen $\lambda_1, \ldots, \lambda_n \geq 0$ gibt, die (41) und (43) erfüllen, und wenn für die mit ihnen gebildete Lagrange-Funktion

$$\mathcal{L}(\mu) := \frac{1}{2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(x,y)\, d\mu(x)\, d\mu(y) + \sum_{i=1}^{n} \lambda_i \int_{\mathbb{R}^d} g_i(x)\, d\mu(x)$$

gilt:

$$\min_{\mu \in W} \mathcal{L}(\mu) = \mathcal{L}(\hat{\mu}).$$

Nun ist $\mathcal{L}$ auf W ein konvexes Funktional sowie differenzierbar wie im Beweis des Satzes 4. $\mathcal{L}$ nimmt daher ein Minimum im Punkt $\hat{\mu}$ genau dann an, wenn die (einseitige) Richtungsableitung in Richtung $\nu - \mu$ für jedes $\nu \in W$ im Punkt $\hat{\mu}$ nichtnegativ ist:

$$\int_{\mathbb{R}^d} \tilde{l}(x)\, d(\nu - \hat{\mu})(x) \geq 0, \quad \nu \in W \tag{44}$$

mit

$$\tilde{l}(x) := \int_{\mathbb{R}^d} K(x,y)\, d\hat{\mu}(y) + \sum_{i=1}^{n} \lambda_i g_i(x).$$

Nun sind aber die Bedingungen (44) und (42) äquivalent:

Gilt (44), so ist

$$-\alpha := \int_{\mathbb{R}^d} \tilde{l}(x)\, d\hat{\mu}(x) \leq \int_{\mathbb{R}^d} \tilde{l}(x)\, d\nu(x)$$

insbesondere wenn ν ein beliebiges Dirac-Maß ist, also

$$l(x) = \tilde{l}(x) + \alpha \geq 0, \quad x \in \mathbb{R}^d. \tag{45}$$

l ist stetig (siehe den Beweis von Satz 4), und nach Definition von α gilt

$$\int_{\mathbb{R}^d} l(x)\, d\hat{\mu}(x) = 0.$$

Zusammen mit (45) folgt (42).

Gilt umgekehrt (42), so ist wegen $\tilde{l} = l - \alpha$

$$\int_{\mathbb{R}^d} \tilde{l}\, d(\nu - \hat{\mu}) = \int_{\mathbb{R}^d} (l - \alpha)\, d(\nu - \hat{\mu}) = \int_{\mathbb{R}^d} l\, d(\nu - \hat{\mu}) = \int_{\mathbb{R}^d} l\, d\nu \geq 0 \qquad (46)$$

für $\nu \in W$, d.h. (44) ist erfüllt. ∎

Mit Satz 5 kann nun der Satz von Plackett erneut bewiesen werden. Ist nämlich μ die Gleichverteilung im Intervall $[-\sqrt{3}, \sqrt{3}]$, so haben wir

$$\int_{\mathbb{R}} |x - y|\, d\mu(y) = \frac{1}{2\sqrt{3}} \int_{-\sqrt{3}}^{\sqrt{3}} |x - y|\, dy = \begin{cases} \frac{1}{2\sqrt{3}}(x^2 + 3) & (|x| \leq \sqrt{3}) \\ |x| & (|x| \geq \sqrt{3}) \end{cases}$$

und damit gelten die Bedingungen (41),(42) und (43) des obigen Satzes mit $n = 1$, $g_1(x) = x^2$, $a_1 = 1$, $K(x,y) = -|x - y|$, $\alpha = \frac{\sqrt{3}}{2}$, $\lambda_1 = \frac{1}{2\sqrt{3}}$. μ ist also eine Lösung des Plackettschen Extremalproblems, und zwar die einzige, wegen der aus (39) ersichtlichen strengen Konvexität des Funktionals φ_0.

Der eben gegebene Beweis ist natürlich nur deswegen so kurz, weil wir für μ die richtige Verteilung eingesetzt haben. Hätten wir diese nicht schon vorher gekannt, so hätten wir zu ihrer Auffindung ähnliche Überlegungen wie in 2.5 anstellen müssen. Dennoch hat der Satz 5, falls er anwendbar ist, gegenüber dem Satz 4 zwei Vorteile: Erstens braucht kein separater Existenzbeweis geführt zu werden, und zweitens braucht man sich nicht mit nur theoretisch denkbaren irregulären Lösungen zu befassen (wie etwa in 2.5, Nachweis der Stetigkeit von F und damit der Differenzierbarkeit von l).

3.2 Eine Klasse strikt negativ definiter Integralkerne

Die Überlegungen des vorhergehenden Abschnittes zeigen, daß es nützlich ist, über möglichst einfache Kriterien für die Konvexität von Funktionalen von Wahrscheinlichkeitsmaßen zu verfügen. Bei den in dieser Arbeit auftretenden Funktionalen handelt es sich speziell um quadratische Formen der Gestalt

$$Q(\mu) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(x,y)\, d\mu(x)\, d\mu(y).$$

Wegen der für reelle Zahlen α und beliebige Maße μ, ν unter geeigneten Integrierbarkeitsvoraussetzungen gültigen Identität

$$\alpha Q(\mu) + (1 - \alpha) Q(\nu) - Q(\alpha\mu + (1 - \alpha)\nu) = \alpha(1 - \alpha) Q(\mu - \nu)$$

ist die Konvexität von Q auf der Menge der Wahrscheinlichkeitsmaße äquivalent zu der Bedingung

$$Q(\mu - \nu) \geq 0 \qquad (\mu, \nu \text{ Wahrscheinlichkeitsmaße}).$$

Setzt man $\tilde{K} := -K$ und benutzt man, daß sich jedes signierte Maß als Differenz zweier endlicher Maße darstellen läßt, so erhält man als weitere äquivalente Formulierung

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \tilde{K}(x,y)\, d\sigma(x)\, d\sigma(y) \leq 0 \qquad (\sigma \text{ signiertes Maß mit } \sigma(\mathbb{R}^d) = 0). \qquad (47)$$

Schreibt man sich die letzte Beziehung speziell für molekulare signierte Maße auf[6], so ergibt sich

$$\sum_{j,k=1}^{n} c_j c_k \tilde{K}(x_j, x_k) \leq 0 \qquad (n \in \mathbb{N}, x_1, \ldots, x_n \in \mathbb{R}^d, c_1, \ldots, c_n \in \mathbb{R}, \sum_{j=1}^{n} c_j = 0) \tag{48}$$

Es ist üblich, reelle symmetrische Kerne $\tilde{K} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ als negativ definite Kerne zu bezeichnen, wenn $\mathcal{X}$ eine nichtleere Menge ist und die Bedingung (48) mit $\mathcal{X}$ an Stelle von $\mathbb{R}^d$ erfüllt ist (siehe [5, Kapitel 3], wo allgemeiner komplexwertige Kerne betrachtet werden). Ist $\tilde{K}$ negativ definit und von der speziellen Form

$$\tilde{K}(x, y) = \psi(x - y) \qquad (x, y \in \mathbb{R}^d)$$

mit einer Funktion $\psi : \mathbb{R}^d \to \mathbb{R}$ mit $\psi(x) = \psi(-x)$ für $x \in \mathbb{R}^d$, so bezeichnet man ψ als negativ definite Funktion[7]. Ein sehr einfach nachzurechnendes Beispiel ist durch

$$\psi(x) = |x|^2 \qquad (x \in \mathbb{R}^d) \tag{49}$$

gegeben.

Es existiert nun eine Theorie negativ definiter Funktionen und Kerne, die insbesondere eine große Zahl von konkreten Beispielen hevorbringt. Diese Theorie, wie sie etwa in [5] dargelegt ist, vernachlässigt aber Fragen, welche die Gültigkeit von Ungleichungen wie (47) für nicht notwendig molekulare signierte Maße oder das Eintreten des Gleichheitszeichens betreffen. Die Bedeutung dieser Fragen sei kurz anhand eines prominenten Beispiels erläutert.

In der für die Funktionentheorie wichtigen Potentialtheorie der Ebene beweist man durch spezielle Argumente und nicht ganz mühelos, daß es zu jeder kompakten Teilmenge C des $\mathbb{R}^2$ genau ein Wahrscheinlichkeitsmaß ν mit Träger in C gibt, welches das sogenannte Energieintegral $\int\int \log \frac{1}{|x-y|} \, d\nu(x) \, d\nu(y)$ minimiert, sofern es überhaupt ein Wahrscheinlichkeitsmaß mit Träger in C gibt, für welches das Integral endlich ist. Die Existenz einer minimierenden Verteilung folgt dabei leicht aus dem Lemma von Helly. Probleme bereitet dagegen der Beweis der Eindeutigkeit dieser sogenannten Gleichgewichtsverteilung[8].

Aus der in [5, Kapitel 3] beschriebenen Theorie folgt nun aber mühelos, daß für jedes $\epsilon > 0$ der durch

$$\tilde{K}_\epsilon(x, y) = \log(\epsilon + |x - y|^2)$$

[6] Ein signiertes Maß heißt molekular, wenn es sich als endliche Linearkombination von Diracmaßen schreiben läßt.

[7] Genauer: Negativ definit auf der Gruppe $(\mathbb{R}^d, +)$ bezüglich der Involution $x \to -x$. Siehe [5, Kapitel 4].

[8] Er wird in der grundlegenden Arbeit von Frostman [12] erst nach längerer Vorbereitung auf Seite 61 gegeben. Ähnlich muß bei der Lektüre des Standardwerkes [32] zwischen Behauptung (Seite 55) und Beweis (Seite 80) einiges bewältigt werden. Einen anderen, nicht ganz vollständigen, Beweis gab Hille in [16] (Theorem 16.4.3, p. 282). Er läßt sich vervollständigen, indem man sein vorbereitendes, aus [12][Seite 61] übernommenes, Theorem 16.4.2 auf Seite 281 durch das Lemma 2–1 aus [2, Seite 26] ersetzt (die Beweise sind im wesentlichen identisch). Dann erkennt man, daß aus der Endlichkeit von $I(\mu)$ und $I(\nu)$ die Endlichkeit von $I(\mu, \nu)$ folgt, womit die Zeile (16.4.12) auf Seite 283 in [16] gerechtfertigt ist.

definierte Kern $\tilde{K}_\epsilon$ negativ definit im Sinne der Beziehung (48) ist. Denn es ist

$$\tilde{K}_\epsilon(x,y) = \log(1 + \frac{1}{\epsilon}|x-y|^2) + \log\epsilon$$

und die negative Definitheit des ersten Summanden auf der rechten Seite folgt mit dem Beispiel (49) aus [5, Korollar 2.10, Seite 78], während die Konstante $\log\epsilon$ trivialerweise negativ definit ist.

Mittels Approximation beliebiger endlicher Maße durch molekulare Maße folgt daraus die Beziehung (47) mit $\tilde{K}_\epsilon$ an Stelle von $\tilde{K}$ für jedes $\epsilon > 0$, was sich in der Form

$$\begin{aligned}&\frac{1}{2}\Big(\int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\log(\epsilon+|x-y|^2)\,d\mu(x)\,d\mu(y)+\int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\log(\epsilon+|x-y|^2)\,d\nu(x)\,d\nu(y)\Big)\\ \leq\ &\int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\log(\epsilon+|x-y|^2)\,d\frac{\mu+\nu}{2}(x)\,d\frac{\mu+\nu}{2}(y)\end{aligned}$$

schreiben läßt. Durch Grenzübergang folgt mit dem Satz über die monotone Konvergenz

$$\begin{aligned}&\int\int\frac{1}{\log|x-y|}\,d\frac{\mu+\nu}{2}(x)d\frac{\mu+\nu}{2}(y)\\ &\leq \frac{1}{2}\left(\int\int\frac{1}{\log|x-y|}\,d\mu(x)d\mu(y)+\int\int\frac{1}{\log|x-y|}\,d\nu(x)d\nu(y)\right),\end{aligned} \tag{50}$$

das zu minimierende Funktional ist also konvex (insbesondere ist die Menge der Maße mit endlicher Energie konvex). Wenn man zeigen könnte, daß es sogar streng konvex ist, hätte man sofort die Eindeutigkeit der minimierenden Verteilung. Aber selbst wenn man das Gleichheitszeichen in (48) im vorliegenden Fall ausschließen könnte[9] , würde man mit dem obigen Beweis von (50) aufgrund der vollzogenen Grenzübergänge nicht die strenge Konvexität erhalten.

Wie sich zeigt, kann man aber einen Satz aufstellen, welcher der oben implizit benutzten Proposition 2.9 aus [5] entspricht, und welcher die strenge Konvexität des Energieintegrals als Spezialfall enthält.

Wir erinnern daran, daß eine auf einem Intervall I definierte Funktion g vollständig monoton heißt, wenn sie beliebig oft differenzierbar ist und $(-1)^n g^{(n)}(t) \geq 0$ für alle $x \in I$ und $n = 0, 1, \ldots$ gilt. Nach einem Satz von Bernstein (siehe Kapitel 8 dieser Arbeit) ist jede in $(0,\infty)$ vollständig monotone Funktion g Laplace-Transformierte eines nicht notwendig endlichen, positiven Borelmaßes m:

$$g(t) = \int_{[0,\infty)} e^{-\lambda t}\,dm(\lambda). \tag{51}$$

Wir setzen im folgenden $\varphi_+ := \max(\varphi, 0)$ und $\varphi_- := -\min(\varphi, 0)$ für $[-\infty,\infty]$-wertige Funktionen φ.

---

[9] Dies kann durch Verweis auf das Theorem 3', p. 823 in [24] geschehen.

Satz 6 *Die Funktion $\varphi : [0, \infty) \to [-\infty, \infty)$ sei stetig und habe in $(0, \infty)$ eine vollständig monotone Ableitung. μ und ν seien Wahrscheinlichkeitsmaße im $\mathbb{R}^d$. Für Funktionen $h : [0, \infty) \to [-\infty, \infty]$ sei zur Abkürzung*

$$E_{\mu\nu} h := \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} h(|x - y|^2)\, d\mu(x)\, d\nu(y)$$

gesetzt, sofern wenigstens Quasiintegrierbarkeit vorliegt. Dann gelten die folgenden Aussagen.

a) Die folgenden drei Bedingungen sind äquivalent:

(i) $E_{\mu\nu}\varphi_+ < \infty$

(ii) $E_{\mu\mu}\varphi_+ < \infty$ *und* $E_{\nu\nu}\varphi_+ < \infty$

(iii) $E_{\frac{\mu+\nu}{2}\frac{\mu+\nu}{2}}\varphi_+ < \infty$

b) Sind die Erwartungswerte aus a) endlich, so gilt

$$\frac{1}{2}(E_{\mu\mu}\varphi + E_{\nu\nu}\varphi) \le E_{\frac{\mu+\nu}{2}\frac{\mu+\nu}{2}}\varphi \le E_{\mu\nu}\varphi, \tag{52}$$

wobei $-\infty$ vorkommen kann.

c) Sind die Erwartungswerte aus a) endlich und ist zusätzlich $E_{\frac{\mu+\nu}{2}\frac{\mu+\nu}{2}}\varphi_- < \infty$, so sind alle in b) vorkommenden Erwartungswerte endlich und es gilt überall $<$, es sei denn, daß φ' konstant oder $\mu = \nu$ ist.

d) Die Menge

$$\left\{\mu : \mu \,\textit{Wahrscheinlichkeitsmaß im}\ \mathbb{R}^d\ \textit{mit}\ E_{\mu\mu}|\varphi| < \infty\right\}$$

ist konvex. Sofern sie nicht leer ist und φ' nicht konstant ist, ist $\mu \to E_{\mu\mu}\varphi$ streng konkav und

$$\rho(\mu, \nu) = \sqrt{2E_{\mu\nu}\varphi - E_{\mu\mu}\varphi - E_{\nu\nu}\varphi} = \sqrt{-E_{(\mu-\nu)(\mu-\nu)}\varphi}$$

definiert eine Metrik auf dieser Menge.

Zum Beweis des Satzes benötigen wir ein einfaches Lemma über konkave Funktionen.

Lemma 3 *Es sei $\varphi : (0, \infty) \to \mathbb{R}$ eine konkave Funktion. Dann gibt es ein $C \in \mathbb{R}$, so daß*

$$\varphi_+(s + t) \le C + \varphi_+(s) + \varphi_+(t)$$

für alle $s, t > 0$ gilt.

Beweis Es sei φ nach oben unbeschränkt (andernfalls ist nichts zu zeigen). Dann ist φ wachsend. Wir nehmen zunächst zusätzlich an, daß φ eine Nullstelle s_0 besitzt. Dann kann $C = \varphi(2s_0)$ gesetzt werden: Für $s \leq s_0$ und $t \leq s_0$ ist die behauptete Ungleichung offensichtlich. In den anderen Fällen nutzt man aus, daß $\varphi(t+s) - \varphi(s)$ für konkaves φ fallend in s ist. Im Fall $s \geq s_0, t \leq s_0$ ergibt sich

$$\varphi_+(s+t) = \varphi(s+t) - \varphi(s) + \varphi(s) \leq \varphi(s_0+t) - \varphi(s_0) + \varphi(s) \leq C + \varphi(s)$$

und damit die Behauptung, während im Fall $s \geq s_0, t \geq s_0$ entsprechend

$$\varphi_+(s+t) \leq \varphi(s_0+t) + \varphi(s) = \varphi(s_0+t) - \varphi(t) + \varphi(t) + \varphi(s) \leq \varphi(2s_0) + \varphi(t) + \varphi(s)$$

auf das Gewünschte führt.

Hat schließlich φ keine Nullstelle, so ist φ bei 0 stetig ergänzbar und man kann das vohergehende auf $\varphi - \varphi(0)$ mit $s_0 = 0$ anwenden. Das liefert die Behauptung mit $C = -\varphi(0)$. ∎

Beweis des Satzes 6 a) φ ist als Funktion mit vollständig monotoner Ableitung insbesondere wachsend und konkav. Mit der Ungleichung $|x-y|^2 \leq 2|x|^2 + 2|y|^2$ folgt aus dem Lemma

$$\varphi_+(|x-y|^2) \leq \varphi_+(2|x|^2 + 2|y|^2) \leq 3C + 2(\varphi_+(|x|^2) + \varphi_+(|y|^2)) \tag{53}$$

für $x, y \in \mathbb{R}^d$.

Es gelte nun die Bedingung (ii) in Aussage a). Dann existiert ein y_0 mit $\int \varphi_+(|x-y_0|^2)\, d\mu(x) < \infty$. Ersetzt man in (53) x durch $x - y_0$ und y durch $-y_0$, so erhält man die Endlichkeit von $\int \varphi_+(|x|^2)\, d\mu(x)$. Entsprechend ist $\int \varphi_+(|y|^2)\, d\nu(y)$ endlich. Zusammen folgt durch nochmaliges Anwenden von (53) die Aussage (i).

Mit der gleichen Schlußweise erkennt man, daß die Bedingung (i) die Bedingung (ii) zur Folge hat. Daß beide Bedingungen zusammen genommen äquivalent zur Bedingung (iii) sind, ist klar.

b) Aus der Darstellung (51) für $g = \varphi'$ erhält man für $t \in [0, \infty)$ und $\epsilon > 0$

$$\begin{aligned}
\varphi_\epsilon(t) &:= \varphi(\epsilon + t) \\
&= \varphi(\epsilon) + \int_\epsilon^{t+\epsilon} \varphi'(\tau)\, d\tau \\
&= \varphi(\epsilon) + \int_{[0,\infty)} \int_\epsilon^{t+\epsilon} e^{-\lambda\tau}\, d\tau\, dm(\lambda) \\
&= \varphi(\epsilon) + m_0 t + \int_{(0,\infty)} (1 - e^{-\lambda t}) \frac{e^{-\epsilon\lambda}}{\lambda}\, dm(\lambda),
\end{aligned} \tag{54}$$

wobei wir zur Abkürzung

$$m_0 := m(\{0\})$$

gesetzt haben. Dabei war die vorgenommene Änderung der Integrationsreihenfolge wegen $e^{-\lambda\tau} \geq 0$ legitim. Nach Lemma 3 hat man

$$\varphi(\epsilon) \leq \varphi_\epsilon(t) \leq C + \varphi(\epsilon)_+ + \varphi_+(t)$$

für ein $C \in \mathbb{R}$ und jedes $t \geq 0$. Mit der vorausgesetzten Endlichkeit der Erwartungswerte aus Teil a) folgt, daß die Erwartungswerte $E_{\mu\nu}\varphi_\epsilon$, $E_{\mu\mu}\varphi_\epsilon$, $E_{\nu\nu}\varphi_\epsilon$ und $E_{\frac{\mu+\nu}{2}\frac{\mu+\nu}{2}}\varphi_\epsilon$ im eigentlichen Sinne existieren (d. h. sie sind endlich). Mit (54) ergibt sich

$$\begin{aligned} E_{\mu\nu}\varphi_\epsilon &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \varphi_\epsilon(|x-y|^2)\, d\mu(x)d\nu(y) \\ &= \varphi(\epsilon) + m_0 \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |x-y|^2\, d\mu(x)d\nu(y) + I_\epsilon(\mu,\nu), \end{aligned} \tag{55}$$

mit

$$I_\epsilon(\mu,\nu) := \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{(0,\infty)} \left(1 - e^{-\lambda|x-y|^2}\right) \frac{e^{-\epsilon\lambda}}{\lambda}\, dm(\lambda)d\mu(x)d\nu(y).$$

Wir setzen nun

$$e^{-\lambda|x|^2}\, d\mu(x) =: d\mu_\lambda(x), \quad e^{-\lambda|y|^2}\, d\nu(y) =: d\nu_\lambda(y)$$

und erhalten wegen der Nichtnegativität des Integranden

$$\begin{aligned} I_\epsilon(\mu,\nu) &= \int_{(0,\infty)}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} (1 - e^{-\lambda|x-y|^2})\, d\mu(x)d\nu(y)\, \frac{e^{-\epsilon\lambda}}{\lambda}\, dm(\lambda) \\ &= \int_{(0,\infty)} \left(1 - \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} e^{2\lambda x'y} e^{-\lambda|x|^2}\, d\mu(x) e^{-\lambda|y|^2}\, d\nu(y)\right) \frac{e^{-\epsilon\lambda}}{\lambda}\, dm(\lambda) \\ &= \int_{(0,\infty)} \left(1 - \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \sum_{k=0}^{\infty} \frac{(2\lambda)^k}{k!}(x'y)^k\, d\mu_\lambda(x)d\nu_\lambda(y)\right) \frac{e^{-\epsilon\lambda}}{\lambda}\, dm(\lambda), \end{aligned} \tag{56}$$

wobei $x'y$ das Skalarprodukt bezeichnet. Wegen

$$\begin{aligned} & \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \sum_{k=0}^{\infty} \left|\frac{(2\lambda)^k}{k!}(x'y)^k\right| d\mu_\lambda(x)d\nu_\lambda(y) \\ \leq\ & \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \sum_{k=0}^{\infty} \frac{(2\lambda)^k}{k!}|x|^k|y|^k\, d\mu_\lambda(x)d\nu_\lambda(y) \\ =\ & \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} e^{\lambda(2|x||y|-|x|^2-|y|^2)}\, d\mu(x)d\nu(y) \\ =\ & \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} e^{-\lambda(|x|-|y|)^2}\, d\mu(x)d\nu(y) \\ \leq\ & 1 \\ <\ & \infty \end{aligned}$$

für $\lambda > 0$ können nach dem Satz von Fubini die Doppelintegration und die Summation in (56) miteinander vertauscht werden. Das liefert, wenn wir mit $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0{}^d$ Multiindizes, mit $|\alpha| = \alpha_1 + \ldots + \alpha_d$ ihre Länge und mit $\binom{\alpha}{k} = \frac{\alpha_1! \cdot \ldots \cdot \alpha_d!}{k!}$ den Multinomialkoeffizienten bezeichnen, sowie $x^\alpha := x_1{}^{\alpha_1} \cdot \ldots \cdot x_d{}^{\alpha_d}$ setzen,

$$I_\epsilon(\mu,\nu) = \int_{(0,\infty)} \left(1 - \sum_{k=0}^{\infty} \frac{(2\lambda)^k}{k!} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} (x'y)^k\, d\mu_\lambda(x)d\nu_\lambda(y)\right) \frac{e^{-\epsilon\lambda}}{\lambda}\, dm(\lambda)$$

$$= \int_{(0,\infty)} \left(1 - \sum_{k=0}^{\infty} \frac{(2\lambda)^k}{k!} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \sum_{|\alpha|=k} \binom{\alpha}{k} x^\alpha y^\alpha \, d\mu_\lambda(x) d\nu_\lambda(y)\right) \frac{e^{-\epsilon\lambda}}{\lambda} \, dm(\lambda)$$

$$= \int_{(0,\infty)} \left(1 - \sum_{k=0}^{\infty} \frac{(2\lambda)^k}{k!} \sum_{|\alpha|=k} \binom{\alpha}{k} \int_{\mathbb{R}^d} x^\alpha \, d\mu_\lambda(x) \int_{\mathbb{R}^d} y^\alpha \, d\nu_\lambda(y)\right) \frac{e^{-\epsilon\lambda}}{\lambda} \, dm(\lambda)$$

Setzt man diesen Ausdruck für $I_\epsilon(\mu, \nu)$ in die Formel (55) ein, so erhält man mit den entsprechenden Ausdrücken für $E_{\mu\mu}\varphi_\epsilon$ und $E_{\nu\nu}\varphi_\epsilon$

$$\begin{aligned}
& 2E_{\mu\nu}\varphi_\epsilon - E_{\mu\mu}\varphi_\epsilon - E_{\nu\nu}\varphi_\epsilon \\
= & \left|\int x \, d\mu(x) - \int y \, d\nu(y)\right|^2 \\
& + \int_{(0,\infty)} \sum_{k=0}^{\infty} \frac{(2\lambda)^k}{k!} \sum_{|\alpha|=k} \binom{\alpha}{k} \left(-2\int x^\alpha d\mu_\lambda \int y^\alpha d\nu_\lambda + \left(\int x^\alpha d\mu_\lambda\right)^2 + \left(\int y^\alpha d\nu_\lambda\right)^2\right) \frac{e^{-\epsilon\lambda}}{\lambda} \, dm(\lambda) \\
= & \left|\int x \, d\mu(x) - \int y \, d\nu(y)\right|^2 \\
& + \int_{(0,\infty)} \sum_{k=0}^{\infty} \frac{(2\lambda)^k}{k!} \sum_{|\alpha|=k} \binom{\alpha}{k} \left(\int x^\alpha \, d\mu_\lambda(x) - \int y^\alpha \, d\nu_\lambda(y)\right)^2 \frac{1}{\lambda} e^{-\epsilon\lambda} \, dm(\lambda), \qquad (57)
\end{aligned}$$

und das ist offensichtlich nichtnegativ. Es folgt die Ungleichungskette

$$\frac{1}{2}(E_{\mu\mu}\varphi_\epsilon + E_{\nu\nu}\varphi_\epsilon) \leq E_{\frac{\mu+\nu}{2}\frac{\mu+\nu}{2}}\varphi_\epsilon \leq E_{\mu\nu}\varphi_\epsilon .$$

Für $\epsilon \longrightarrow 0$ ergibt sich daraus die behauptete Ungleichung (52) durch Anwendung des Satzes über die monotone Konvergenz (auf die wachsende Folge der nichtnegativen Funktionen $\varphi_1 - \varphi_{\frac{1}{n}}$).

c) Nun werde zusätzlich zur Endlichkeit der Erwartungswerte aus a) angenommen, daß $E_{\frac{\mu+\nu}{2}\frac{\mu+\nu}{2}}\varphi_-$ endlich ist. Wegen b) ist dann auch $E_{\mu\nu}\varphi$ endlich. Damit müssen wegen

$$E_{\frac{\mu+\nu}{2}\frac{\mu+\nu}{2}}\varphi = \frac{1}{4}(E_{\mu\mu}\varphi + E_{\nu\nu}\varphi) + \frac{1}{2}E_{\mu\nu}\varphi \qquad (58)$$

auch $E_{\mu\mu}\varphi$ und $E_{\nu\nu}\varphi$ endlich sein.

Aus (57) ergibt sich mit dem Satz über die monotone Konvergenz

$$\begin{aligned}
& 2E_{\mu\nu}\varphi - E_{\mu\mu}\varphi - E_{\nu\nu}\varphi \\
= & \lim_{\epsilon\to 0} 2E_{\mu\nu}\varphi_\epsilon - E_{\mu\mu}\varphi_\epsilon - E_{\nu\nu}\varphi_\epsilon \\
= & \left|\int x \, d\mu(x) - \int y \, d\nu(y)\right|^2 \\
& + \int_{(0,\infty)} \sum_{k=0}^{\infty} \frac{(2\lambda)^k}{k!} \sum_{|\alpha|=k} \binom{\alpha}{k} \left(\int x^\alpha \, d\mu_\lambda(x) - \int y^\alpha \, d\nu_\lambda(y)\right)^2 \frac{1}{\lambda} \, dm(\lambda). \qquad (59)
\end{aligned}$$

Tritt nun an einer Stelle in (52) eine Gleichheit auf, so muß sie wegen (58) überall auftreten. Insbesondere muß damit der Ausdruck in (59) verschwinden. Ist nun aber $\mu \neq \nu$, so muß es zu jedem $\lambda > 0$ ein α geben, für welches $\int x^\alpha \, d\mu_\lambda(x) \neq \int y^\alpha \, d\nu_\lambda(y)$ ist, da die durch

$$
\begin{aligned}
M_{\mu_\lambda}(t) &= \int e^{tx} \, d\mu_\lambda(x) \\
&= \int e^{tx - \lambda|x|^2} \, d\mu(x), \\
M_{\nu_\lambda}(t) &= \int e^{ty - \lambda|y|^2} \, d\nu(y),
\end{aligned}
$$

$t \in \mathbb{R}^d$, gegebenen momentenerzeugenden Funktionen von μ_λ und ν_λ offenbar im ganzen $\mathbb{R}^d$ endlich sind und somit die Maße μ_λ und ν_λ durch ihre Momente eindeutig festgelegt sind. Der Integrand im dm-Integral in (59) ist also in ganz $(0, \infty)$ positiv. Ist nun φ' nicht konstant, so ist $m((0, \infty)) > 0$ und damit muß das dm -Integral in (59) positiv sein.

d) Daß die angegebene Menge konvex ist, folgt aus den Aussagen a) und b). Die strenge Konkavität folgt aus c). Die Gültigkeit der Dreiecksungleichung folgt mit der Minkowski-Ungleichung aus der Darstellung (59). Die Definitheit wurde gerade in c) festgestellt. ∎

4 Charakterisierung gegebener eindimensionaler Wahrscheinlichkeitsverteilungen durch Extremalprobleme

4.1 Ein negatives Resultat

Bevor wir im nächsten Abschnitt die Frage 1 durch Angabe eines geeigneten Kernes K bejahen, wollen wir ein nicht uninteressantes negatives Resultat beweisen. Betrachtet man den Kern $K(x, y) = |x-y|$, so kann man auf die Idee kommen, daß man zur Charakterisierung einer gegebenen Verteilung einen Kern der Form $K(x, y) = u(x - y)$ mit einer geeigneten Funktion u wählen kann. Überraschenderweise kann man viele „vernünftige" Verteilungen, darunter die Normalverteilung, so nicht charakterisieren, jedenfalls dann nicht, wenn man u einer naheliegenden Wachstumsbeschränkung unterwirft.

Es bezeichne $\mathcal{S}$ den Raum der schnell fallenden Funktionen ([22], Seite 168). Für $f \in \mathcal{S}$ sei $\hat{f}$ die zugehörige Fourier–Transformierte. Die Normalverteilungsdichte ist ein Element von $\mathcal{S}$ und erfüllt auch die übrigen Voraussetzungen des folgenden Satzes.

Satz 7 *Es sei $K(x,y) = u(x - y)$ mit einer geraden, Borel-meßbaren Funktion u, die der Bedingung*

$$|u(x)| \leq c(1 + |x|^2) \qquad (x \in \mathbb{R}) \tag{60}$$

für ein $c \in \mathbb{R}$ genügt. Es sei weiter μ ein Wahrscheinlichkeitsmaß auf $\mathbb{R}$ mit Dichte f. Es gelte

$$\operatorname{supp} \mu = \mathbb{R},$$

$$f \in \mathcal{S},$$

$$\hat{f}(x) \neq 0 \qquad (x \in \mathbb{R}).$$

Ist dann μ eine Lösung der Aufgabe (E) mit den Nebenbedingungen $E[X] = 0$, $E[X^2] = 1$, so ist u λ^1-fast überall gleich einem Polynom zweiten Grades.

Es folgt, daß alle Wahrscheinlichkeitsmaße mit λ^1-Dichte sowie Erwartungswert 0 und Varianz 1 ebenfalls das Extremalproblem (E) lösen, welches somit zur Charakterisierung von μ untauglich ist.

Beweis Ist μ eine Lösung mit den genannten Eigenschaften, so folgt ähnlich wie in Abschnitt 2.5

$$\int_{\mathbb{R}} u(x - y) f(y)\, dy + p(x) = 0 \qquad (x \in \mathbb{R})$$

mit einem Polynom p zweiten Grades. Man kann nun u als temperierte Distribution auffassen und erhält durch Fourier–Transformation (Theorem 7.19 c) in [22], Seite 179)

$$\hat{f}\hat{u} = -\hat{p}$$

Da $\hat{f}$ nach Voraussetzung keine Nullstellen hat, können wir die letzte Gleichung mit der glatten Funktion $\frac{1}{\hat{f}}$ multiplizieren und erhalten

$$\hat{u} = -\frac{1}{\hat{f}}\hat{p}.$$

Nun ist $\hat{p} = \sum_{j=0}^{2} \alpha_j \delta^{(j)}$ mit gewissen komplexen Zahlen α_j (siehe [22, Seite 177, Beispiel 7.16, Gleichung (5)]), wobei $\delta^{(j)}$ für die j-te Ableitung des Diracmaßes steht. Es folgt leicht

$$\hat{u} = \sum_{j=0}^{2} \tilde{\alpha}_j \delta^{(j)}$$

mit gewissen $\tilde{\alpha}_j$. Rücktransformation liefert

$$u = \tilde{p},$$

ein Polynom zweiten Grades, und die Behauptung folgt. ∎

Bezüglich der Voraussetzungen des Satzes 7 sei bemerkt, daß man die Bedingung $f \in \mathcal{S}$ vermutlich abschwächen kann. Vielleicht braucht man nicht einmal die Existenz einer Dichte von μ zu fordern. Wir wollen hier auf diese Fragen nicht weiter eingehen, da in Zusammenhang mit der Frage 1 eine Abschwächung der Voraussetzungen an μ keine wesentlich neuen Erkenntnisse mit sich brächte.

4.2 Ein positives Resultat

Trotz des Ergebnisses des vorangegangenen Abschnittes wird man sich in Hinblick auf eine positive Beantwortung der Frage 1 weiter nach Kernen umsehen, die dem Plackettschen ähneln. Eine sich anbietende Wahl ist

$$K(x,y) = |h(x) - h(y)| \tag{61}$$

mit einer noch zu bestimmenden Funktion h. Man kann nun die Multiplikatorenregel Satz 4 auch dazu benutzen, bei vorgegebener Lösung ein geeignetes h zu bestimmen.

Wir formulieren vorbereitend ein einfaches Lemma.

Lemma 4 *Ist $h : \mathbb{R} \to \mathbb{R}$ stetig und streng wachsend und ist F Verteilungsfunktion eines Wahrscheinlichkeitsmaßes, so ist*

$$\int_{\mathbb{R}} |h(x) - h(y)|\, dF(y) = \int_{h(x)}^{\infty} 1 - F(h^{-1}(t))\, dt + \int_{-\infty}^{h(x)} F(h^{-1}(t))\, dt,$$

wobei $h^{-1}(t) := +\infty$ für $t \geq \sup h$ und $h^{-1}(t) := -\infty$ für $t \leq \inf h$ gesetzt sei.

Beweis Es ist

$$\begin{aligned}
\int_{\mathbb{R}} (h(x) - h(y))_- \, dF(y) &= \int_{\mathbb{R}} \int_{\mathbb{R}} I(0 < t) I(t < h(y) - h(x)) \, dt \, dF(y) \\
&= \int_0^\infty \int_{\mathbb{R}} I(y > h^{-1}(t + h(x)) \, dF(y) dt \\
&= \int_0^\infty 1 - F(h^{-1}(h(x) + t)) \, dt \\
&= \int_{h(x)}^\infty 1 - F(h^{-1}(t)) \, dt
\end{aligned}$$

und

$$\begin{aligned}
\int_{\mathbb{R}} (h(x) - h(y))_+ \, dF(y) &= \int_{\mathbb{R}} \int_{\mathbb{R}} I(h(y) - h(x) \le t) I(t < 0) \, dt dF(y) \\
&= \int_{-\infty}^0 \int_{\mathbb{R}} I(y \le h^{-1}(h(x) + t) \, dF(y) dt \\
&= \int_{-\infty}^{h(x)} F(h^{-1}(t) \, dt.
\end{aligned}$$

■

Wir betrachten nun bei gegebenem μ_0 mit zugehöriger Verteilungsfunktion F_0 die Aufgabe (E) mit K wie in (61), aber mit noch zu bestimmendem h, und den Nebenbedingungen

$$E[X] = \int x \, dF_0(x), \qquad E[X^2] = \int x^2 \, dF_0(x).$$

Nehmen wir an, daß μ_0 eine Lösung dieser Aufgabe ist, so folgt durch formales Anwenden des Satzes 4 mit $\lambda_0 = 1$

$$l(x) = \int |h(x) - h(y)| \, dF_0(y) + p(x) \begin{cases} \ge 0 & (x \in \mathbb{R}) \\ = 0 & (x \in \operatorname{supp} \mu_0) \end{cases}$$

für ein Polynom p zweiten Grades. Nach Lemma 2 ergibt sich formal für $x \in \operatorname{supp} \mu$

$$-\int_{h(x)}^\infty 1 - F_0(h^{-1}(t)) \, dt - \int_{-\infty}^{h(x)} F(h^{-1}(t)) \, dt + p(x) = 0$$

und somit, falls differenzieren erlaubt ist,

$$(1 - 2F_0(x)) h'(x) = -p'(x),$$

also

$$h'(x) = \frac{ax + b}{F_0(x) - \frac{1}{2}}$$

für gewisse a, b. Daß dieses so gefundene h tatsächlich geeignet ist, beinhaltet der folgende Satz.

Satz 8 *F_0 sei eine stetige Verteilungsfunktion mit einem eindeutig bestimmten Median m_0 und endlichem zweiten Moment. Die durch*

$$\eta_0(x) := \frac{x - m_0}{F_0(x) - \frac{1}{2}} \tag{62}$$

definierte Funktion sei in einer Umgebung von m_0 integrierbar. h_0 bezeichne eine Stammfunktion von η_0. h sei eine streng wachsende stetige Funktion mit

$$h(x) = \begin{cases} = h_0(x) & (x \in \operatorname{supp} \mu_0) \\ \geq h_0(x) & (x \leq m_0) \\ \leq h_0(x) & (x \geq m_0). \end{cases}$$

Sind dann X_0 und Y_0 unabhängig und jeweils nach F_0 verteilt und X und Y unabhängig und identisch verteilt nach einer beliebigen Verteilung mit endlichem zweiten Moment, so folgt aus

$$E[(X - m_0)^2] \leq E[(X_0 - m_0)^2] \tag{63}$$

das Bestehen von

$$E[|h(X) - h(Y)|] \leq E[|h(X_0) - h(Y_0)|], \tag{64}$$

und Gleichheit tritt in der letzten Ungleichung genau dann ein, wenn X und X_0 dieselbe Verteilung besitzen.

Beweis Wir betrachten die Aufgabe (E) mit $d = 1$,

$$K(x, y) = -|h(x) - h(y)|,$$

$n = 1$, $g_1(x) = (x - m_0)^2$, $a_1 = E[(X_0 - m_0)^2]$ und mit der Nebenbedingung vom Ungleichungstyp (33). Es soll der Satz 5 benutzt werden.

Unter Berücksichtigung der Monotonie von h erhält man leicht

$$h(x) = O(|h_0(x)|) \qquad (|x| \to \infty).$$

Aus (62) folgt $\eta_0(x) \sim 2|x|$ für $|x| \to \infty$ und damit

$$h(x) = O(x^2) \qquad (|x| \to \infty).$$

Daher gilt für ein $c \in \mathbb{R}$

$$|K(x, y)| \leq c(1 + x^2)(1 + y^2) \qquad (x, y \in \mathbb{R}). \tag{65}$$

Damit sind die Voraussetzungen des Satzes 4 erfüllt.

Weiter ist das zu minimierende Funktional (40) streng konvex auf der Menge der Wahrscheinlichkeitsmaße mit endlicher Varianz: Es läßt sich nämlich in der Form

$$\mu \to \int_{\mathbb{R}} \int_{\mathbb{R}} |x - y| \, d\mu^h(x) \, d\mu^h(y) \tag{66}$$

schreiben, wobei μ^h das durch $\mu^h(B) := \mu(h^{-1}(B))$ definierte Bildmaß von μ unter h bezeichnet. Die Abbildung $\mu \longrightarrow \mu^h$ ist nun linear und injektiv (letzteres wegen der strengen Monotonie von h), und (66) ist als Komposition des bereits als streng konvex erkannten Plackettschen Funktionals (39) mit dieser Abbildung folglich streng konvex.

Wegen $a_1 > 0$ (F_0 ist ja stetig) ist auch die Slater-Bedingung erfüllt. Der Satz 5 ist also anwendbar, und wir haben nur zu zeigen, daß es ein $\alpha \in \mathbb{R}$ und ein $\lambda \geq 0$ gibt, mit

$$l(x) = -\int_{\mathbb{R}} |h(x) - h(y)|\, dF_0(y) + \alpha + \lambda(x - m_0)^2 \begin{cases} \geq 0 & (x \in \mathbb{R}) \\ = 0 & (x \in \operatorname{supp} \mu). \end{cases} \tag{67}$$

Es sei zunächst $h = h_0$. Dann ist Lemma 3 anwendbar und wir erhalten

$$l(x) = -\int_{h_0(x)}^{\infty} 1 - F_0(h_0^{-1}(t))\, dt - \int_{-\infty}^{h_0(x)} F_0(h_0^{-1}(t))\, dt + \alpha + \lambda(x - m_0)^2$$

und daraus

$$\begin{aligned} l'(x) &= (1 - 2F_0(x))\eta_0(x) + 2\lambda(x - m_0) \\ &= -2(x - m_0) + 2\lambda(x - m_0). \end{aligned}$$

l' verschwindet also für $\lambda = 1$ identisch, und damit ist gezeigt, daß (67) für ein $\lambda \geq 0$ und ein $\alpha \in \mathbb{R}$, welche nicht beide $= 0$ sind, gilt.

Nun sei nicht notwendig $h = h_0$. Es ist dann mit λ und α wie oben

$$\begin{aligned} l(x) &= -\int_x^{\infty} h(y) - h(x)\, dF_0(y) - \int_{-\infty}^{x} h(x) - h(y)\, dF_0(y) + \alpha + \lambda(x - m_0)^2 \\ &= \int_{-\infty}^{x} h(y)\, dF_0(y) - \int_x^{\infty} h(y)\, dF_0(y) + h(x)(1 - 2F_0(x)) + \alpha + \lambda(x - m_0)^2 \\ &= \int_{-\infty}^{x} h_0(y)\, dF_0(y) - \int_x^{\infty} h_0(y)\, dF_0(y) + h(x)(1 - 2F_0(x)) + \alpha + \lambda(x - m_0)^2. \end{aligned}$$

Dies ist nun im Träger von μ_0 gleich, und ansonsten größer oder gleich, dem Wert von $l(x)$, den wir mit h_0 an Stelle von h erhalten haben. Folglich gilt wiederum (67), womit Satz 8 bewiesen ist. ∎

Wir hätten in der Formulierung des Satzes $h = h_0$ setzen können, um zu einer übersichtlicheren Formulierung zu gelangen. Es wurde die allgemeinere Fassung gewählt, um den Satz von Plackett mit $h(x) = 2\sqrt{3}x$,

$$h_0(x) = \begin{cases} -x^2 - 3 & (x \leq -\sqrt{3}) \\ 2\sqrt{3}x & (-\sqrt{3} \leq x \leq \sqrt{3}) \\ x^2 + 3 & (x \geq \sqrt{3}) \end{cases}$$

als Spezialfall zu erhalten.

Es sei noch bemerkt, daß eine leichte Änderung der im obigen Beweis betrachteten Aufgabe (E) zur Nichtlösbarkeit führen kann: F_0 erfülle die Voraussetzungen des Satzes 8 und sei zusätzlich streng wachsend in $\mathbb{R}$ (also supp $\mu = \mathbb{R}$). Dann hat die Aufgabe (E) mit $d = 1$, $K(x,y) = -|h_0(x) - h_0(y)|$, $n = 1$ und $g_1(x) = (x - m_0)^2$ keine Lösung, wenn die Nebenbedingung in Gleichungsform (32) mit einem $a_1 > E[(X_0 - m_0)^2]$ gestellt wird. Zum Beweis nimmt man die Existenz einer Lösung F beziehungsweise μ an und überlegt sich, ähnlich wie in Abschnitt 2.5, daß die auf Seite 35 zur Motivation des Satzes 8 durchgeführten „formalen Schlüsse" mit F an Stelle von F_0 korrekt sind und auf

$$F(x) - \frac{1}{2} = a(F_0(x) - \frac{1}{2}) \qquad (x \in \text{supp } \mu)$$

führen. Aus dieser Beziehung leitet man leicht

$$\int_{-\infty}^{\infty} (x - m_0)^2 \, dF(x) \leq \int_{-\infty}^{\infty} (x - m_0)^2 \, dF_0(x)$$

her, was die Nichtzulässigkeit von F zeigt.

5 Charakterisierung mehrdimensionaler Gleichverteilungen

In diesem Kapitel soll die Frage 2 der Einleitung beantwortet werden. Wir suchen also eine Funktion $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, so daß die Aufgabe (E) etwa mit der Nebenbedingung

$$E[\|X\|^2] = a,$$

a eine positive Konstante, als Lösung nur die Gleichverteilung in einer Kugel um den Ursprung besitzt.

Um eine Idee zu erhalten, wie eine geeignete Funktion K aussehen könnte, betrachten wir erneut den in Abschnitt 2.5 gegebenen Beweis des Satzes von Plackett. Dort wurde im wesentlichen aus der für $x \in \operatorname{supp} \mu$ gültigen Gleichung

$$-\int |x-y|\, d\mu(y) = -p(x), \tag{68}$$

wobei p ein Polynom zweiten Grades bezeichnet, geschlossen, daß μ eine Gleichverteilung ist. Dies geschah durch einmaliges Differenzieren von (68), welches aufgrund einer vorherigen Umformung des Integrals auf elementare Weise möglich war. Man sieht aber deutlicher, warum dies auf die Gleichverteilung führen mußte, wenn man beide Seiten von (68) „direkt" zweimal differenziert: Im Sinne der Distributionentheorie ist

$$\frac{d^2}{dx^2}|\cdot| = 2\delta \tag{69}$$

und damit ergibt sich im Inneren des Gültigkeitsbereiches von (68), wenn $\star$ das Faltungsprodukt von Distributionen bedeutet[10],

$$\begin{aligned}
\mu &= \frac{1}{2} 2\delta \star \mu \\
&= \frac{1}{2}\left(\frac{d^2}{dx^2}|\cdot|\right) \star \mu \\
&= \frac{1}{2}\frac{d^2}{dx^2}(|\cdot| \star \mu) \\
&= \frac{1}{2}p'',
\end{aligned}$$

d. h. μ hat, zumindest im Inneren seines Trägers, eine konstante Dichte.

[10] Da diese Rechnung nur zur Motivation des folgenden durchgeführt wird, kann der Leser sie als Heuristik ansehen. Sie läßt sich aber streng rechtfertigen: Die ersten beiden Gleichheiten sind offensichtlich. Zur Begründung der dritten Gleichheit nehmen wir der Einfachheit halber an, daß μ kompakten Träger hat. Dann folgt sie aus Theoreme IX in [25, Seite 160] oder Theorem 6.37 in [22, Seite 160]. Die letzte Gleichheit schließlich gilt mindestens im Inneren des Gültigkeitsbereiches von (68), da Gleichheit von Distributionen eine lokale Eigenschaft ist, die sich auf ihre Ableitungen überträgt.

Von technischen Feinheiten einmal abgesehen kann man also die Beziehung (69) und die Tatsache, daß p'' konstant ist, als entscheidende Ursache für die Gültigkeit des Satzes von Plackett ansehen. Im Fall beliebiger Raumdimension $d \geq 1$ erhielte man für jede Lösung μ der betrachteten Extremalaufgabe (E) aus der Multiplikatorenregel Satz 4 formal

$$-\int K(x,y)\,d\mu(x) = a + b|x|^2 \qquad (x \in \text{supp}\,\mu) \tag{70}$$

mit reellen Konstanten a, b. Wendet man nun den Laplace-Operator $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}$ auf (70) an, so wird die rechte Seite konstant, und man wünscht sich in Analogie zu (69)

$$\Delta K(\cdot, y) = c_d \delta_y$$

mit δ_y = Diracmaß im Punkt y und einer reellen Konstanten c_d. Bekanntlich (siehe [25, Seite 44-45]) leistet dies der durch

$$K(x,y) = K_d(x,y) = \begin{cases} -|x-y| & (d=1) \\ \log \frac{1}{|x-y|} & (d=2) \\ \frac{1}{|x-y|^{d-2}} & (d \geq 3) \end{cases} \tag{71}$$

definierte Kern, wobei wir

$$K(x,x) = \infty \qquad (d \geq 2)$$

setzen wollen, und zwar mit den Konstanten

$$c_d = \begin{cases} -2 & (d=1) \\ -2\pi & (d=2) \\ -(d-2)\frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})} & (d \geq 3). \end{cases}$$

Tatsächlich gilt in Verallgemeinerung des Satzes von Plackett der folgende

Satz 9 *Es sei d eine natürliche Zahl und K_d wie in (71) definiert. Sind dann X und Y unabhängige und identisch verteilte Zufallsvektoren mit Werten im $\mathbb{R}^d$ und gilt*

$$E[|X|^2] \leq \frac{d}{d+2}, \tag{72}$$

so ist

$$E[K_d(X,Y)] \geq \begin{cases} -\frac{2}{3} & (d=1) \\ \frac{1}{2} & (d=2) \\ \frac{2d}{d+2} & (d \geq 3), \end{cases} \tag{73}$$

und Gleichheit tritt genau dann ein, wenn X und Y in der Einheitskugel $\{x \in \mathbb{R}^d : |x| \leq 1\}$ gleichverteilt sind.

Beweis Wir können uns auf den Fall $d \geq 2$ beschränken.

Wir betrachten die Aufgabe (E) mit $K = K_d$ und der einen Nebenbedingung (72). Obwohl wir in der Folge keinen Gebrauch davon machen werden, sei bemerkt, daß die Existenz einer Lösung sofort aus dem Lemma 1 folgt.

Da die durch $\varphi(t) = \frac{1}{2}\log t$ beziehungsweise $\varphi(t) = -t^{-(d-2)/2}$, $d \geq 3$, gegebenen Funktionen offenbar in $(0, \infty)$ vollständig monotone Ableitungen besitzen, ist der Satz 6 anwendbar. Wir haben also ein streng konvexes Funktional auf der konvexen Menge

$$W := \{\mu : \mu \text{ Wahrscheinlichkeitsmaß}, \int |x|^2 \, d\mu(x) \leq \frac{d}{d+2}, \quad \int\int |K_d| \, d\mu d\mu < \infty\}$$

zu minimieren. Daraus folgt schon, daß die Aufgabe (E) höchstens eine Lösung hat.

Nun ist der Satz 5 scheinbar nicht anwendbar, da die dort verlangten Regularitätsvoraussetzungen von K_d nicht erfüllt werden. Dennoch überträgt sich der damalige Beweis dafür, daß die Analoga der dortigen Bedingungen (41), (42) und (43) hinreichend für das Vorliegen eines Minimums sind, fast wörtlich auf den hier vorliegenden Fall:

Nach dem Satz von Kuhn–Tucker ist die Gleichverteilung in $\{x \in \mathbb{R}^d : |x| \leq 1\}$, die wir mit $\hat{\mu}$ bezeichnen wollen, sicher dann eine Lösung unserer Aufgabe (E), wenn es eine reelle Zahl $\lambda \geq 0$ gibt mit

$$\lambda \left(\int_{\mathbb{R}^d} |x|^2 \, d\hat{\mu}(x) - \frac{d}{d+2} \right) = 0 \tag{74}$$

derart, daß für die durch

$$\mathcal{L}(\mu) := \frac{1}{2} \int\int K_d \, d\mu d\mu + \lambda \int |x|^2 \, d\mu(x) \qquad (\mu \in W)$$

definierte Lagrangefunktion gilt:

$$\min_{\mu \in W} \mathcal{L}(\mu) = \mathcal{L}(\hat{\mu}). \tag{75}$$

Die Bedingung (74) ist für jedes λ erfüllt, da der Ausdruck in der Klammer verschwindet. Wegen der Konvexität des Funktionals $\mathcal{L}$ gilt (75) genau dann, wenn die hier offenbar existierende einseitige Richtungsableitung

$$\begin{aligned}
&\lim_{t \to 0+} \frac{1}{t} \left(\mathcal{L}(\hat{\mu} + t(\nu - \hat{\mu})) - \mathcal{L}(\hat{\mu}) \right) \\
&= \lim_{t \to 0+} \frac{1}{t} \left(\frac{t^2}{2} \int\int K \, d(\nu - \hat{\mu}) d(\nu - \hat{\mu}) + t \int\int K_d \, d\hat{\mu} d(\nu - \hat{\mu}) + t\lambda \int |x|^2 \, d(\nu - \hat{\mu})(x) \right) \\
&= \int\int K_d \, d\hat{\mu} d(\nu - \hat{\mu}) + \lambda \int |x|^2 \, d(\nu - \hat{\mu})(x) \\
&= \int \left(\int K_d(x, y) \, d\hat{\mu} + \lambda |x|^2 \right) d(\nu - \hat{\mu})(x)
\end{aligned}$$

von $\mathcal{L}$ an der Stelle $\hat{\mu}$ in Richtung $\nu - \hat{\mu}$ für jedes $\nu \in W$ nichtnegativ ist. Dafür reicht es hin, daß für ein $\alpha \in \mathbb{R}$

$$l(x) := \int K_d(x, y) \, d\hat{\mu}(y) + \lambda |x|^2 + \alpha \begin{cases} \geq 0 & x \in \mathbb{R}^d \\ = 0 & x \in \operatorname{supp} \hat{\mu} \end{cases} \tag{76}$$

gilt (Beweis wie in (46)).

Es bezeichne nun $\omega_d = 2\pi^{\frac{d}{2}}/\Gamma(\frac{d}{2})$ den Oberflächeninhalt der Einheitssphäre in $\mathbb{R}^d$. Damit ist für $d \geq 3$

$$
\begin{aligned}
\int_{\mathbb{R}^d} K_d(x,y)\, d\hat{\mu}(y) &= \frac{d}{\omega_d} \int_{\{|y| \leq 1\}} \frac{1}{|x-y|^{d-2}}\, dy \\
&= \frac{d}{\omega_d} \begin{cases} \frac{\omega_d}{2}(1 + |x|^2 \frac{2-d}{d}) & (|x| \leq 1) \\ \omega_d \frac{1}{d} |x|^{2-d} & (|x| \geq 1) \end{cases} && (77) \\
&= \begin{cases} \frac{d}{2} + |x|^2 \frac{2-d}{2} & (|x| \leq 1) \\ |x|^{2-d} & (|x| \geq 1) \end{cases}, && (78)
\end{aligned}
$$

wobei man eine Herleitung der bei (77) benutzten Formel etwa in [10, Seite 40] findet. Damit gilt (76) mit $\lambda = \frac{d-2}{2} > 0$ und $\alpha = -\frac{d}{2}$. Für $d \geq 3$ ist demnach die Gleichverteilung $\hat{\mu}$ tatsächlich extremal, und wir erhalten den Wert der rechten Seite von (73) aus (78):

$$
\begin{aligned}
\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K_d\, d\hat{\mu} d\hat{\mu} &= \frac{d}{\omega_d} \int_{\{|x| \leq 1\}} (\frac{d}{2} + \frac{2-d}{2}|x|^2)\, dx \\
&= \frac{d}{2} + \frac{2-d}{2} d \int_0^1 r^2 r^{d-1}\, dr \\
&= \frac{2d}{d+2}.
\end{aligned}
$$

Nun sei $d = 2$. Mit

$$
\int_0^{2\pi} \log|x - r(\cos t, \sin t)|\, dt = \begin{cases} 2\pi \log r & (|x| \leq r) \\ 2\pi \log|x| & (|x| \geq r) \end{cases}
$$

(siehe [16, Lemma 14.4.1, Seite 214]) erhält man für $x \in \mathbb{R}^2$

$$
\begin{aligned}
\int_{\mathbb{R}^2} K_2(x,y)\, d\hat{\mu}(y) &= \frac{1}{\pi} \int_{\{|y| \leq 1\}} \frac{1}{\log|x-y|}\, dy \\
&= -\frac{1}{\pi} \int_0^1 \int_0^{2\pi} \log|x - r(\cos t, \sin t)|\, dt\, r\, dr \\
&= \begin{cases} -2\int_0^{|x|} r \log|x|\, dr - 2\int_{|x|}^1 r \log r\, dr & (|x| \leq 1) \\ -2\log|x| \int_0^1 r\, dr & (|x| \geq 1) \end{cases} \\
&= \begin{cases} 1 - |x|^2 & (|x| \leq 1) \\ -\log|x| & (|x| \geq 1). \end{cases}
\end{aligned}
$$

Man erkennt, daß (76) mit $\lambda = 1 > 0$ und $\alpha = -1$ gilt. Also ist auch für $d = 2$ die Gleichverteilung extremal, und mit

$$
\int_{\mathbb{R}^2} \int_{\mathbb{R}^2} K_2\, d\hat{\mu} d\hat{\mu} = \frac{1}{\pi} \int_{\{|x| \leq 1\}} (1 - |x|^2)\, dx = \frac{1}{2}
$$

ist der Satz 9 vollständig bewiesen. ∎

6 Maximierung des erwarteten Abstandes bei gegebener mittlerer Varianz

In diesem Kapitel soll die Frage 3 der Einleitung behandelt werden. Wir betrachten also im $\mathbb{R}^d$ die Aufgabe (E) mit $K(x,y) = -|x-y|$ und der einen Nebenbedingung

$$E[|X|^2] \leq 1. \tag{79}$$

Wir können uns auf den Fall $d \geq 2$ beschränken, da der Fall $d = 1$ durch den Satz von Plackett geklärt ist.

Nach Lemma 1 (Seite 15) existiert jedenfalls eine Lösung der Aufgabe (E). Aus Homogenitätsgründen muß für jede Lösung in der Nebenbedingung (79) das Gleichheitszeichen eintreten. Weiter ist Satz 6 anwendbar, da die Funktion $\sqrt{\ }: [0,\infty) \to \mathbb{R}$ stetig ist und in $(0,\infty)$ eine vollständig monotone Ableitung besitzt. Das zu minimierende Funktional ist also sicherlich streng konvex auf der Menge der Wahrscheinlichkeitsmaße μ mit $\int_{\mathbb{R}^d} |x|^2 \, d\mu(x) < \infty$. Zusammen mit der bereits festgestellten Existenz einer Lösung erhalten wir die eindeutige Lösbarkeit von (E).

Weiter folgt, daß die Lösung μ invariant unter orthogonalen Transformationen ist. Denn für jede derartige Transformation T gilt

$$\int_{\mathbb{R}^d} |x|^2 \, d\mu(x) = \int_{\mathbb{R}^d} |x|^2 \, d\mu^T(x)$$

und

$$\int\int |x-y| \, d\mu(x) d\mu(y) = \int\int |x-y| \, d\mu^T(x) d\mu^T(y)$$

und somit ist μ^T auch Lösung von (E), also mit μ identisch.

Mit Satz 5, dessen Voraussetzungen hier erfüllt sind, erhält man zusammenfassend das folgende Zwischenergebnis:

Die Aufgabe (E) *hat für jedes* $d \geq 2$ *genau eine Lösung* μ. *Diese ist rotationssymmetrisch und erfüllt die Gleichung*

$$\int_{\mathbb{R}^d} |x|^2 \, d\mu(x) = 1. \tag{80}$$

Ein beliebiges Wahrscheinlichkeitsmaß μ *ist genau dann Lösung von* (E), *wenn* (80) *gilt und es ein* $\alpha \in \mathbb{R}$ *und ein* $\lambda \in [0,\infty)$ *gibt, so daß*

$$\int_{\mathbb{R}^d} |x-y| \, d\mu(y) - \alpha - \lambda |x|^2 \begin{cases} \leq 0 & (x \in \mathbb{R}^d) \\ = 0 & (x \in \operatorname{supp} \mu) \end{cases} \tag{81}$$

gilt.

Man kann nun, ähnlich wie in Kapitel 5, den Distributionenkalkül heranziehen, um zu einer Vermutung zu gelangen, wie μ aussehen muß. Für ungerade Raumdimensionen d ist nämlich (siehe [25, Seite 47, Formel (II,3;16)])

$$\Delta^{\frac{d+1}{2}} |\cdot| = c_d \delta$$

mit einer gewissen Konstanten $c_d > 0$.

Ist nun $d \in \{3, 5, \ldots\}$ und μ eine Lösung von (E), so ergibt sich durch formale Anwendung des Operators $\Delta^{\frac{d+1}{2}}$ auf die Gleichung

$$\int |x - y| \, d\mu(y) = \alpha + \lambda |x|^2 \tag{82}$$

im Inneren ihres Gültigkeitsbereiches die Beziehung

$$c_d \mu = \Delta^{\frac{d+1}{2}} (|\cdot| \star \mu) = 0,$$

wegen $\frac{d+1}{2} \in \{2, 3, \ldots\}$. μ verschwindet demnach im Inneren seines Trägers, was nur dann der Fall sein kann, wenn das Innere des Trägers leer ist. Bedenkt man die Rotationssymmetrie von μ, so drängt sich für $d = 3, 5, \ldots$ die Vermutung auf, daß μ die Gleichverteilung auf der Kugeloberfläche $\{x \in \mathbb{R}^d : |x| = 1\}$ sein muß. Der folgende Satz besagt, daß diese Vermutung sogar für jede Dimension $d \geq 3$ richtig ist.

Satz 10 *Es sei d eine natürliche Zahl ≥ 3. Sind dann X und Y unabhängige und identisch verteilte Zufallsvektoren im $\mathbb{R}^d$ mit*

$$E[|X|^2] \leq 1,$$

so ist

$$E[|X - Y|] \leq \sqrt{2} \frac{\Gamma^2(\frac{d}{2})}{\Gamma(\frac{d}{2} - \frac{1}{4}) \Gamma(\frac{d}{2} + \frac{1}{4})}$$

und Gleichheit tritt genau dann auf, wenn X und Y auf der Oberfläche der Einheitskugel $\{x \in \mathbb{R}^d : |x| = 1\}$ gleichverteilt sind.

Für $d = 2$ ergibt sich eine andersartige Extremalverteilung, die zwischen den beiden gegensätzlichen Fällen $d \geq 3$ (Verteilung von $|X|$ ist ein Diracmaß) und $d = 1$ (Verteilung von $|X|$ hat eine konstante Dichte in $[0, \sqrt{3}]$) vermittelt. Wir werden wieder durch Differenzieren der Beziehung (82) auf sie geführt. Im $\mathbb{R}^2$ gilt nämlich im Distributionensinne (siehe [25, Seite 45, Formel (II,3;8)])

$$\Delta |\cdot| = \frac{1}{|\cdot|}.$$

Die Anwendung des Laplace-Operators im Inneren von (82) sollte demnach so etwas wie

$$\int_{\mathbb{R}^2} \frac{d\mu(y)}{|x - y|} = \alpha \qquad (x \in \operatorname{supp} \mu \subset \mathbb{R}^2)$$

liefern. Diese Beziehung kann man nun auch „dreidimensional" lesen: Unter der plausiblen Annahme, daß der Träger von μ eine Kreisscheibe ist, besagt sie gerade, daß das räumliche

Potential von μ auf ihr konstant und somit μ die elektrostatische Gleichgewichtsverteilung auf der Kreisscheibe sein muß. Diese hat, für den Fall, daß der Radius gleich 1 ist, bekanntlich die Dichte (siehe [20, Seite 35, Hilfssatz V mit $\lambda = -1$])

$$f_0(x) = \frac{1}{2\pi} \frac{1}{\sqrt{1-|x|^2}} I(|x| < 1) \qquad (x \in \mathbb{R}^2) \tag{83}$$

bezüglich des zweidimensionalen Lebesguemaßes.

Tatsächlich gilt der

Satz 11 *Sind X und Y unabhängige und identisch verteilte Zufallsvektoren im $\mathbb{R}^2$ mit $E[|X|^2] \leq 1$, so ist*

$$E[|X-Y|] \leq \frac{\pi}{\sqrt{6}}$$

und Gleichheit tritt genau dann auf, wenn die Verteilung von X die durch

$$f(x) = \frac{1}{2\pi} \frac{\frac{2}{3}}{\sqrt{1-\frac{2}{3}|x|^2}} I(|x| < \sqrt{\frac{3}{2}}) \qquad (x \in \mathbb{R}^2)$$

gegebene Dichte f bezüglich des zweidimensionalen Lebesgue-Maßes hat.

Zum Beweis dieser Sätze benötigen wir Eigenschaften der für $d = 2, 3, \ldots$ durch

$$g_d(|x|) := \int_{\mathbb{R}^d} |x-y| \, d\omega(y)$$

definierten Funktion $g_d : [0, \infty) \to \mathbb{R}$ (ω bezeichnet wieder die Gleichverteilung auf $\{x \in \mathbb{R}^d : |x| = 1\}$). Das Integral hängt offenbar wirklich nur von $r := |x|$ ab und es ergibt sich für $d \geq 2$ (siehe etwa [18, Seite 8, Formel (1,2)], wo ω gleich ω_d mal dem hier verwendeten ω ist)

$$\begin{aligned} g_d(r) &= \int_{\mathbb{R}^d} \sqrt{1 + r^2 + 2xy} \, d\omega(y) \\ &= \frac{1}{B(\frac{d-1}{2}, \frac{1}{2})} \int_{-1}^{1} \sqrt{1 + r^2 + 2rt} \, (1-t^2)^{\frac{d-3}{2}} \, dt \qquad (r \in [0, \infty)), \end{aligned}$$

wobei B die Beta-Funktion bezeichnet.

Lemma 5 *Für $d = 2, 3, \ldots$ ist die oben definierte Funktion $g_d : [0, \infty) \to \mathbb{R}$ stetig differenzierbar. Ihre Ableitung g_d' ist beschränkt und in $[1, \infty)$ positiv und streng konkav. Für $d \geq 3$ ist zusätzlich g_d' in $[0, \infty)$ konkav.*

Beweis Es sei zunächst $d \in \{2, 3, \ldots\}$. Für $-1 < t < 1$, $0 \leq r < \infty$ und

$$g(r, t) := \sqrt{1 + r^2 + 2rt}$$

ist

$$
\begin{aligned}
\frac{\partial}{\partial r} g(r,t) &= (r+t)(1+r^2+2rt)^{-\frac{1}{2}} && (84)\\
\left|\frac{\partial}{\partial r} g(r,t)\right| &\leq 1 && (85)\\
\frac{\partial^2}{\partial r^2} g(r,t) &= (1-t^2)(1+r^2+2rt)^{-\frac{3}{2}} && (86)\\
&= (1+r^2)^{-\frac{3}{2}}(1-t^2)(1+\frac{2r}{1+r^2}t)^{-\frac{3}{2}}\\
&\leq (1-t^2)(1-|t|)^{-\frac{3}{2}}\\
&= (1+|t|)^{\frac{3}{2}}(1-t^2)^{-\frac{1}{2}}\\
&\leq 2\sqrt{2}(1-t^2)^{-\frac{1}{2}} && (87)\\
\frac{\partial^3}{\partial r^3} g(r,t) &= -3(1-t^2)(r+t)(1+r^2+2rt)^{-\frac{5}{2}}. && (88)
\end{aligned}
$$

Dabei ergibt sich (85) durch Quadrieren von (84) mittels $t^2 \leq 1$.

Nun ist an jeder Stelle $r \in [0,\infty) \setminus \{1\}$

$$
g_d(r) = \frac{1}{B(\frac{d-1}{2},\frac{1}{2})} \int_{-1}^{1} g(r,t)(1-t^2)^{\frac{d-3}{2}}\, dt
$$

dreimal differenzierbar und die Ableitungen können unter dem Integralzeichen ausgeführt werden (denn für $|r-1| \geq \delta > 0$ und $|t| \leq 1$ ist $0 \leq (1+r^2+2rt)^{-1} \leq (1+r^2-2r)^{-1} \leq \delta^{-2}$, womit die benötigten partiellen Ableitungen von $g(r,t)$ dem Betrage nach durch Konstanten abgeschätzt werden können (siehe (84), (86) und (88)), was wegen der Integrierbarkeit von $(1-t^2)^{\frac{d-3}{2}} \leq (1-t^2)^{-\frac{1}{2}}$ die Anwendung des Satzes über die majorisierte Konvergenz ermöglicht):

$$
g_d^{(l)}(r) = \frac{1}{B(\frac{d-1}{2},\frac{1}{2})} \int_{-1}^{1} \frac{\partial^l}{\partial r^l} g(r,t) \left(1-t^2\right)^{\frac{d-3}{2}} dt \qquad (r \in [0,\infty) \setminus \{1\}, l = 1,2,3) \tag{89}
$$

Man erkennt, daß g_d in $[0,\infty)$ stetig differenzierbar ist, denn g_d ist stetig und die für $r \neq 1$ existierende Ableitung ist stetig und bei $r = 1$ stetig ergänzbar (Satz über die majorisierte Konvergenz angewandt auf die rechte Seite von (89) für $l = 1$, nach (85) kann $(1-t^2)^{\frac{d-3}{2}}$ als integrierbare Majorante gewählt werden).

Die behauptete Beschränktheit von g_d' ergibt sich mit (89) unmittelbar aus (85), während die Positivität in $[1,\infty)$ aus (84) folgt. Schließlich ist g_d' streng konkav in $[1,\infty)$, weil g_d' an der Stelle 1 stetig ist und nach (88) und (89) mit $l = 3$ in $(1,\infty)$ eine negative zweite Ableitung besitzt.

Nun sei $d \geq 3$. Dann ist g_d' stetig differenzierbar (Schlußweise wie oben, Majorante jetzt nach (87) $2\sqrt{2}(1-t^2)^{-\frac{1}{2}}(1-t^2)^{\frac{d-3}{2}} \leq 2\sqrt{2}(1-t^2)^{-\frac{1}{2}}$). Eine stetig differenzierbare Funktion auf $[0,\infty)$ ist dort genau dann konkav, wenn sie in $(0,1)$ und in $[1,\infty)$ konkav ist. Zum Nachweis der Konkavität von g_d' in $[0,\infty)$ braucht also nur noch

$$g_d'''(r) \leq 0 \qquad (0 < r < 1,\ d = 3,4,\ldots) \tag{90}$$

gezeigt zu werden.

Wegen

$$\begin{aligned} g_3(r) &= \frac{1}{B(1,\frac{1}{2})}\int_{-1}^{1}\sqrt{1+r^2+2rt}\,dt \\ &= \frac{1}{2}\frac{1}{3r}\left((1+r)^3 - |1-r|^3\right) \\ &= \begin{cases} 1+\frac{1}{3}r^2 & (0 \leq r \leq 1) \\ \frac{1}{3r}+r & (r \geq 1) \end{cases} \end{aligned}$$

ist

$$g_3'''(r) = \begin{cases} 0 & (0 \leq r < 1) \\ -\frac{2}{r^4} & (r > 1), \end{cases}$$

was die Gültigkeit von (90) für $d = 3$ zeigt.

Es sei nun $d > 3$ und $r \in (0,1)$ fest gewählt. Die durch

$$\begin{aligned} h(t) &:= \frac{\partial^3}{\partial r^3}g(r,t) + \frac{\partial^3}{\partial r^3}g(r,-t) \\ &= -3(1-t^2)\left((r+t)(1+r^2+2rt)^{-\frac{5}{2}} + (r-t)(1+r^2-2rt)^{-\frac{5}{2}}\right) \end{aligned} \tag{91}$$

definierte Funktion besitzt in (0,1) genau eine Nullstelle $t_0 = t_0(r)$ und h ist in $(0,t_0)$ negativ und in $(t_0,1)$ positiv. Denn es ist $h(0) < 0$ und die Funktion

$$\eta(t) = \frac{t-r}{t+r} - \frac{(1+r^2-2rt)^{\frac{5}{2}}}{(1+r^2+2rt)^{\frac{5}{2}}}$$

hat in $(0,1)$ dieselben Nullstellen wie h und ist wegen

$$\eta(t) = 1 - \frac{2r}{t+r} - \left(-1 + \frac{2(1+r^2)}{1+r^2+2rt}\right)^{\frac{5}{2}}$$

dort streng wachsend mit $\eta(0) = -2 < 0$ und $\eta(1) = \frac{1-r}{1+r} - \left(\frac{1-r}{1+r}\right)^5 > 0$.

Mit diesem t_0 ist nach (89) mit $l = 3$ und h wie in (91) unter Ausnutzung des gerade behandelten Falles $d = 3$

$$\begin{aligned} g_d'''(r) &= \frac{1}{B(\frac{d-1}{2},\frac{1}{2})}\int_0^1 h(t)(1-t^2)^{\frac{d-3}{2}}\,dt \\ &\leq \frac{1}{B(\frac{d-1}{2},\frac{1}{2})}\left(\int_0^{t_0} h(t)(1-t_0^2)^{\frac{d-3}{2}}\,dt + \int_{t_0}^1 h(t)(1-t_0^2)^{\frac{d-3}{2}}\,dt\right) \\ &= \frac{1}{B(\frac{d-1}{2},\frac{1}{2})}(1-t_0^2)^{\frac{d-3}{2}} g_3'''(r) \\ &= 0. \end{aligned}$$

Damit gilt (90) auch für $d > 3$ und das Lemma ist vollständig bewiesen. ∎

Beweis des Satzes 10 Da die gleich herzuleitende Formel (92) auch im Beweis des Satzes 11 Verwendung findet, werde zunächst nur $d \geq 2$ angenommen. μ bezeichne die nach dem Zwischenergebnis existierende Lösung. Wegen der sphärischen Symmetrie von μ existiert ein Wahrscheinlichkeitsmaß ν auf $[0, \infty)$, so daß sich μ (bis auf ein eventuell vorhandenes Atom im Nullpunkt) als Produktmaß von ν mit ω schreiben läßt. Im Fall $d \geq 3$ ist zu zeigen, daß ν im Punkt 1 konzentriert ist.

Jedenfalls ist für $r = |x| \geq 0$

$$\begin{aligned} l_0(r) &:= \int_{\mathbb{R}^d} |x-y|\, d\mu(y) \\ &= \nu(\{0\})r + \int_{(0,\infty)} \rho \int_{\{y\in\mathbb{R}^d:|y|=1\}} |\frac{1}{\rho}x - y|\, d\omega(y)\, d\nu(\rho). \\ &= \nu(\{0\})r + \int_{(0,\infty)} \rho g_d(\frac{r}{\rho})\, d\nu(\rho) \end{aligned}$$

mit g_d wie im Lemma 5. Damit ergibt sich die Differenzierbarkeit von l_0 mit der Ableitung

$$l_0'(r) = \nu(\{0\}) + \int_{(0,\infty)} g_d'(\frac{r}{\rho})\, d\nu(\rho) \qquad (r \geq 0,\ d \geq 2) \tag{92}$$

mit Hilfe des Satzes über die majorisierte Konvergenz, da nach Lemma 5 die Funktion g_d differenzierbar und ihre Ableitung und damit auch der Integrand in

$$\frac{l_0(r+\epsilon) - l_0(r)}{\epsilon} = \nu(\{0\}) + \int_{(0,\infty)} \rho \frac{g_d(\frac{r+\epsilon}{\rho}) - g_d(\frac{r}{\rho})}{\epsilon}\, d\nu(\rho)$$

dem Betrage nach durch eine Konstante beschränkt ist.

Von nun an nutzen wir aus, daß $d \geq 3$ ist. Da g_d' nach Lemma 5 konkav in $[0, \infty)$ und streng konkav in $[1, \infty)$ ist, ist für jedes $\rho > 0$ die Funktion $r \longrightarrow g_d'(\frac{r}{\rho})$ konkav in $[0, \infty)$ und streng konkav in $[\rho, \infty)$. Damit folgt aus (92): Ist für ein $a \in (0, \infty)$ $\nu((0, a+\epsilon]) > 0$ für jedes $\epsilon > 0$, so ist l_0' streng konkav in $[a, \infty)$.

Um einen Widerspruch zu erhalten nehmen wir nun an, daß es Zahlen $r_1, r_2 \in \operatorname{supp} \nu$ mit $0 < r_1 < r_2 < \infty$ gibt. Dann gilt nach dem Vorhergehenden für die Funktion

$$l(r) := l_0(r) - \alpha - \lambda r^2 \qquad (r \in [0, \infty)),$$

daß ihre Ableitung

$$l'(r) = l_0'(r) - 2\lambda r \qquad (r \in [0, \infty)) \tag{93}$$

in $[r_1, \infty)$ streng konkav ist.

Andererseits ist $l(r)$ die linke Seite von (81), für $r = |x|$ und gewisse α, λ. Damit ist $l(r_1) = l(r_2) = 0$, was die Existenz eines $\xi \in (r_1, r_2)$ mit $l'(\xi) = 0$ zur Folge hat. Zusammen mit $l'(r_1) = l'(r_2)$ (r_1 und r_2 sind nach (81) Maximalstellen von l) ergibt sich daraus der gewünschte Widerspruch zur strengen Konkavität.

ν kann also nur noch von der Form $\nu(\{0\})\delta_0 + (1-\nu(\{0\}))\delta_r$ mit einem $r > 0$ sein. Nimmt man $\nu(\{0\}) > 0$ an, so folgt aus (81) $l'(0) \leq 0$. Andererseits ist nach (92), (93) und Lemma 5 $l'(0) = \nu(\{0\}) + (1-\nu(\{0\}))g_g'(1) \geq \nu(\{0\})$. Also muß doch $\nu(\{0\}) = 0$ sein und damit tatsächlich (wegen (80)) $\nu = \delta_1$ und somit ω extremal.

Es bleibt der Extremalwert zu berechnen:

$$\begin{aligned}
\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |x-y|\, d\omega(y)d\omega(x) &= \int_{\mathbb{R}^d} g_d(|x|)\, d\omega(x) \\
&= g_d(1) \\
&= \frac{1}{B(\frac{d-1}{2},\frac{1}{2})}\int_{-1}^{1}\sqrt{2+2t}(1-t^2)^{\frac{d-3}{2}}\,dt \\
&= \frac{\sqrt{2}}{B(\frac{d-1}{2},\frac{1}{2})}\int_{-1}^{1}(1+t)^{\frac{d-2}{2}}(1-t)^{\frac{d-3}{2}}\,dt \\
&= \frac{\sqrt{2}}{B(\frac{d-1}{2},\frac{1}{2})}2^{d-\frac{3}{2}}\int_{0}^{1}t^{\frac{d}{2}-1}(1-t)^{\frac{d-1}{2}-1}\,dt \\
&= \sqrt{2}\,2^{d-\frac{3}{2}}\frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d-1}{2})\Gamma(\frac{1}{2})}\frac{\Gamma(\frac{d}{2})\Gamma(\frac{d-1}{2})}{\Gamma(d-\frac{1}{2})} \\
&= \sqrt{2}\,2^{2(\frac{d}{2}-\frac{1}{4})-1}\frac{\Gamma^2(\frac{d}{2})}{\pi^{\frac{1}{2}}\Gamma(2(\frac{d}{2}-\frac{1}{4}))} \\
&= \sqrt{2}\frac{\Gamma^2(\frac{d}{2})}{\Gamma(\frac{d}{2}-\frac{1}{4})\Gamma(\frac{d}{2}+\frac{1}{4})}.
\end{aligned}$$

Dabei ergibt sich das letzte Gleichheitszeichen aus der Legendreschen Verdopplungsformel $2^{2z-1}\Gamma(z)\Gamma(z+\frac{1}{2}) = \pi^{\frac{1}{2}}\Gamma(2z)$ mit $z = \frac{d}{2}-\frac{1}{4}$. ■

Zum Beweis des Satzes 11 wird ein weiteres Lemma benötigt.

Lemma 6 *Für jede komplexe Zahl z ist*

$$l_0(z) := \frac{1}{2\pi}\int_0^1\int_0^{2\pi}(1-r^2)^{-\frac{1}{2}}|z-re^{i\varphi}|\,r\,drd\varphi = \begin{cases} \frac{\pi}{4}+\frac{\pi}{8}|z|^2 & (|z|\leq 1) \\ |z|F(-\frac{1}{2},-\frac{1}{2};\frac{3}{2};\frac{1}{|z|^2}) & (|z|\geq 1), \end{cases}$$

wobei F die Hypergeometrische Funktion bezeichnet[11].

Beweis Aus Stetigkeitsgründen braucht der Fall $|z| = 1$ nicht betrachtet zu werden.

Es sei zunächst $|z| < 1$. Wie in [20, Seite 35] schreiben wir $\zeta = re^{i\varphi}$ und wenden die Substitution

$$\zeta = \frac{\zeta' + z}{1+\bar{z}\zeta'} \tag{94}$$

[11] Für $|t| \leq 1$ ist $F(-\frac{1}{2},-\frac{1}{2};\frac{3}{2};t) = \frac{3}{4}(1-t)^{\frac{1}{2}} + \frac{1}{4}(1+2t)t^{-\frac{1}{2}}\arcsin t^{\frac{1}{2}}$, was jedoch für das folgende keine Vereinfachung mit sich bringt.

an, welche das Innere des Einheitskreises auf sich abbildet, womit sich

$$l_0(z) = \frac{1}{2\pi}(1-|z|^2)^{\frac{5}{2}} \int_0^1 \int_0^{2\pi} \frac{r(1-r^2)^{-\frac{1}{2}}}{|1+\bar{z}re^{i\varphi}|^4} r\,dr d\varphi$$

ergibt.

Mit

$$\frac{1}{2\pi}\int_0^{2\pi} \frac{d\varphi}{|1+ae^{i\varphi}|^4} = \frac{1+|a|^2}{|1-|a|^2|^3} \qquad (a \in \mathbf{C},\ |a| \neq 1) \tag{95}$$

angewandt auf $a = \bar{z}r$ erhalten wir

$$\begin{aligned} l_0(z) &= (1-|z|^2)^{\frac{5}{2}} \int_0^1 \frac{r^2(1-r^2)^{-\frac{1}{2}}}{(1-|z|^2r^2)^3}(1+|z|^2r^2)\,dr \\ &= \frac{1}{2}(1-|z|^2)^{\frac{5}{2}} \int_0^1 \frac{t^{\frac{1}{2}}(1-t)^{-\frac{1}{2}}}{(1-|z|^2t)^3}(1+|z|^2t)\,dt \\ &= \frac{1}{2}(1-|z|^2)^{\frac{5}{2}} \left\{ \int_0^1 \frac{t^{\frac{1}{2}}(1-t)^{-\frac{1}{2}}}{(1-|z|^2t)^2}\,dt + 2|z|^2 \int_0^1 \frac{t^{\frac{3}{2}}(1-t)^{-\frac{1}{2}}}{(1-|z|^2t)^3}\,dt \right\}. \end{aligned}$$

Wendet man nun die Formel

$$\int_0^1 \frac{t^{a-1}(1-t)^{b-1}}{(1-\epsilon t)^{a+b}}\,dt = \frac{B(a,b)}{(1-\epsilon)^a} \qquad (a>0, b>0, \epsilon<1),$$

welche sich ergibt, wenn man in $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$ die Substitution $x = \frac{(1-\epsilon)t}{1-\epsilon t}$ vornimmt, auf die beiden Integrale in den geschweiften Klammern an (mit $b = \frac{1}{2}$ und $\epsilon = |z|^2$ sowie $a = \frac{3}{2}$ im ersten und $a = \frac{5}{2}$ im zweiten Integral), so folgt schließlich wie behauptet

$$\begin{aligned} l_0(z) &= \frac{1}{2}(1-|z|^2)^{\frac{5}{2}} \left\{ \frac{B(\frac{3}{2},\frac{1}{2})}{(1-|z|^2)^{\frac{3}{2}}} + 2|z|^2 \frac{B(\frac{5}{2},\frac{1}{2})}{(1-|z|^2)^{\frac{5}{2}}} \right\} \\ &= \frac{\pi}{4}(1-|z|^2) + \frac{3\pi}{8}|z|^2 \\ &= \frac{\pi}{4} + \frac{\pi}{8}|z|^2. \end{aligned}$$

Nun sei $|z| > 1$. Wir wenden wieder die Substitution (94) an, welche jetzt aber das $l_0(z)$ definierende Integral in ein Integral über das Äußere des Einheitskreises überführt:

$$l_0(z) = \frac{1}{2\pi}(|z|^2-1)^{\frac{5}{2}} \int_1^\infty \int_0^{2\pi} \frac{r}{(r^2-1)^{\frac{1}{2}}} \frac{1}{|1+\bar{z}re^{i\varphi}|^4} r\,dr d\varphi.$$

Mit (95) erhalten wir

$$l_0(z) = (|z|^2-1)^{\frac{5}{2}} \int_1^\infty \frac{r^2}{(r^2-1)^{\frac{1}{2}}} \frac{1+|z|^2r^2}{(|z|^2r^2-1)^3}\,dr.$$

Durch die Substitution $r^2 = \frac{1}{t}$ wird

$$
\begin{aligned}
l_0(z) &= \frac{1}{2}(|z|^2-1)^{\frac{5}{2}} \int_0^1 \frac{1}{(1-t)^{\frac{1}{2}}} \frac{|z|^2+t}{(|z|^2-t)^3}\, dt \\
&= \frac{1}{2}(|z|^2-1)^{\frac{5}{2}} \left\{ \int_0^1 (1-t)^{-\frac{1}{2}}(|z|^2-t)^{-2}\, dt + 2\int_0^1 t(1-t)^{-\frac{1}{2}}(|z|^2-t)^{-3}\, dt \right\}.
\end{aligned}
$$

Wir setzen noch zur Abkürzung

$$r := |z|$$

und erhalten mit einer Integraldarstellung Hypergeometrischer Funktionen [1, Seite 558, Formel 15.3.1]

$$
\begin{aligned}
l_0(z) &= \frac{(r^2-1)^{\frac{5}{2}}}{2r^4} \left\{ \int_0^1 t^{1-1}(1-t)^{\frac{3}{2}-1-1}(1-\frac{1}{r^2}t)^{-2}\, dt + \frac{2}{r^2}\int_0^1 t^{2-1}(1-t)^{\frac{5}{2}-1-1}(1-\frac{1}{r^2}t)^{-3}\, dt \right\} \\
&= \frac{1}{2} r(1-\frac{1}{r^2})^{\frac{5}{2}} \left\{ \frac{\Gamma(1)\Gamma(\frac{3}{2}-1)}{\Gamma(\frac{3}{2})} F(2,1;\frac{3}{2};\frac{1}{r^2}) + \frac{2}{r^2}\frac{\Gamma(1)\Gamma(\frac{5}{2}-1)}{\Gamma(\frac{5}{2})} F(3,2;\frac{5}{2};\frac{1}{r^2}) \right\} \\
&= r(1-\frac{1}{r^2})^{\frac{5}{2}} \left\{ F(2,1;\frac{3}{2};\frac{1}{r^2}) + \frac{4}{3}\frac{1}{r^2} F(3,2;\frac{5}{2};\frac{1}{r^2}) \right\}.
\end{aligned}
$$

Mit der Bezeichnung $(a)_n := a(a+1)\cdot \ldots \cdot (a+n-1)$ ist für $|x| < 1$

$$
\begin{aligned}
F(2,1;\frac{3}{2};x) + \frac{4}{3}xF(3,2;\frac{5}{2};x) &= \sum_{n=0}^{\infty} \frac{(2)_n(1)_n}{(\frac{3}{2})_n}\frac{x^n}{n!} + \sum_{n=0}^{\infty} \frac{(2)_{n+1}(2)_n}{(\frac{3}{2})_{n+1}}\frac{x^{n+1}}{n!} \\
&= \sum_{n=0}^{\infty} \frac{(2)_n(1)_n}{(\frac{3}{2})_n}\frac{x^n}{n!} + \sum_{n=0}^{\infty} n\frac{(2)_n(1)_n}{(\frac{3}{2})_n}\frac{x^n}{n!} \\
&= \sum_{n=0}^{\infty} \frac{(2)_n(2)_n}{(\frac{3}{2})_n}\frac{x^n}{n!} \\
&= F(2,2;\frac{3}{2};x) \\
&= (1-x)^{-\frac{5}{2}} F(-\frac{1}{2},-\frac{1}{2};\frac{3}{2};x)
\end{aligned}
$$

(für die letzte Umformung siehe [1, Formel 15.3.3]).

Somit erhalten wir schließlich wie behauptet

$$
\begin{aligned}
l_0(z) &= r(1-\frac{1}{r^2})^{\frac{5}{2}}(1-\frac{1}{r^2})^{-\frac{5}{2}} F(-\frac{1}{2},-\frac{1}{2};\frac{3}{2};\frac{1}{r^2}) \\
&= |z| F(-\frac{1}{2},-\frac{1}{2};\frac{3}{2};\frac{1}{|z|^2}).
\end{aligned}
$$

■

Beweis des Satzes 11 Zur Vereinfachung der folgenden Rechnungen betrachten wir an Stelle von $E[|X|^2] \le 1$ die Nebenbedingung

$$E[|X|^2] \le \frac{2}{3}.$$

Eine leichte Rechnung zeigt nämlich, daß

$$E[|X|^2] = \frac{2}{3} \tag{96}$$

gilt, wenn X die Dichte f_0 aus (83) besitzt.

Aus Homogenitätsgründen ist die Extremalität der in Satz 11 genannten Dichte $f(x) = \frac{2}{3}f_0(\sqrt{\frac{2}{3}}\,x)$ daher nachgewiesen, wenn gezeigt ist, daß das zu f_0 gehörige Wahrscheinlichkeitsmaß μ die hinreichende Bedingung (81) mit geeigneten α und λ erfüllt.

Nun ist für $r = |x|$ mit $x \in \mathbb{R}^2$ nach Lemma 6

$$\begin{aligned} l(r) &:= \int_{\mathbb{R}^2} |x-y|\,d\mu(y) - \frac{\pi}{4} - \frac{\pi}{8}r^2 \\ &= l_0(r) - \frac{\pi}{4} - \frac{\pi}{8}r^2 \\ &= \begin{cases} 0 & (0 \le r \le 1) \\ rF(-\frac{1}{2},-\frac{1}{2};\frac{3}{2};\frac{1}{r^2}) - \frac{\pi}{4} - \frac{\pi}{8}r^2 & (r \ge 1). \end{cases} \end{aligned} \tag{97}$$

Es ist also nur noch zu zeigen, daß der für $r \ge 1$ erhaltene Ausdruck kleiner oder gleich 0 ist.

Nach (92) existiert l' in $[0,\infty)$ und ist in $[1,\infty)$ konkav. Explizit erhält man aus (97) unter Benutzung von [1, Formel 15.2.1]

$$l'(r) = \begin{cases} 0 & (0 \le 1 \le 1) \\ F(-\frac{1}{2},-\frac{1}{2};\frac{3}{2};\frac{1}{r^2}) - \frac{1}{3r^2}F(\frac{1}{2},\frac{1}{2};\frac{5}{2};\frac{1}{r^2}) - \frac{\pi}{4}r & (r \ge 1). \end{cases}$$

Es wird behauptet, daß l' in $[0,\infty)$ konkav ist. Dann muß nämlich $l' \le 0$ in $[0,\infty)$ sein, woraus die gewünschte Ungleichung $l \le 0$ in $[1,\infty)$ folgt.

Da nun l' in $[0,1]$ und $[1,\infty)$ jeweils konkav ist, braucht nur die Differenzierbarkeit an der Stelle 1 nachgewiesen zu werden, welche sich aber aus der stetigen Ergänzbarkeit von

$$l''(r) = \begin{cases} 0 & (0 \le r < 1) \\ \frac{1}{3r^2}F(\frac{1}{2},\frac{1}{2};\frac{5}{2};\frac{1}{r^2}) + \frac{1}{15r^5}F(\frac{3}{2},\frac{3}{2};\frac{7}{2};\frac{1}{r^2}) - \frac{\pi}{4} & (r > 1) \end{cases}$$

an der Stelle 1 ergibt: Nach [1, Formel 15.1.20] ist nämlich

$$\begin{aligned} \lim_{r\to 1+} l''(r) &= \frac{1}{3}F(\frac{1}{2},\frac{1}{2};\frac{5}{2};1) + \frac{1}{15}F(\frac{3}{2},\frac{3}{2};\frac{7}{2};1) - \frac{\pi}{4} \\ &= \frac{1}{3}\frac{\Gamma(\frac{5}{2})\Gamma(\frac{3}{2})}{\Gamma(2)\Gamma(2)} + \frac{1}{15}\frac{\Gamma(\frac{7}{2})\Gamma(\frac{1}{2})}{\Gamma(2)\Gamma(2)} - \frac{\pi}{4} \\ &= \frac{\pi}{8} + \frac{\pi}{8} - \frac{\pi}{4} \\ &= 0. \end{aligned}$$

Damit ist die Extremalität der angegebenen Dichte f bewiesen.

Es bleibt der Extremalwert zu berechnen. Wenn X und Y unabhängig sind und jeweils die Dichte f_0 besitzen, so ist wegen Lemma 6 und (96)

$$\begin{aligned} E[|X-Y|] &= \int_0^1 \int_0^{2\pi} (\frac{\pi}{4} + \frac{\pi}{8} r^2) \frac{1}{2\pi} \frac{1}{\sqrt{1-r^2}} r \, dr \, d\varphi \\ &= \frac{\pi}{4} + \frac{\pi}{8} \cdot \frac{2}{3} \\ &= \frac{\pi}{3}. \end{aligned}$$

Für die bezüglich der Nebenbedingung $E[|X|^2] \leq 1$ extremale Dichte f_0 ergibt sich somit als Extremalwert wie behauptet $\sqrt{\frac{3}{2}} \frac{\pi}{3} = \frac{\pi}{\sqrt{6}}$. ■

7 Eine Charakterisierung mehrdimensionaler Gleichverteilungen mittels Varianzabschätzungen für unimodale Verteilungen

Es soll eine im eindimensionalen Fall wohlbekannte Varianzabschätzung für gewisse unimodale Verteilungen [17] auf höhere Dimensionen übertragen werden. Dabei wird ein von Anderson [3] eingeführter Unimodalitätsbegriff verwendet.

Definition *Eine Wahrscheinlichkeitsverteilung im $\mathbb{R}^d$ heißt unimodal, wenn sie eine Dichte f besitzt, so daß für jedes $\eta \geq 0$ die Menge*

$$\{x \in \mathbb{R}^d : f(x) \geq \eta\}$$

konvex ist. Sie heißt unimodal mit Modalwert 0, wenn zusätzlich die oben angegebene Menge für jedes η den Ursprung enthält oder leer ist. Entsprechend heißt dann die Dichte f unimodal beziehungsweise unimodal mit Modalwert 0.

Für eine Diskussion verschiedener Unimodalitätsbegriffe verweisen wir auf [9].

Im folgenden bezeichnet $\mathrm{tr}(A)$, $det(A)$ und A' die Spur, Determinante und Transponierte der Matrix A und

$$B := \{x \in \mathbb{R}^d : |x| \leq 1\}$$

die abgeschlossene euklidische Einheitskugel.

Satz 12 *Der Zufallsvektor X habe im $\mathbb{R}^d$ eine unimodale Verteilung mit Modalwert 0, deren Träger in B enthalten sei.*

Dann ist

$$\frac{1}{d} tr(Cov(X)) \leq \frac{1}{d+2}, \tag{98}$$

wobei Gleichheit genau dann eintritt, wenn X in B gleichverteilt ist.

Mit der Ungleichung $(det(Cov(X)))^{\frac{1}{d}} \leq \frac{1}{d} tr(Cov(X))$, welche sich nach Diagonalisierung aus der Ungleichung vom geometrischen und arithmetischen Mittel ergibt, erhalten wir das

Korollar *Satz 12 bleibt gültig, wenn die mittlere Varianz $\frac{1}{d} tr(Cov(X))$ durch die verallgemeinerte Varianz $(det(Cov(X)))^{\frac{1}{d}}$ ersetzt wird.*

Im Beweis des Satzes 12 wird klar, daß die Ungleichung (98) auch unter einer schwächeren Unimodalitätsannahme gültig ist, wobei aber die Charakterisierung der der Gleichverteilung verloren geht. Siehe dazu die Bemerkungen nach dem Beweis des Satzes 12.

Es sei darauf hingewiesen, daß man bereits im eindimensionalen Fall die Voraussetzung „mit Modalwert 0" des Satzes 12 nicht ersatzlos streichen kann. Dies wird in [17] und [26] diskutiert. In der letztgenannten Arbeit werden auch Irrtümer anderer Arbeiten korrigiert.

Der Beweis des Satzes 12 stützt sich auf vier Lemmata, von denen die ersten beiden die naheliegende Beweisidee enthalten, während die letzten beiden nur zum Beweis der Eindeutigkeitsaussage verwendet werden.

Lemma 7 *Es sei f eine sphärisch symmetrische unimodale Dichte eines Zufallsvektors X, deren Träger in B enthalten ist. Dann ist*

$$E[|X|^2] \leq \frac{d}{d+2},$$

und Gleichheit tritt genau für die Dichte der Gleichverteilung ein.

Beweis Es ist $f(x) = h(|x|)$ mit einer fallenden Funktion h, welche als rechtsseitig stetig in $(0, \infty)$ angenommen werden kann und somit in $[1, \infty)$ verschwindet. Wir bezeichnen mit ω_d den Oberflächeninhalt der Einheitssphäre und erhalten

$$\begin{aligned} E[|X|^2] &= \omega_d \int_{(0,1]} r^{d+1} h(r) dr \\ &= -\omega_d \int_{(0,1]} \frac{1}{d+2} r^{d+2} dh(r) \\ &= \frac{d}{d+2} \int_{(0,1])} r^2 \frac{\omega_d}{d} r^d d(-h(r)). \end{aligned}$$

Nun entspricht $\frac{\omega_d}{d} r^d d(-h(r))$ der Integration bezüglich eines Wahrscheinlichkeitsmaßes auf $(0, 1]$, was man durch eine partielle Integration einsieht, und wir erhalten offensichtlich genau dann den Maximalwert von $E[|X|^2]$, wenn diese Wahrscheinlichkeitsmaß an der Stelle 1 konzentriert ist, was der Funktion $h(r) = \frac{d}{\omega_d} I(r < 1)$ entspricht. ∎

Übrigens erhält man in der gleichen Weise entsprechende Abschätzungen für $E[\varphi(|X|)]$, wenn φ eine wachsende und etwa nichtnegative Funktion ist.

Lemma 8 *Ist X ein Zufallsvektor mit endlichen zweiten Momenten und hat X^* die Verteilung welche man aus der von X durch sphärische Symmetrisierung erhält, so ist*

$$Cov(X^*) = \frac{1}{d}(tr(Cov(X)) + |E[X]|^2) I_d \quad ,$$

wobei I_d die d-dimensionale Einheitsmatrix bezeichnet.

Beweis X^* ist verteilt wie UX, wobei U eine zufällige und von X stochstisch unabhängige Matrix ist, welche in der Orthogonalen Gruppe gleichverteilt ist (damit ist gemeint: Für jede feste orthogonale Matrix U_0 haben die zufälligen Matrizen U und U_0U dieselbe Verteilung). Daraus folgt $E[UX] = 0$ und somit

$$Cov(X^*) = E[UX(UX)'] = E[UXX'U'] = U_0 E[UXX'U'] U_0'$$

für jedes feste U_0. Die symmetrische Matrix $Cov(X^*)$ ist daher reelles Vielfaches von I_d, und Berechnug der Spur liefert

$$Cov(X^*) = \frac{1}{d} tr(E[XX'])I_d.$$

Die Behauptung des Lemmas folgt nun aus $E[XX'] = Cov(X) + E[X]E[X']$. ∎

Lemma 9 *Es sei g eine Wahrscheinlichkeitsdichte im $\mathbb{R}^d$ mit existierendem Erwartungswertvektor a. Wenn dann $K = \{x : g(x) > 0\}$ konvex ist, so ist a ein innerer Punkt von K.*

Beweis Es sei Y ein Zufallsvektor mit der oben genannten Dichte g. Es werde angenommen, daß a kein innerer Punkt von K ist. Dann gibt es ein von 0 verschiedens $l \in \mathbb{R}^d$ und ein $c \in \mathbb{R}$ mit $l'a \leq c$ und $l'x \geq c$ für $x \in K$ (siehe etwa die Korollare zu Satz 2.2.1 in [7]). Aus $E[l'X] = l'\mu \leq c$ folgt dann aber $l'X = c$ fast sicher, was der vorausgesetzten Existenz einer Dichte von X widerspricht. ∎

Lemma 10 *Zu jeder unimodalen Wahrscheinlichkeitsdichte f gibt es eine nach unten halbstetige und unimodale Version $\underline{f}$.*

Beweis Wir setzen $\underline{f}(x) := \liminf_{y \to x} f(y) := \lim_{\epsilon \to 0} \inf_{|x-y|<\epsilon} f(y)$. Dann ist $\underline{f}$ offenbar nach unten halbstetig und wegen

$$\{x : \underline{f}(x) \geq h\} = \bigcap_{\epsilon > 0} int\{x : f(x) \geq h - \epsilon\}$$

auch unimodal. Um einzusehen, daß $\underline{f}$ nur auf einer Nullmenge von f verschieden ist, schreibt man

$$\{\underline{f} \neq f\} = \{\liminf_{y \to x} f(y) < f(x)\} = \bigcup_{r \in \mathbb{Q}} \{\liminf_{y \to x} < r \leq f(x)\} \subset \bigcup_{r \in \mathbb{Q}} \{f \geq r\} \setminus int\{f \geq r\}$$

und benutzt die Tatsache, daß der Rand einer konvexen Menge das Maß 0 hat (siehe [13, Satz 3.1.20]). ∎

Im folgenden benutzen wir die Bezeichnung „Radius" für jede Menge der Form

$$\{tx : 0 < t < 1\}$$

mit einem $x \in \mathbb{R}^d$ mit $|x| = 1$.

Beweis des Satzes 12 Zu gegebenem X mit unimodaler Dichte f mit Modalwert 0 ist mit U und X^* wie im Beweis des Lemmas 8

$$f^*(x) = E[f(Ux)] \tag{99}$$

eine Dichte von X^*. f ist nach Voraussetzung fallend auf jedem Radius. Wegen (99) hat auch f^* diese Eigenschaft. Folglich ist Lemma 7 auf f^* und X^* anwendbar, und wir erhalten durch Bildung der Spur in Lemma 8

$$tr(Cov(X)) = tr(Cov(X^*)) - |E[X]|^2 \leq tr(Cov(X^*)) \leq \frac{d}{d+2},$$

womit (98) bewiesen ist.

Nun werde angenommen, daß Gleichheit in (98) eintritt. Dann muß

$$E[X] = 0 \tag{100}$$

gelten, und wegen der Eindeutigkeitsaussage des Lemmas 7 muß f^* Dichte der Gleichverteilung sein. Mit der radialen Monotonie von f und der Beziehung (99) folgt daraus

$$f \text{ ist konstant auf fast jedem Radius.} \tag{101}$$

Es bleibt zu zeigen, daß die Konstanten in (101) nicht von den Radien abhängen. Dazu dürfen wir nach Lemma 10 annehmen, daß f eine Dichte von X ist, welche nach unten halbstetig und unimodal ist und für die (100) und (101) gelten (eventuell ist aber 0 im Sinne unserer Definition kein Modalwert für diese Version der Dichte von X).

Wir führen einen indirekten Beweis und nehmen also an, daß

$$m := \text{ess inf}\,\{f(y) : y \in B\} < f(x)$$

auf einer Menge positiven Maßes gilt („ess inf" steht für das „essentielle Infimum", siehe [6, Seite 77]). Wir betrachten die durch

$$g(x) := \frac{1}{1 - m\int_B dx}(f(x) - m)$$

definierte Wahrscheinlichkeitsdichte. Nach Lemma 9 ist der zu g gehörige Erwartungswertvektor a ein innerer Punkt von $\{x \in \mathbb{R}^d : f(x) > m\}$. Nun ist wegen (100) und $\int_B x\,dx = 0$

$$\begin{aligned} 0 &= E[X] \\ &= \int_B x f(x)\,dx \\ &= \int_B x(f(x) - m)\,dx \\ &= (1 - m\int_B dx)a, \end{aligned}$$

also $a = 0$ und damit 0 innerer Punkt von $\{x \in \mathbb{R}^d : f(x) > m\}$. Das hat zur Folge, daß es keinen Radius gibt, auf dem f kleiner oder gleich m ist. Im Widerspruch dazu läßt sich aber ein solcher Radius wegen (101), der Bedeutung von m und der Halbstetigkeit von f durch ein offensichtliches Folgenkompaktheitsargument konstruieren. ∎

Bemerkungen (i) Der obige Beweis der Ungleichung (98) benutzt außer der Beschränktheit des Trägers nur, daß eine Dichte von X auf jedem Radius fallend ist. Diese Eigenschaft bezeichnet man als Stern-Unimodalität (engl. star unimodality). Allerdings wird in der Klasse der Stern-unimodalen Verteilungen mit Träger in B die Gleichverteilung nicht durch Maximierung der mittleren Varianz charakterisiert, da für jede derartige Dichte mit Erwartungswertvektor 0, die auf jedem Radius konstant ist, das Gleichheitszeichen in (98) eintritt.

(ii) Wir hätten uns technische Details im Beweis der Eindeutigkeitsaussage ersparen können, wenn wir von vornherein die Stetigkeit von f angenommen hätten. Dann folgt nämlich aus (101) sofort die Konstanz von f in B. Durch diese Schlußweise wäre jedoch die Bedeutung der Konvexität nicht klar geworden und man wäre darauf geführt worden, die Unimodalitätsvoraussetzung des Satzes 12 durch Stern-Unimodalität und Stetigkeit zu ersetzen, was aber nach der vorhergehenden Bemerkung unnatürlich erscheint.

(iii) Wie eingangs erwähnt, erhält man im Fall $d = 1$ eine von der Gleichverteilung verschiedene Extremalverteilung, wenn man die Varianz innerhalb der Klasse *aller* unimodalen Verteilungen mit Träger in $[-1, 1]$ maximieren möchte. Es wäre interessant, das entsprechende Extremalproblem für höhere Dimensionen zu lösen.

8 Satz von Bernstein, Umkehrformel von Post–Widder und Eindeutigkeitssatz für die Laplace–Transformation

Ist μ ein Wahrscheinlichkeitsmaß auf $[0,\infty)$, so bezeichnet man die durch

$$\varphi(t) = \int_{[0,\infty)} e^{-tx}\, d\mu(x) \tag{102}$$

definierte Funktion $\varphi : [0,\infty) \to \mathbb{R}$ als Laplace–Transformierte von μ. Es gilt dann:

(i) φ ist stetig in $[0,\infty)$ mit $\varphi(0) = 1$.

(ii) φ ist in $(0,\infty)$ beliebig oft differenzierbar mit $(-1)^n\varphi^{(n)}(t) \geq 0$ $(t > 0, n = 0, 1, \ldots)$.

Satz 13 (Bernstein) *Jede Funktion $\varphi : [0,\infty) \to \mathbb{R}$, welche die obigen Bedingungen (i) und (ii) erfüllt, ist Laplace-Transformierte eines Wahrscheinlichkeitsmaßes.*

Beweis: Es kann zusätzlich zu (i) und (ii) das Bestehen von

(iii) $\varphi(\infty) := \lim_{t\to\infty} \varphi(t) = 0$

angenommen werden (andernfalls betrachte man $\frac{\varphi-\varphi(\infty)}{1-\varphi(\infty)}$ an Stelle von φ). Dabei existiert der Limes, weil φ fallend und nichtnegativ ist.

Wir fixieren ein $t > 0$. Die Taylor–Entwicklung von φ um einen Punkt $A > 0$ mit Restglied in Integralform ergibt für jede natürliche Zahl n

$$\varphi(t) = \sum_{k=0}^{n-1} \frac{(-1)^k\varphi^{(k)}(A)}{k!}(A-t)^k + \int_t^A \frac{(s-t)^{n-1}}{(n-1)!}(-1)^{(n)}\varphi^{(n)}(s)\, ds.$$

Läßt man bei festem $n \in \mathbb{N}$ in dieser Formel A gegen unendlich streben, so konvergiert das Integral gegen einen endlichen Grenzwert, da der Integrand nach Voraussetzung nichtnegativ ist und die Integrale durch $\varphi(t)$ beschränkt sind. Folglich konvergiert für jedes n die Summe, was zur Folge hat, daß die einzelnen Summanden konvergieren. Deren Grenzwerte können aber wegen $\lim_{A\to\infty} \frac{A-t}{A} = 1$ nicht von t abhängen. Es gibt also Konstanten c_n, so daß

$$\begin{aligned}
\varphi(t) &= c_n + \int_t^\infty \frac{(s-t)^{n-1}}{(n-1)!}(-1)^n\varphi^{(n)}(s)\, ds \\
&= c_n + \int_0^\infty \left(1-\frac{t}{s}\right)_+^{n-1} \frac{(-1)^n}{(n-1)!} s^{n-1}\varphi^{(n)}(s)\, ds
\end{aligned} \tag{103}$$

für jedes $t > 0$ und $n \in \mathbb{N}$ gilt, wobei in (103) und im folgenden x_+ für das Maximum von x und 0 steht. Läßt man bei festem $n \geq 2$ nun t gegen unendlich streben, so konvergiert das

Integral in (103) nach dem Satz über die majorisierte Konvergenz gegen 0. Die c_n müssen wegen (iii) also alle verschwinden.

Mit der Substitution $s = \frac{n}{x}$ ergibt sich demnach

$$\varphi(t) = \int_0^\infty \left(1 - \frac{tx}{n}\right)_+^{n-1} f_n(x)\, dx \tag{104}$$

für $t > 0$, wobei

$$f_n(x) := \frac{(-1)^n}{n!}\left(\frac{n}{x}\right)^{n+1} \varphi^{(n)}\left(\frac{n}{x}\right) I(x > 0) \tag{105}$$

gesetzt wurde.

Mit Hilfe des Satzes über die monotone Konvergenz erkennt man, daß (104) auch für $t = 0$ gilt. Die f_n sind somit Wahrscheinlichkeitsdichten. Nach dem Auswahlsatz von Helly gibt es eine Teilfolge (f_{n_k}) von (f_n) derart, daß die Folge der zugehörigen Maße vag gegen ein nichtnegatives Maß μ auf $[0, \infty)$ mit $\mu([0, \infty)) \leq 1$ konvergiert. Zusammen mit der leicht einzusehenden Tatsache, daß $(1 - \frac{y}{n})_+^{n-1}$ in $[0, \infty)$ *gleichmäßig* gegen e^{-y} konvergiert, folgt daraus das Bestehen von (102). Für $t = 0$ ergibt sich nachträglich, daß μ tatsächlich ein Wahrscheinlichkeitsmaß ist. ∎

Wenn man nun den Eindeutigkeitssatz für Laplace-Transformierte heranzieht, wird im nachhinein klar, daß der Übergang zu einer Teilfolge im obigen Beweis unnötig war. Wir erhalten somit als Korollar den

Satz 14 (Umkehrformel von Post-Widder) *Ist φ die Laplace-Transformierte des Wahrscheinlichkeitsmaßes μ, so ist die durch (105) definierte Funktion für jede natürliche Zahl n eine Wahrscheinlichkeitsdichte auf $[0, \infty)$, und die Folge der zugehörigen Wahrscheinlichkeitsmaße konvergiert nach Verteilung gegen μ.*

Andererseits kann man, ähnlich wie in [6, Seite 29], die Post-Widdersche Umkehrformel direkt nachweisen (siehe hierzu auch [11, Kapitel 7.6]), um damit den hier sonst nicht verwendeten Eindeutigkeitssatz zu gewinnen: Bei Gültigkeit von (102) ist ja

$$f_n(x) = \frac{(-1)^n}{n!}\left(\frac{n}{x}\right)^{n+1} \int_0^\infty (-1)^n y^n e^{-\frac{n}{x}y}\, d\mu(y),$$

womit sich die zu f_n gehörige Verteilungsfunktion F_n in der Form

$$F_n(A) = \int_0^A f_n(x)\, dx = \int_0^\infty \int_{\frac{1}{A}}^\infty \frac{(ny)^n}{(n-1)!} x^{n-1} e^{-nyx}\, dx\, d\mu(y)$$

darstellen läßt. Man erkennt nun, daß das innere Integral in der obigen Formel die Wahrscheinlichkeit dafür ist, daß eine Gamma-verteilte Zufallsvariable mit Erwartungswert $\frac{1}{y}$ und Varianz $\frac{1}{ny^2}$ einen Wert $\geq \frac{1}{A}$ annimmt (siehe [11, Seite 47]). Diese Wahrscheinlichkeit konvergiert für $A \neq y$ bei $n \to \infty$ gegen $I(y < A)$. Damit ergibt sich, wenn A eine Stetigkeitsstelle der Verteilungsfunktion F von μ ist, die Konvergenz von $F_n(A)$ gegen $F(A)$, und das ist die Behauptung des Satzes 14.

Mit Hilfe des Eindeutigkeitssatzes erhält man nun in bekannter Weise (siehe [5, Korollar 6.14, Seite 135]) das folgende Korollar zu Satz 14, welches beim Beweis des Satzes 6 dieser Arbeit verwendet wurde:

Korollar *Erfüllt* $\varphi : (0, \infty) \to \mathbb{R}$ *die Bedingung (ii), so ist* φ *die Laplace-Transformierte eines nichtnegativen Maßes auf* $[0, \infty)$.

Bemerkung Der oben ausgeführte Beweis des Satzes von Bernstein ist nicht wirklich neu. Er findet sich im wesentlichen in [33, Seite 139–147]. In der hier aufgeschriebenen Weise hat er aber den Vorzug, leicht motivierbar zu sein und zugleich einen bequemen Zugang zum Eindeutigkeitssatz und zu einer der vielen Umkehrformeln für die Laplace-Transformation zu eröffnen.

Literatur

[1] M. Abramowitz und I. A. Stegun (1965). *Handbook of Mathematical Functions.* Dover, New York.

[2] L. V. Ahlfors (1973). *Conformal Invariants. Topics in Geometric Function Theory.* McGraw-Hill, New York. [MR **50**(1975)10211]

[3] T. W. Anderson (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6**, 170-176. [MR **16**(1955) 1005; Zbl **66**(1956)374]

[4] L. Baringhaus und N. Henze (1990). A consistent test for uniformity with unknown limits based on D'Agostinos D. *Statist. Probab. Lett.* **9**, No. 4.

[5] C. Berg, J. P. R. Christensen und P. Ressel (1984). *Harmonic Analysis on Semigroups.* Springer-Verlag, New York. [MR **86**b:43001]

[6] P. Billingsley (1986). *Probability and Measure. Second Edition.* Wiley, New York. [MR **87**f:60001]

[7] D. Blackwell und M. A. Girshick (1954). *Theory of games and statistical decisions.* Dover Publications, New York. [MR **16**(1955)1135]

[8] H. A. David (1981). *Order Statistics. Second Edition.* Wiley, New York. [MR **82**i:62073]

[9] S. W. Dharmadhikari und K. Joag-dev (1988). *Unimodality, Convexity and Applications.* Academic Press, San Diego. [MR **89**k:60020]

[10] W. F. Donoghue (1969). *Distributions and Fourier Transforms.* Academic Press, New York.

[11] W. Feller (1971). *An Introduction to Probability and Its Applications Vol. 2, Second Edition.* Wiley, New York. [MR **42**(1971)5292; Zbl **158**(1969)349]

[12] O. Frostmann (1935). *Potentiel d'équilibre et capacité des ensemble.* Thèse, Lund.

[13] P. Gänssler und W. Stute (1977). *Wahrscheinlichkeitstheorie.* Springer-Verlag, Berlin Heidelberg New York. [MR **58**(1979)18632]

[14] E. J. Gumbel (1954). The maxima of the mean largest value and of the range. *Ann. Math. Statist.* **25**, 76-84. [MR **15**(1954)725]

[15] H. O. Hartley und H. A. David (1954). Universal bounds for mean range and extreme observation. *Ann. Math. Statist.* **25**, 85-99. [MR **15**(1954)725]

[16] E. Hille (1962). *Analytic Function Theory Vol. 2.* Blaisdell Publishing Company, Waltham, Massachusetts. [MR **34**(1967)1490]

[17] H. I. Jacobson (1969). The maximum variance of restricted unimodal distributions. *Ann. Math. Statist.* **40**, 1746-1752. [MR **40**(1970)5054]

[18] F. John (1955). *Plane Waves and Spherical Means Applied to Partial Differential Equations.* Interscience, New York. [MR **17**(1956)746]

[19] S. Moriguti (1951). Extremal properties of extreme value distributions. *Ann. Math. Statist.* **22**, 523-536. [MR **13**(1952)570]

[20] G. Polya und G. Szegö (1931). Über den transfiniten Durchmesser (Kapazitätskonstante) von ebenen und räumlichen Punktmengen. *J. Reine Angew. Math.* **165**, 4-49.

[21] R. L. Plackett (1947). Limits of the ratio of mean range to standard deviation. *Biometrika* **34**, 120–122. [MR **8**(1947)395]

[22] W. Rudin (1973). *Functional Analysis.* McGraw-Hill, New York. [MR **51**(1976)1315]

[23] J. S. Rustagi (1976). *Variational Methods in Statistics.* Academic Press, New York. [MR **53**(1977)6388]

[24] I. J. Schoenberg (1938). Metric spaces and completely monoton functions. *Ann. Math.* **39**, 811-841. Nachgedruckt in: I. J. Schoenberg (1988). *Selected Papers. Vol. 1.* Birkhäuser, Boston, Seiten 115-145.

[25] L. Schwartz (1966). *Théorie des Distributions.* Hermann, Paris. [MR **35**(1968)730]

[26] J. W. Seaman, D. M. Young und D. W. Turner (1987). On the variance of certain bounded random variables. *The Mathematical Scientist* **12**, 109-116.

[27] N. Sugiura (1962). On the orthogonal inverse expansion with an application to the moments of order statistics. *Osaka Math. J.* **14**, 253-263. [MR **26**(1963)859]

[28] N. Sugiura (1964). The bivariate orthogonal inverse expansion and the moments of order statistics. *Osaka J. Math.* **1**, 45-59. [MR **29**(1965)6589]

[29] G. J. Székely und T. F. Móri (1985). An extremal property of rectangular distributions. *Statist. Probab. Lett.* **3**, 107-109. [MR **87**d:62026]

[30] G. R. Terrell (1983). A characterization of rectangular distributions. *Ann. Probab.* **11**, 823-826. [MR **84**i:62020]

[31] V. M. Tichomirov (1982). *Grundprinzipien der Theorie der Extremalaufgaben.* Teubner, Leipzig. [MR **83**m:49002]

[32] M. Tsuji (1959). *Potential Theory in Modern Function Theory.* Maruzen, Tokyo. [MR **22**(1961)5712]

[33] D. V. Widder (1934). The inversion of the Laplace integral and the related moment problem. *Trans. Amer. Math. Soc.* **36**, 107-200.