Kapitel 6: Konzentrations- und Disparitätsmessung

1. Konzentrationsbegriff und Konzentrationsmessung 141
a) Absolute und relative, statische und dynamische Konzentration ... 141
b) Konstruktion von Konzentrations- und Disparitätsmaßen 142
2. Eigenschaften von Konzentrations- und Disparitätsmaßen 146
a) Axiome 146
b) Erläuterungen zu den Axiomen 148
3. Messung der (absoluten) Konzentration 150
a) Konzentrationskurve und Rosenbluth-Index 150
b) Herfindahl-Index 157
c) Exponentialindex und Entropie 160
4. Messung der relativen Konzentration (Disparität) 163
a) Lorenzkurve und Gini-Koeffizient 163
aa) Berechnung bei Einzelbeobachtungen 163
bb) Berechnung bei gruppierten und klassierten Daten 170
b) Der Variationskoeffizient als Disparitätsmaß 175
c) Disparität und verwandte Konzepte 176
5. Dominanzmaße: Entdeckung oligopolistischer Strukturen 178
6. Zur Vertiefung des Verständnisses der Lorenzkurve 181
a) Momentverteilung und Häufigkeitsverteilung 181
b) Stetige Lorenzkurve 184
c) Schutz-Koeffizient (Maximaler Nivellierungssatz) 186
d) Gleichmäßig normierte Maße 188

1. Konzentrationsbegriff und Konzentrationsmessung

a) Absolute und relative, statische und dynamische Konzentration

Konzentration im wirtschaftlichen Sinne kann zweierlei bedeuten:

1. eine Ballung von Verfügungsmacht, Marktanteilen o.ä. auf wenige Einheiten (z.B. Marktbeherrschung) und
2. die Existenz erheblicher Größenunterschiede zwischen den Einheiten ("Ungleichheit").

Einmal wird auf die absolut geringe Anzahl der wirtschaftlichen Einheiten abgestellt (Anzahlaspekt der Konzentration), im anderen Fall auf die Ungleichheit der auf die Einheiten entfallenden Anteile am gesamten Merkmalsbetrag (Disparitätsaspekt der Konzentration). Die statistischen Maße der absoluten Konzentration (Konzentration im engeren Sinne) berücksichtigen beide Aspekte, die der Disparität (oder: relativen Konzentration) nur den zweiten.

Beispiele

Eine Aussage im Sinne der relativen Konzentration ist zum Beispiel:

- 1,7% der Bevölkerung haben mehr als 70% des Produktivvermögens (sowohl die Merkmalsträger [Bevölkerung] als auch die auf sie entfallenden Merkmalbeträge [Anteile am Produktivvermögen] sind relativiert [Prozentangaben]).

Eine Aussage im Sinne der absoluten Konzentration ist dagegen

- auf einem bestimmten Markt haben nur 3 Unternehmen (die drei größten Unternehmen) zusammen einen Marktanteil von 90% (die Merkmalsträger sind in absoluter Zahl angegeben, es kommt auf die absolut geringe Anzahl an).

Im Falle der sog. egalitären Verteilung (Verteilung der Merkmalssumme auf n Einheiten zu gleichen Anteilen 1/n) wird der Unterschied besonders deutlich. Die relative Konzentration (oder: Disparität) ist jeweils Null, die absolute Konzentration dagegen umso größer, je kleiner n ist.

Neben der soweit dargestellten statischen Betrachtung eines Konzentrationszustands wird auch häufig eine **dynamische Konzentrationsmessung** gefordert, in der die Veränderung der Verteilung (des Konzentrationsmerkmals) im Zeitablauf geeignet beschrieben wird. Die Darstellung dieser Art von Mobilität macht Verlaufsanalysen erforderlich und führt zu sehr komplizierten Modellen der **Unternehmensdemographie**, die Zu- und Abgänge, Wachstum, Diversifikation, Aktivitätsverlagerung usw. berücksichtigen müßten (auch unter Berücksichtigung von Methoden des Kapitels 12) und auch aus Gründen des Datenschutzes kaum für empirische Untersuchungen angewandt werden können, so dass z.Zt. noch statische oder komparativ-statische Betrachtungen dominieren.

b) Konstruktion von Konzentrations- und Disparitätsmaßen

1. Notwendigkeit von Gedankenexperimenten

In der wirtschaftlichen Realität sind absolute und relative Konzentration nicht zwei streng unterschiedene Erscheinungen, sondern zwei in der Regel gemeinsam auftretende Aspekte <u>eines</u> Vorgangs. Neugründungen, Fusionen, Auflösungen oder -teilungen, ungleiches Größenwachstum usw. berühren meist beide Arten von Konzentration und damit auch beide Arten von statistischen Maßzahlen gleichzeitig, wenngleich häufig in unterschiedlicher Weise. Dem steht jedoch nicht entgegen, dass man modellmäßige Vorgänge konstruieren kann, die sich isoliert auf einen der beiden Aspekte der Konzentration auswirken. Das geschieht bei der Entwicklung einer Axiomatik. In Form von sog. "Proben" werden die Auswirkungen solcher Vorgänge auf die Maßzahlen der Konzentration und Disparität untersucht.

Auch die Unterscheidung zwischen einer statischen und dynamischen Betrachtung ist nur eine gedankliche Abstraktion. In der Realität entwickeln sich Konzentrationszustände aus Konzentrationsprozessen und es ist fraglich, ob eine Momentaufnahme den Sachverhalt überhaupt hinreichend beschreiben kann. Wegen der Komplexität dieser Prozesse ist es auch hier notwendig, in Gedankenexperimenten isoliert den Effekt einfacher Veränderungen an den Daten "durchzuspielen".

2. Daten, im folgenden verwendete Symbole

Grundlage konzentrationsstatistischer Betrachtung sind die im folgenden definierten Größen n, h_i, q_i, bzw. c_i:

Def. 6.1: Anteile, Merkmalsanteile

Es sollen die folgenden Symbole verabredet werden:

1. Die Anzahl der **Merkmalsträger** ist n, ihre Anteile an der Gesamtheit der Merkmalsträger $h_i = 1/n$ (bei Einzelbeobachtungen) bzw. $h_i = n_i /n$ bei gruppierten oder klassierten Daten

2. Die Merkmalsanteile, d.h. die Anteile an dem **Merkmalsbetrag** (Merkmalssumme der zu verteilenden Größe) lauten q_i. Wenn besonders hervorgehoben werden soll, dass die Merkmalsanteile sich auf in einer bestimmten Weise geordnete Merkmalsträger beziehen kann statt q_i auch das Symbol c_i verwendet werden (vgl. Gl. 6.2).

Die folgenden Betrachtungen beziehen sich auf Einzelwerte. Unterschiede für den Fall klassierter Daten werden an späterer Stelle behandelt.

3. Konzentrationsmaße

Aufgabe der Deskriptiven Statistik ist es auch hier, Zustände

a) graphisch, bzw. tabellarisch
b) durch eine zusammenfassende Maßzahl (ein "Konzentrationsmaß")
 zu beschreiben.

Konzentrationsmaße, in deren Berechnung alle Anteile q_i eingehen, heißen summarische, solche dagegen, die nur einen Teil der Anteile q_i berücksichtigen heißen diskrete Konzentrationsmaße (vgl. auch Abschn. 6).

4. Axiomatik

Die Entwicklung einer Axiomatik erfolgt, wie gesagt, durch Gedankenexperimente in Form von sog. "Proben". So beschreibt z.B. die "Ergänzungsprobe" (Hinzufügen von Einheiten, deren Merkmalsbeträge jeweils Null sind) eine isolierte fiktive Vergrößerung der Disparität. Entsprechend untersucht die "Proportionalitätsprobe" (Aufteilung jeder Einheit in k genau gleich große Einheiten) die Auswirkungen des reinen Anzahleffekts auf die statistischen Maßzahlen. Maße der Konzentration und Disparität sollen auf diese beiden Proben unterschiedlich reagieren. Sie sind also auch durch ihre im Rahmen einer Axiomatik (zu der die beiden genannten "Proben" gehören) geforderten Eigenschaften unterschieden.

Übersicht 6.1

Darstellung	(absolute) Konzentration	(relative Konzentration) Disparität
a) graphisch	Konzentrationskurve	Lorenzkurve
b) Maße - summarisch - diskret	 K_R(Rosenbluth)-Index*) Herfindahl-Index K_H concentration ratios	 D_G Gini-Koeffizient **) Variationskoeffizient Schutz-Koeffizient***)

*) aus der Konzentrationskurve entwickelt,

**) aus der Lorenzkurve entwickelt.

***) auch maximaler Nivellierungssatz von Lindahl genannt

5. Wirtschaftsstatistische Fragen

Bei der Durchführung einer Konzentrationsmessung im wirtschaftlichen Bereich sind eine Reihe wirtschaftsstatistischer Fragen zu klären: Wahl eines geeigneten Konzentrationsmerkmals (z.B. Umsatz oder Beschäftigtenzahl als Maß der Unternehmensgröße) oder z.B. die "Abgrenzung des relevanten Marktes" (d.h. der zugrundegelegten Masse)[1].

1 Ausführlicher zu diesen Aspekten vgl. von der Lippe, P.: "Wirtschaftsstatistik" UTB Bd. 209, 4. Aufl., S. 230f.

6. Anforderungen an ein Konzentrationsmerkmal

Ein Konzentrationsmerkmal muss sein

- nichtnegativ
- extensiv und
- mindestens intervallskaliert.

7. Extremfälle der Konzentration

Die Entwicklung eines Axiomensystems wird durch die Existenz von zwei klar definierten Extremfällen der Konzentration erleichtert. Es ist insbesondere möglich, so Konzentrations-, bzw. Disparitätsmaße zu normieren. Es sind die folgenden zwei Extremsituationen zu unterscheiden:

a. **egalitäre Verteilung** (Einpunktverteilung oder extreme Gleichverteilung), die Situation der minimalen Disparität: jeder der n Merkmalsträger hat den gleichen Merkmalsbetrag x und damit auch den gleichen Merkmalsanteil

$$q_i = \frac{1}{n} \qquad i=1,2,...,n.$$

wobei diese Anteile auch in einem Spaltenvektor **q** zusammengefaßt werden.
Bei einer Disparitäts von Null wird auch mißverständlich von "Gleichverteilung"[2] geprochen.

b. **vollkommene Ungleichheit** (extreme Ungleichverteilung, maximale Konzentration, Zweipunktverteilung): ein Merkmalsträger vereinigt die gesamte Merkmalssumme auf sich, sein Anteil q ist somit 1 und die Anteile q der übrigen n-1 Merkmalsträger sind demnach alle jeweils Null.

Def. 6.2: Disparitäts- und Gleichheitsmaß

Ist D ein Disparitätsmaß, so ist G=1-D ein Gleichheitsmaß

Die im nächsten Abschnitt besprochenen Axiome für Konzentrations- und Disparitätsmaße lassen sich oft leichter formulieren, wenn auf Gleichheit anstatt Disparität abgestellt wird.

---

[2] Der beschriebene Fall der egalitären Verteilung heißt in der Statistik auch **Einpunkt-verteilung** (alle Einheiten haben die gleiche [und damit einzige] Merkmals-ausprägung, deren prozentuale Häufigkeit 100% ist). **Gleichverteilung** heißt dagegen: jede Merkmalsausprägung kommt gleich häufig vor (z.B. die Verteilung der Augenzahl beim Würfeln [eine Wahrscheinlichkeits- nicht Häufigkeitsverteilung]).

2. Eigenschaften von Konzentrations- und Disparitätsmaßen

a) Axiome

Es sind Eigenschaften zu unterscheiden

a) die sowohl Konzentrations- als auch Disparitätsmaße haben sollten und
b) solche, bei denen sich Konzentrations- und Disparitätsmaße unterscheiden.

Ist speziell ein Konzentrationsmaß gemeint, so wird es hier K genannt, ein Disparitätsmaß im allgemeinen Sinne heißt entsprechend D und bei einer Aussage, die sich sowohl auf Konzentrations- als auch auf Disparitätsmaße bezieht, soll das entsprechende Maß C genannt werden.

zu a) Eigenschaften von Konzentrations- und Disparitätsmaßen

K1: Unabhängigkeit von der Maßeinheit des Konzentrationsmerkmals

Ein Konzentrations- oder Dispariatätsmaß C soll invariant sein bei proportionaler Transformation: Ist $y_i = bx_i$ $(b > 0)$, so ist $C(y) = C(x)$.

K2: Verschiebungsprobe (Transfer)

Wird ein Betrag d mit $0 < d < h/2$ transferiert von einem Merkmalsträger i (mit dem Merkmalsbetrag $x_{(i)}$) zum Merkmalsträger j mit $x_{(j)} = x_{(i)} - h$ also $x_{(j)} < x_{(i)}$, so soll C abnehmen (regressiver [egalisierender, negativer, d.h. die Konzentration verringernder] Transfer).

Die Umkehrung sollte entsprechend bei einem progressiven [positiven] also die Konzentration (und damit auch das Konzentrationsmaß) erhöhenden Transfer ("von arm zu reich") gelten, d.h. das Konzentrationsmaß sollte dann steigen.

K3: (Verschiebung, Niveauänderung)

Sei $y_i = a + x_i$, dann ist bei egalitärer Verteilung des Merkmals X die Konzentration des Merkmals Y gleich, also

$$C(y) = C(x)$$

und in den sonstigen Fällen soll gelten

$$C(y) = \begin{cases} < C(x) & \text{wenn } a > 0 \quad \text{(abnehmende Konzentration)} \\ > C(x) & \text{wenn } a < 0 \quad \text{(zunehmende Konzentration)} \end{cases}$$

zu b) Eigenschaften, bei denen sich Konzentrations- und Disparitätsmaße unterscheiden

Im Unterschied zu den Axiomen K1 bis K3 wird bei den Axiomen K4 und K5 postuliert, dass Konzentrations- und Disparitätsmaße auf die in den "Proben" unterstellten Vorgänge unterschiedlich reagieren. Die Proportionalitätsprobe (Axiom K4) beschreibt den reinen Anzahleffekt, weshalb Disparitätsmaße hierauf nicht reagieren sollen. Die Forderung K5 ist damit motiviert, dass man gedanklich eine immer größer werdende Ungleichheit durch Hinzufügen von "Nulleinheiten" entstehen lassen kann.

K4: (Proportionalitätsprobe):

Ersetzt man jeden einzelnen Merkmalsträger i mit dem Anteil q_i am Merkmalsbetrag durch $k > 1$ gleich große Merkmalsträger mit den Anteilen q_i/k, so soll für das neue Disparitätsmaß D* gelten:

$$\boxed{D^* = D(\mathbf{q}) \qquad \text{(Disparität bleibt unverändert)}}$$

und für das "neue" Konzentrationsmaß K* im Vergleich zum "alten"

$$\boxed{K^* = \frac{K}{k} = \frac{1}{k}K(\mathbf{q}) \qquad \text{(Fall der Dekonzentration).}}$$

Entsprechend soll im "umgekehrten" Fall einer Fusion von k gleich großen Einheiten zu einer Einheit gelten $D^*=D$ und $K^*=kK$ (Fusion).

K5: (Ergänzungsprobe, Nullergänzung):

Fügt man einer Verteilung m Einheiten, deren Merkmalsbeträge jeweils Null sind ("Nullträger") hinzu, so soll gelten

$$\boxed{K^* = K \text{ und } D^* > D.}$$

K6: Wertebereiche

Die folgenden Wertebereiche sollen gelten: für

Konzentrationsmaße	$\frac{1}{n} \leq K \leq 1$
Disparitätsmaße	$0 \leq D \leq 1 - \frac{1}{n}$.

b) Erläuterungen zu den Axiomen

1. Die Axiome K1 bis K3 werden sowohl für Konzentrationsmaße, als auch für Disparitätsmaße gefordert. Das Axiom K1 besagt, dass ein Konzentrations- und ein Disparitätsmaß unabhängig sein soll von der Maßeinheit der zu verteilenden Größe (des Konzentrationsmerkmals). Die "Ungleichheit" der Einkommensverteilung wird nicht davon berührt, ob das Einkommen in DM oder in Pfennigen gerechnet wird. Sie ist in beiden Fällen gleich groß.

   Diese Invarianz gegenüber proportionalen Transformationen (auch "Bresciani-Turoni-Bedingung" genannt) wird dadurch sichergestellt, dass zur Berechnung von Konzentrations- und Disparitätsmaßen nur Merkmals**anteile**, nicht absolute Merkmals**beträge** benutzt werden.

2. Das Axiom K2 (auch "Transfer", "Pigou-Dalton-Effekt" oder [wegen K3] mißverständlich "Verschiebungsprobe" genannt) wird nicht selten nur für Disparitätsmaße gefordert. Bei einem Transfer "von reich zu arm" soll die Disparität abnehmen (negativer -, regressiver - oder egalisierender Transfer). Umgekehrt soll die Disparität zunehmen, wenn d von j auf i transferiert wird (positiver - oder progressiver Transfer).

   Transfers sind stets Umverteilungen bei gleichbleibendem gesamtem Merkmalsbetrag und wegen $d < h/2$ auch bei gleichbleibender Rangfolge der Merkmalsträger (unter denen der Transfer stattfindet). Die Bedingung $d < h/2$ wird deshalb eingeführt ("Überholverbot").
   Es ist auch üblich, eine Reaktion auf einen Transfer für Konzentrations-, nicht nur für Dispariatätsmaße zu fordern, oder aber zu unterscheiden:
   bei einem regressiven Transfer soll K abnehmen, D dagegen strikt abnehmen.
   Gelegentlich wird der Transfer auch "bewertet" in dem Sinne, dass sich ein positiver Transfer auf ein Konzentrations- oder Disparitätsmaß stärker erhöhend auswirken soll, als sich ein betragsmäßig gleich großer negativer Transfer verringernd auswirkt ("**starke Verschiebungsprobe**").

3. In der Literatur wird K3 häufig nur für Disparitätsmaße gefordert. Anders als die in K1 postulierte gleiche **relative** Änderung (z.B. Erhöhung aller Einkommen um 20%, so dass $b = 1{,}2$ ist) soll sich eine gleiche **absolute** Änderung der Merkmalsbeträge durchaus auf die Disparität auswirken. Nach Axiom K3 bedeutet eine für alle Merkmalsträger gleich große Zunahme ($a>0$) der Merkmalsbeträge eine Verringerung der Disparität und entsprechend eine Abnahme ($a<0$) eine Vergrößerung der Disparität.

4. Der mit der Proportionalitätsprobe beschriebene Anzahleffekt soll sich auf ein Disparitätsmaß nicht auswirken, die Konzentration soll sich jedoch proportional zu einer Verringerung (Fusion) der Anzahl der Merkmalsträger vergrößern, bzw. zu einer Vergrößerung (Dekonzentration) der Anzahl der Merkmalsträger verringern.

5. Hinter der Ergänzungsprobe (Axiom K5) steht - wie gesagt - die Vorstellung, dass man sich eine "Ungleichverteilung" durch Hinzufügen von Nullträgern aus der Gleichverteilung entstanden denken kann. In gleicher Weise wird eine bestehende "Ungleichheit" durch Hinzufügen von Nullträgern vergrößert.
   Der Vorgang der Nullergänzung soll dagegen ein Maß der absoluten Konzentration K unverändert lassen, weil die Anzahl derjenigen Merkmalsträger, die sich bisher die Merkmalssumme aufteilten und deren Anteile $q_i>0$ sind, gleich bleiben.
   Mit dem Axiom K5 ist noch nicht gesagt, um wieviel sich D bei Hinzufügen (Wegnehmen) von Nulleinheiten vergrößert (verringert). Eine solche Aussage wird mit der folgenden spezielleren Fassung des Axioms getroffen, die wir Axiom K5a nennen wollen

   Axiom K5a:

   Manchmal wird auch gefordert, dass sich ein Disparitätsmaß in einem bestimmten Ausmaß vergrößert. Fügt man einer Verteilung m=(k-1)n Nullelemente hinzu, so dass der Vektor der Merkmalsanteile $\mathbf{q} = [q_1...q_n]$ in den Vektor $\mathbf{q}_e = [q_1...q_n\ 0...0]$ mit kn Elementen übergeht, so sollte für ein Disparitätsmaß D und für ein Gleichheitsmaß G = 1 - D gelten

$$(6.1) \qquad 1 - D(\mathbf{q}_e) = \frac{1 - D(\mathbf{q})}{k}$$

$$(6.1a) \qquad G(\mathbf{q}_e) = \frac{G(\mathbf{q})}{k}\ .$$

   Das bedeutet z.B., dass durch eine Verdoppelung (k=2) der Anzahl der Merkmalsträger das Gleichheitsmaß halbiert wird. Es handelt sich aber nicht einfach um eine Verdoppelung der ursprünglichen Einheiten (was ein Anzahleffekt im Sinne von K4 wäre), sondern um ein Hinzufügen von n "Nullträgern" zu n Einheiten, deren Merkmalsbeträge größer oder gleich Null sind.

6. Es wird nicht nur gefordert, dass die oben genannte Grenzen gelten, sondern auch, dass sie in genau definierten Situationen angenommen werden. Danach gilt bei den bereits dargestellten extremen Zuständen:

maximale Konzentration	$K_{max} = 1$ und $D_{max} = 1 - \frac{1}{n}$
egalitäre Verteilung	$K_{min} = \frac{1}{n}$ und $D_{min} = 0$.

Man beachte, dass die *minimale* (nicht aber die maximale) Konzentration und die *maximale* (nicht aber die minimale) Disparität von n abhängig ist. Deshalb kann man auch nicht einfach fordern: $0 \leq D,K \leq 1$, auch nicht wenn man bedenkt, dass 1/n "gegen Null strebt", zumal n bei Konzentrationsuntersuchungen oft nicht sehr groß ist.

- *Konzentration:* War ein Markt bisher von 4 gleich großen Unternehmen beherrscht, so würde ein Übergang zu 3 gleich großen Unternehmen die Disparität nicht berühren (da stets alle Unternehmen gleich groß sind, bleibt D=0), wohl aber die Konzentration von 1/4 auf 1/3 erhöhen (die wegen D=0 minimale Konzentration ist also von n abhängig: je kleiner n desto größer K_{min}), denn im Sinne der absoluten Konzentration ist die Konzentration auf 3 statt 4 Unternehmen natürlich eine Steigerung

- *Disparität:* Sie kann nur maximal sein, wenn einer alles und der Rest (n - 1 Einheiten) nichts bekommt. Es ist klar, dass diese Ungleichheit umso größer ist je größer "der Rest" ist. Es ist deshalb sinnvoll für den Fall, dass eine Einheit die gesamte Merkmalssumme auf sich vereint danach zu differenzieren, wie groß n ist (die Obergrenze von D ist deshalb abhängig von n).

3. Messung der (absoluten) Konzentration

a) Konzentrationskurve und Rosenbluth-Index

Ordnung nach abnehmender Größe

Die Daten liegen in Form von Einzelbeobachtungen $x_1, x_2, \ldots, x_n$ vor. Da Konzentration bedeutet, dass eine geringe Anzahl von Merkmalsträgern einen großen Anteil an der Merkmalssumme auf sich vereinigt, werden die Merkmalsbeträge nach abnehmender Größe geordnet (d.h. die Merkmalsträger werden nach abnehmenden Merkmalsbeträgen geordnet), so dass

$x^{(1)} \geq x^{(2)} \geq \ldots \geq x^{(n)}$ (abnehmende Merkmalsbeträge).

Bei der Disparitätsmessung ist es dagegen üblich, von einer Ordnung nach zunehmender Größe auszugehen, so dass gilt

$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ (zunehmende Merkmalsbeträge).

Konzentrationsraten, Konzentrationskurve und Rosenbluth-Index

Wir sind jetzt in der Lage, Konzentrationsraten, deren graphische Darstellung, die Konzentrationskurve und den hieraus abgeleiteten Rosenbluth-Index als Konzentrationsmaß (Maß der absoluten Konzentration) zu definieren.

Def. 6.3 : Konzentrationsraten, -kurve, Rosenbluth-Index

a) Wird der Merkmalsanteil des i-ten Merkmalsträgers

$$\boxed{(6.2) \quad c_i = \frac{x^{(i)}}{\Sigma x^{(i)}} = \frac{x^{(i)}}{\Sigma x_j}}$$

genannt (oder äquivalent q_i; vgl. Def. 6.1) dann ist

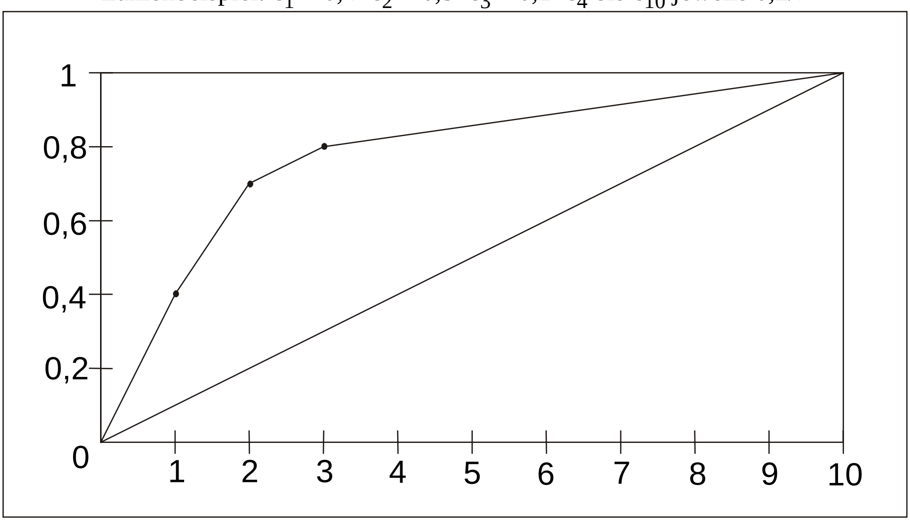
$$\boxed{(6.3) \quad C_i = \sum_{j=1}^{i} c_j \qquad C_0 = 0 < C_1 <...< C_n = 1}$$

der kumulierte Anteil der i größten Merkmalsträger und heißt **Konzentrationsrate** (concentration ratio).

b) Zeichnet man die geordneten Paare (i, C_i) in ein kartesisches Koordinatensystem ein und verbindet man die Punkte mit den Koordinaten (0,0), $(1,C_1)$,..., (n,1), so heißt der daraus resultierende Polygonzug **Konzentrationskurve** (vgl. Abb. 6.1).

Abb. 6.1: Konzentrationskurve

Zahlenbeispiel: $c_1 = 0{,}4$ $c_2 = 0{,}3$ $c_3 = 0{,}1$ c_4 bis c_{10} jeweils 0,2/7



c) Der **Rosenbluth-Index** (Konzentrationskoeffizient von Rosenbluth) ist wie folgt definiert:

$$(6.4) \quad K_R = \frac{1}{2\Sigma i c_i - 1} \qquad \frac{1}{n} \leq K_R \leq 1 \qquad (i=1,2,...,n).$$

Bemerkungen zur Def. 6.3:

1. Zwischen den Ordnungsstatistiken $x^{(i)}$ und den bisher betrachteten Ordnungsstatistiken $x_{(i)}$ gilt die Beziehung
   $x^{(i)} = x_{(n-i+1)}$ $\quad i = 1,2,...,n.$

2. Die Konzentrationsrate C_r ist der Anteil, den die r größten Merkmalsträger auf sich vereinigen. Wenn z.B. das Merkmal Umsatz betrachtet wird ist C_5 der Umsatzanteil der fünf größten Unternehmen. Konzentrationsraten C_i sind diskrete Konzentrationsmaße, weil sie nur einen Teil der Merkmalsanteile c_i berücksichtigen. Sinnvoll kann nur $r < n$ gewählt werden, weil

   - (absolute) Konzentration bedeutet, dass eine geringere Anzahl r von Merkmalsträgern einen kumulierten Merkmalsanteil C_r hat.
   - für $r = n$ notwendig gilt $C_r = 1$, was keine Information liefert.

   Die Wahl von r ist willkürlich. Die Konzentrationsrate ist darüberhinaus zwar einfach und anschaulich, sie nutzt aber die in der Konzentrationskurve steckende Information nicht voll aus. Bei Untersuchungen über die Konzentration von Unternehmen in bezug auf Umsätze; Marktanteile usw. ist es üblich (z.B. in den Untersuchungen der Monopolkommission) die Konzentrationsraten C_3, C_6, C_{10} zu betrachten.

3. Die Konzentrationskurve ist eine graphische Darstellung der Konzentrationsraten. Aus $x^{(i)} \geq x^{(i+1)}$ folgt

   $$(6.5) \quad c_i \geq c_{i+1} \ \text{ und } \ C_i < C_{i+1} \ .$$

   Die Steigung der Konzentrationskurve beträgt im Intervall von i-1 bis i genau c_i und kann wegen (6.5) nicht zunehmen. Im Fall der egalitären Verteilung, bei dem für alle i gilt $c_i = 1/n$ ist die Konzentrationskurve eine die Punkte (0,0),...,(n,1) verbindende Gerade $C_i = i/n$ mit der Steigung 1/n. In allen anderen Fällen verläuft sie als Polygonzug oberhalb dieser Diagonalen konkav, so dass gilt

(6.6) $C_i \geq \frac{i}{n}$	$(i = 1,2,...,n)$.

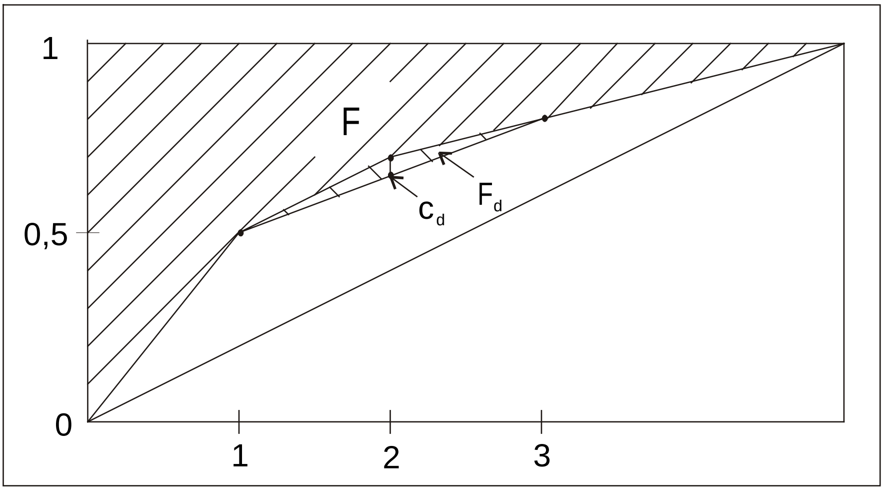
Im Falle der maximalen Konzentration ist $C_1 = 1$ und $C_i = 0$ (für $1 < i < n$).

4. Der Rosenbluth-Index stellt ein Maß für die Wölbung der Konzentrationskurve dar. Man kann zeigen, dass K_R invers von der Fläche F (schraffierte Fläche in Abb. 6.2 bis 6.4) abhängt:

(6.7)	$F = \Sigma ic_i - ½$	$(½ \leq F \leq n/2)$	und
(6.7a)	$K_R = \frac{1}{2F}$		

Aus Gl. 6.7 ist ersichtlich, dass die Reihenfolge der Einheiten nicht beliebig ist. Bei der Konzentrationsmessung werden, wie gesagt, die Merkmalsträger stets nach abnehmenden Merkmalsanteilen c_i geordnet.

Abb. 6.2: Konzentrationskurve und Fläche F (vor und nach negativem Transfer zwischen Einheit 2 und 3)



5. Durch die proportionale Transformation $y_i = b \cdot x_i$ verändern sich die Merkmals<u>anteile</u> c_i nicht, so dass K_R das Axiom K1 erfüllt.

6. Um zu zeigen, dass K_R das Axiom K2 erfüllt, betrachten wir ohne Beschränkung der Allgemeinheit einen negativen (egalisierenden) Transfer des Betrages d und damit des Anteils

$$c_d = \frac{d}{\Sigma x_i} < \frac{1}{2}(c_i\text{-}c_j)$$

von i zu j = i+1. Dadurch wird C_i um c_d geringer und C_{i+1} bleibt gleich. Die Konzentrationskurve verläuft dann zwischen den Punkten (i-1 , C_{i-1}) und (i+1,C_{i+1}) flacher (vgl. Abb. 6.2). Dadurch vergrößert sich die ursprüngliche Fläche F um die Fläche F_d. Mit geometrischen Überlegungen läßt sich zeigen, dass die Fläche F_d dem Merkmalsanteil c_d entspricht. Offensichtlich gilt dies auch bei einem Transfer von i auf j (bei j > i+1). Wegen der in Gl. 6.7a dargestellten inversen Beziehung zwischen F und K_R nimmt K_R ab. Analog nimmt K_R bei positivem Transfer zu.

7. Für $y_i = a+x_i$ erhält man die "neuen" Konzentrationsraten (von y statt von x)

$$C_i,^* = \frac{a(i/n) + C_i\,\overline{x}}{a + \overline{x}} = \frac{a}{a+\overline{x}}\,\frac{i}{n} + \frac{\overline{x}}{a+\overline{x}}\,C_i \qquad (6.8)$$

$C_i,^*$ ist also ein gewogenes Mittel aus dem bisherigen Wert C_i und dem Wert i/n bei egalitärer Verteilung. Ist die Ausgangsverteilung die egalitäre Verteilung, so bleibt diese erhalten. In allen anderen Fällen gilt:

- Für a > 0 ist $C_i^* < C_i$ für i < n, so dass die neue Konzentrationskurve unterhalb der bisherigen verläuft, was bedeutet, dass K_R abnimmt.
- Für a < 0 ist entsprechend $C_i^* > C_i$, wenn i < n ist, so dass K_R zunimmt (das Gewicht $a/(a+\overline{x})$ ist negativ).

Somit erfüllt K_R auch das Axiom K3.

8. Auch Axiom K4 ist, wie leicht zu sehen ist, erfüllt. Eine Dekonzentration mit k = 2 verändert die Konzentrationskurve wie in Abb. 6.3 gezeigt. Den Konzentrationsraten C_i sind jetzt die Abszissenwerte 2i zugeordnet, so dass die neue Fläche F^*=2F ist. Daraus folgt für den neuen Rosenbluth-Index $K_{R*} = (2F^*)^{-1} = ½K_R$. Bei einer Fusion von je zwei Einheiten mit gleichem Merkmalsanteil gilt $K_{R*} = 2K_R$, was sich ebenfalls aus Abb. 6.3 ergibt, wenn man diese von rechts nach links liest.

9. Axiom K5 ist offensichtlich ebenfalls erfüllt, da dies nur darauf hinausläuft, dass das Rechteck über die Punkte (n,0) und (n,1) hinaus

verlängert wird (Abb. 6.4). Es ist klar, dass sich dadurch der Rosenbluth-Index nicht verändert.

Abb. 6.3: Wirkung einer Dekonzentration (Axiom K4)

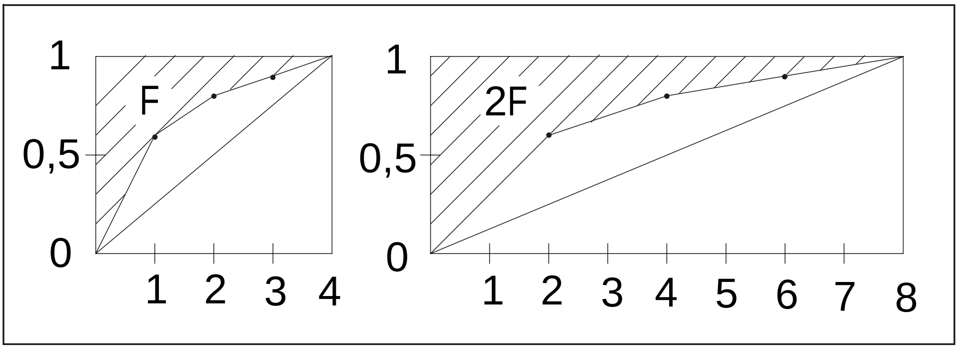
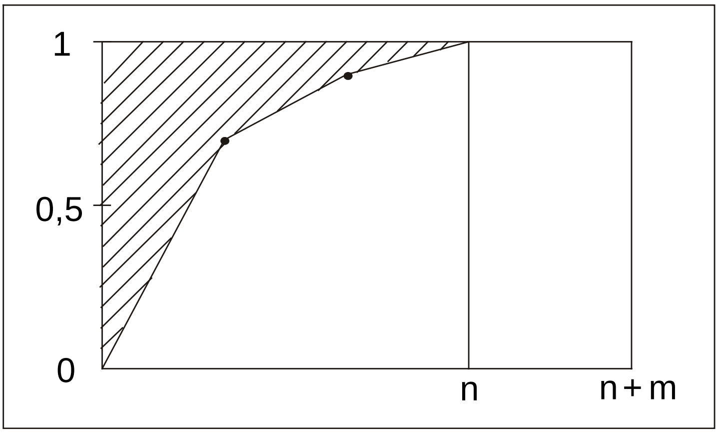


Abb. 6.4: Konzentrationskurve bei Nullergänzung von m Einheiten



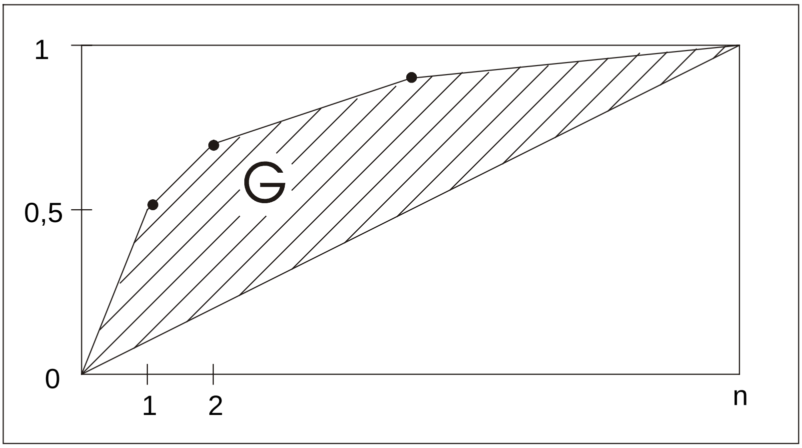
10. Die maximale Konzentration ist erreicht, wenn der größte Merkmalsträger die gesamte Merkmalssumme auf sich vereinigt. Man erhält dann für die Fläche F den Wert ½ und somit $K_R = 1$. Entsprechend gilt bei egalitärer Verteilung (c_i=1/n) F = (1+2+...+n)/n - ½ also 2F = n und somit $K_R = 1/n$.

11. Der Rosenbluth-Index hängt, wie gesagt, invers von der Größe der Fläche F ab. Dies legt den Gedanken nahe, ein Maß aus der Fläche G (vgl. Abb. 6.5) zwischen der Diagonalen und der Konzentrationskurve

$$G = n/2 - F \qquad 0 \leq G \leq (n-1)/2$$

zu konstruieren, analog zur Herleitung des Gini-Koeffizienten aus der Lorenzkurve (vgl. Abschn. 4).

Abb. 6.5: Fläche G als Konzentrationsmaß



Man kann aber leicht sehen, dass eine mit Gl. 2.3 auf das Intervall von 1/n bis 1 normierte Fläche G, also das Flächenmaß

$$G_n = \frac{1+2G}{n} \qquad \frac{1}{n} \leq G_n \leq 1$$

als Maß der absoluten Konzentration unbrauchbar ist, da es die Axiome K4 und K5 nicht erfüllt.

Beispiel 6.1:

An einem Markt bieten fünf Unternehmen an. Ihre Marktanteile sind:

1) $c_1 = 0{,}6$; $c_2 = 0{,}2$; $c_3 = 0{,}1$; $c_4 = 0{,}06$ und $c_5 = 0{,}04$. Man zeichne die Konzentrationskurve und berechne die Fläche F und K_R.

2) Der Marktanteil $c_d = 0{,}04$ wird von Unternehmen 2 auf Unternehmen 3 "übertragen".
Man zeichne wieder die Konzentrationskurve und berechne F und K_R.

3) Angenommen, jedes Unternehmen werde (ausgehend von Situation 1) in zwei gleich große Unternehmen "dekonzentriert". Man zeichne wieder die Konzentrationskurve und berechne F und K_R.

Lösung 6.1:

1) Ordinatenwerte der Konzentrationskurve: $C_1 = 0{,}6$; $C_2 = 0{,}8$; $C_3 = 0{,}9$; $C_4 = 0{,}96$ und $C_5 = 1$. Für die Fläche erhält man $F_1 = 1{,}74 - ½ = 1{,}24$ und für den Rosenbluth-Index $K_{R1} = 0{,}40323$.

2) Nach dem Transfer gilt $C_2 = 0{,}76$. Die anderen Werte bleiben unverändert. Ferner ist $F_2 = 1{,}28$ und $K_{R2} = 0{,}390625$, die Konzentration nimmt also ab (wie im Axiom K2 gefordert).

3) Es gilt jetzt: $C_1 = 0{,}3$; $C_2 = 0{,}6$; $C_3 = 0{,}7;...;C_8 = 0{,}96$, $C_9 = 0{,}98$ und $C_{10} = 1$. Schließlich ist $F_3 = 2{,}48 = 2F_1$ und $K_{R3} = 0{,}2016 = ½K_{R1}$.

b) Herfindahl-Index

Ein sehr einfaches und das am meisten verbreitete Maß der absoluten Konzentration ist der Herfindahl-Index, auf dessen Eigenschaften hier kurz eingegangen werden soll.

<u>Def. 6.4: Herfindahl-Index</u>

Die Summe der quadrierten Merkmalsanteile $q_1,q_2,...,q_n$

$$(6.9) \quad K_H = \Sigma\, q_i^2 = \mathbf{q'q} \qquad \frac{1}{n} \leq K_H \leq 1$$

heißt Herfindahl-Index oder Konzentrationsmaß von Herfindahl (Symbol: statt K_H auch einfach H). Man beachte, dass die Daten hier Einzelbeobachtungen sind. Für gruppierte und klassierte Daten ist K_H gem. Gl. 6.10 zu berechnen.

<u>Interpretation und Eigenschaften des Herfindahl-Index:</u>

1. Der Herfindahl-Index ist ein summarisches Maß der (absoluten) Konzentration. Man erkennt auch leicht, dass es auf die Reihenfolge der Merkmalsanteile (Anteilswerte) q_i nicht ankommt und dass man K_H auch als ein mit den Anteilen q_i gewogenes **arithmetisches** Mittel der Anteile q_i auffassen kann. Daraus folgt auch, dass das Axiom K1 erfüllt ist.

2. Der Herfindahl-Index ist auch in der Form
$K_H = \Sigma(x_i / \Sigma x_i)^2 = \Sigma(x_i{,}^2 / n^2\, x{,}^{-2})$ darstellbar. Daraus folgt auch der bekannte Zusammenhang zwischen K_H und dem Variationskoeffizienten V

$$(6.10) \qquad K_H = \frac{V^2 + 1}{n}\,.$$

Der Zähler V^2+1 charakterisiert die Verteilungsungleichheit und der Nenner mißt den Anzahleffekt. Mithin steigt (sinkt) die absolute Konzentration, wenn sich bei

gegebener relativer Streuung die Anzahl n der Einheiten verringert (vergrößert). Es ist auch plausibel von steigender Konzentration im ökonomischen Sinne zu sprechen, wenn sich die Disparität erhöht und die Anzahl der Merkmalsträger kleiner wird. Wie man daran erkennt, hängt die Aussage der absoluten Konzentration nicht allein von der Anzahl n ab, sondern auch von der Unterschiedlichkeit der Größe der Einheiten:
So kann z.B. für n = 2, bei einem Duopol, K_H alle Werte zwischen ½ und 1 annehmen, je nachdem, welchen Marktanteil q (bzw. p = 1 - q) das erste (bzw. zweite) Unternehmen hat. Die quadratische Funktion $K_H = q^2 + p^2 = 1 - 2q(1-q)$ hat an der Stelle p = q = ½ ihr Minimum.

3. Fusionen von Einheiten vergrößern die absolute Konzentration (aber nicht notwendig auch die Disparität), weil sie die Anzahl der Einheiten verringern.

   Die konzentrationserhöhende Wirkung des Fusionierens der ersten m < n Einheiten ist im Falle des Herfindahl-Index leicht zu sehen, da sich K_H vergrößert wegen: $(\Sigma q_j)^2 > \Sigma q_j^2$ (j=1,2,...,m). Eine Fusion von jeweils k gleich großen Einheiten im Sinne der Proportionalitätsprobe (Axiom K4) bedeutet im Falle von k=2 einen Übergang vom Datenvektor 1: $[q_1\ q_1\ q_2\ q_2\ ...q_n\ q_n]$ zum Vektor 2: $[2q_1\ 2q_2\ ...\ 2q_n]$. Wie man leicht sieht, gilt $K_{H1} = 2\Sigma q_i^2$ und $K_{H2} = 4\Sigma q_i^2$, so dass sich K_H, wie im Axiom K4 gefordert, verdoppelt. Das Axiom ist für beliebiges k bei Fusionen und Dekonzentrationen erfüllt.

4. Der Herfindahl-Index erfüllt auch die übrigen Axiome. Ein negativer Transfer des Betrags d von i zu i+1 (mit den Merkmalsbeträgen x_i und $x_{i+1} = x_i$-h) im Sinne des Axioms K3 verringert K_H um den Betrag $2d(h-d)/(\Sigma x_i)^2$. Entsprechend vergrößert sich K_H bei einem gleich großen positiven Transfer. Es gilt also Axiom K2 (schwache Verschiebungsprobe), nicht aber die starke Verschiebungsprobe. Axiom K5 ist ganz offensichtlich erfüllt, da sich K_H durch Nullergänzung nicht verändert. Mit der im Axiom K3 geforderten Transformation $y_i = a + x_i$ verändert sich der quadrierte Variationskoeffizient wie folgt

   $$(6.11) \quad V_Y^2 = \frac{\bar{x}^2 V_X^2}{(a+\bar{x})^2} \text{ und damit } V_Y = \frac{\bar{x}\, V_X}{a+\bar{x}},$$

   so dass gilt $V_Y < V_X$, wenn a < 0 und $V_Y > V_X$, wenn a > 0, so dass der Herfindahl-Index auch das Axiom K3 erfüllt.

5. Offensichtlich gilt $1/n \leq K_H \leq 1$, wobei die Grenzen genau in den für Konzentrationsmaße beschriebenen Extremzuständen angenommen werden. In diesen Fällen ist $K_H = K_R = s$, wobei s die Steigung der Konzentrationskurve ist. Bei egalitärer Verteilung gilt

$$(6.12)\quad K_H = \sum\left(\frac{1}{n}\right)^2 = \frac{1}{n} \quad .$$

Hier lässt sich auch leicht der Anzahleffekt zeigen: Ein Übergang von n gleich großen Einheiten zu n-1 gleich großen Einheiten bedeutet (analog zu Gl. 6.13) eine Vergrößerung von $K_H = 1/n$ zu $K_H^* = 1/(n-1)$.

6. Weitere Hinweise:

a) K_H nimmt in der Regel recht niedrige Werte an. Die amerikanischen Fusionsricht-linien stellten früher (1968) auf die concentration ratios bezogen auf die vier größten Unternehmen (also die Größe C_4) ab, neuerdings auf den Herfindahl-Index

	US-Fusionsrichtlinien	
Konzentrationsgrad	1968	1982
niedrig	$C_4 < 0{,}5$	$K_H < 0{,}1$
mittelhoch	$0{,}5 < C_4 < 0{,}7$	$0{,}1 < K_H < 0{,}18$
hoch	$C_4 > 0{,}7$	$K_H > 0{,}18$

Man kann zeigen, dass die angegebenen Wertebereiche sich in etwa entsprechen. Der Herfindahl-Index, aber auch die concentration ratios (oder "kritische Konzentrationskurven") werden auch zur Messung der Wettbewerbsintensität benutzt. So gilt z.B. nach §22 GWB als kritische Konzentration: $C_1 > 1/3$, $C_3 > 1/2$ und $C_5 > 2/3$.

b) Das Konzept des Herfindahl-Index läßt sich verallgemeinern zu

$$(6.13)\quad K_\alpha = (\Sigma\, q_i^\alpha)^{1/(\alpha-1)} \quad (\alpha > 0)$$

Es gilt $K_2=K_H$, Strebt α gegen 1 so geht K_α in den Exponentialindex über (vgl. Def. 6.5), wächst α über alle Grenzen ("gegen unendlich"), so strebt K_α gegen C_1, denn mit zunehmendem α wird den größeren Einheiten ein immer größeres Gewicht verliehen. Es gilt die Mittelwertungleichung $K_0 = 1/n \leq K_1 = E \leq K_2 = K_H \leq K_\infty = C_1$

Beispiel 6.2:

Man berechne den Herfindahl-Index für Beispiel 6.1.

Lösung 6.2:

$K_{H1} = 0{,}6^2 + 0{,}2^2 + 0{,}1^2 + 0{,}06^2 + 0{,}04^2 = 0{,}4152$;
$K_{H2} = 0{,}6^2 + 0{,}16^2 + 0{,}14^2 + 0{,}06^2 + 0{,}04^2 = 0{,}4104$ und
$K_{H3} = 2(0{,}3)^2 + 2(0{,}1)^2 + 2(0{,}05)^2 + 2(0{,}03)^2 + 2(0{,}02)^2 = 0{,}2076 = K_{H1}/2$.

c) Exponentialindex und Entropie

Die folgenden beiden Konzentrationsmaße werden nur kurz eingeführt. Die Eigenschaften der Maße werden nicht im einzelnen diskutiert.

Def. 6.5: Exponentialindex

$$(6.14)\quad E = q_1^{q_1} \cdot q_2^{q_2} \ldots q_n^{q_n} = \Pi q_i^{q_i} \qquad \text{mit} \qquad \frac{1}{n} < E < 1 \qquad (i=1,2,...,n)$$

Der Exponentialindex E ist ein mit den Anteilen q_i gewogenes geometrisches Mittel der Anteile q_i. Daraus folgt $E \leq K_H$, da K_H ein entsprechendes arithmetisches Mittel ist. Offensichtlich ist ld(E) = - K_E (K_E = Entropie, vgl. Def. 6.6).

Def. 6.6: Entropie

Die analog dem Streuungsmaß definierte Entropie lautet bei der Konzentrationsmessung:

$$(6.15)\quad K_E = \sum q_i \, \mathrm{ld}\left(\frac{1}{q_i}\right) = - \sum q_i \, \mathrm{ld}(q_i) \qquad \text{(Entropie)}$$

Häufig werden auch die von Theil vorgeschlagenen Maße der Redundanz herangezogen, weil K_E eher als Dekonzentrationmsmaß angesehen werden kann (K_E sinkt [steigt] mit zunehmender [abnehmender] absoluter Konzentration). Es gilt:

$$(6.16a)\quad K_T = \mathrm{ld}(n) - K_E \qquad \text{(absolute Redundanz)}$$

$$(6.16b)\quad K_T^* = \frac{K_T}{\mathrm{ld}(n)} \qquad \text{(relative Redundanz)}$$

Hierbei ist ld(x) der logarithmus dualis (Logarithmus zur Basis 2) der Größe x. Es gilt ldx = logx/log2 = 3,32193logx.

Bemerkungen zur Entropie (Def.6.6)

1. K_E (und entsprechend auch K_T und $K_T,^*$) hängen allein von den Größen q_i ab, sie erfüllen also Axiom K1. Bei vollständiger Konzentration ist also $K_E = -1 \cdot \mathrm{ld}(1) = 0$ [statt 1] und bei egalitärer Verteilung gilt dem-nach $K_E = -\Sigma(1/n)\mathrm{ld}(1/n) = \mathrm{ld}(n)$ [statt 1/n], weshalb auch anstelle von K_E die Größe $K_T,^*$, eine Lineartransformation von K_E, als Konzentrationsmaß vorgeschlagen wird. Man kann K_E - wie gesagt - auch eher als Dekonzentrationmsmaß verstehen, wie die folgende Überlegung zeigt:

Ausgehend von der egalitären Verteilung steigt K_E wenn Egalität zwischen n+m statt n Einheiten besteht (n,m>0), denn ld(n+m) > ld(n). Aus diesem Grunde erfüllt K_E auch nicht Axiom K4. Es ist aber offensichtlich, dass K5 erfüllt ist weil sich K_E (anders als K_T und $K_{T'}^*$) nicht durch Nullergänzung verändert.

2. Bemerkenswerte Vorzüge der Entropie sind ihre einfache Verallgemeinerungsfähigkeit auf zwei und mehr Konzentrationsmerkmale und ihre günstigen Aggregationseigenschaften. Was letztere betrifft, so können diese leicht an dem folgendem Beispiel (der Zusammenhang kann natürlich verallgemeinert werden; vgl. auch Beispiel 6.3) demonstriert werden:

   Angenommen die Einheiten (Unternehmen) 1 und 2 fusionieren und ebenso die Einheiten 3 und 4, so erhält man ausgehend von der "alten" Entropie

$$\text{(6.17a)} \quad K_E^{(1)} = q_1 \mathrm{ld}\left(\frac{1}{q_1}\right) + q_2 \mathrm{ld}\left(\frac{1}{q_2}\right) + q_3 \mathrm{ld}\left(\frac{1}{q_3}\right) + q_4 \mathrm{ld}\left(\frac{1}{q_4}\right)$$

   die "neue" Entropie

$$\text{(6.17b)} \quad K_E^{(2)} = (q_1+q_2)\mathrm{ld}\left(\frac{1}{q_1+q_2}\right) + (q_3+q_4)\mathrm{ld}\left(\frac{1}{q_3+q_4}\right).$$

   Zwischen $K_E^{(1)}$ und $K_E^{(2)}$ besteht folgender Zusammenhang

$$\text{(6.17c)} \quad K_E^{(1)} = K_E^{(2)} + K_E^{(3)} \text{ (Additionstheorem der Entropie),}$$

   wobei $K_E^{(3)}$ wie folgt definiert ist

$$\text{(6.17d)} \quad K_E^{(3)} = (q_1+q_2)\left\{\frac{q_1}{q_1+q_2}\mathrm{ld}\left[\frac{q_1+q_2}{q_1}\right] + \frac{q_2}{q_1+q_2}\mathrm{ld}\left[\frac{q_1+q_2}{q_2}\right]\right\}$$
$$+ \ (q_3+q_4)\left\{\frac{q_3}{q_3+q_4}\mathrm{ld}\left[\frac{q_3+q_4}{q_3}\right] + \frac{q_4}{q_3+q_4}\mathrm{ld}\left[\frac{q_3+q_4}{q_4}\right]\right\}$$

   Die Ausdrücke innerhalb der äußeren Klammern stellen die Konzentrationen innerhalb der neuen durch Fusion (von jeweils zwei Einheiten) entstandenen Gesamtunternehmen dar. Somit kann $K_E^{(3)}$ als "interne" Konzentration bezeichnet werden.
   Sind die fusionierten Unternehmen gleich groß (wenn also z.B. gilt $q_1 = q_2$ und entsprechend $q_3 = q_4$ so sind die Ausdrücke innerhalb der inneren Klammern ld(2) = 1 und mithin $K_E^{(3)}$ = ld(2)=1 und entsprechend gilt bei der Fusion von jeweils k Un-

ternehmen, die gleich groß sind $K_E^{(3)}$ = ld (m), so dass die Entropie durch die Fusion von $K_E^{(1)}$ auf $K_E^{(2)}$ um $K_E^{(3)}$ = ld(k) sinkt. Das bedeutet, dass (wie bereits gesagt) die Entropie das Axiom K4 (Proportionalitätsprobe) nicht erfüllt: wird jede Einheit in k gleich große Einheiten aufgeteilt, so erhöht sich K_E um ld (k) - statt sich zu verringern - (vgl. auch Beispiel 6.3). Axiom K4 würde dagegen bei Dekonzentration $K^* = K/k$ statt $K_E^{(1)} = K_E^{(2)}$ + ld (k) und bei Fusion $K^* = kK$ statt $K_E^{(2)} = K_E^{(1)}$ – ld (k) verlangen.

Das Additionstheorem (Gl. 6.17c) ist danach (analog einer Varianzzerlegung) wie folgt zu lesen

(6.17c)	Gesamtentropie	=	externe Entropie	+	interne Entropie	
	$K_E^{(1)}$	=	$K_E^{(2)}$	+	$K_E^{(3)}$	

Beispiel 6.3

a) Drei Unternehmen teilen sich einen Markt. Ihre Marktanteile sind $q_1 = 1/2$, $q_2 = 1/3$, $q_3 = 1/6$. Man berechne die Entropie.

b) Berechnen Sie die Entropie, wenn sich die drei Unternehmen des Teils a in jeweils
- zwei
- drei

gleich große Unternehmen aufspalten.

c) Man berechne die Entropie sowie die absolute und relative Redundanz für die folgenden Konzentrationszustände (Vektoren der Marktanteile) Q_A = [0,3 0,7] und Q_B = [0,15 0,15 0,35 0,35].

Lösung 6.3:

a) K_E = 1,45915

b) bei zwei: K_E = 1,45915 + ld(2) = 1,45915 + 1 = 2,45915
bei drei: K_E = 1,45915 + ld(3) = 1,45915 + 1,58496 = 3,04411.
(für die relative Redundanz erhält man aber: $K_{T'}^{*} = K_T$/ld4 = K_T/2 = 0,11871/2 = 0,05936 [wie es Axiom K4 fordert]).

c) bei Q_A: K_E = 0,88129, K_T = 1 - 0,88129 = 0,11871 = $K_{T'}^{*}$
bei Q_B K_E = 1,88129, K_T = 2 - 1,88129 = 0,11871
(also unverändert, wie es eigentlich von einem Disparitätsmaß, nicht aber von einem Konzentrationsmaß, gefordert wird.)

4. Messung der relativen Konzentration (Disparität)

a) Lorenzkurve und Gini-Koeffizient

aa) Berechnung bei Einzelbeobachtungen

Bei der Berechnung des Gini-Koeffizienten ist es nützlich zu unterscheiden, ob der Berechnung Einzelbeobachtungen oder eine klassierte Verteilung zugrunde liegen. Es soll deshalb zunächst der Fall geordneter Einzelbeobachtungen betrachtet werden.

1. Definition bei Einzelbeobachtungen

Wie bei der Konzentrationsmessung gibt es auch bei der Disparitätsmessung eine graphische Darstellung und eine summarische Maßzahl (vgl. Übers. 6.1). Bei der Konzentrationsmessung wird grundsätzlich von Einzelbeobachtungen ausgegangen, während für die Disparitätsmessung auch die klassierte Verteilung und als theoretisches Modell die stetige Verteilung (vgl. Abschn. 6b dieses Kapitels) von Interesse ist.

Def. 6.7: Lorenzkurve und Gini-Koeffizient bei Einzelbeobachtungen

a) Lorenzkurve

Wird der Merkmalsanteil des i-ten Merkmalsträgers bei einer Ordnung nach **zunehmender** Größe

$$(6.18)\quad q_i = \frac{x_{(i)}}{\Sigma x_{(j)}} = \frac{x_{(i)}}{\Sigma x_j} \qquad (i,j = 1,2,...,n)$$

genannt (vgl. Def. 6.3), dann ist

$$(6.19)\quad Q_i = \Sigma\, q_j \qquad (j = 1,2,...,i)$$

der kumulierte Anteil der i kleinsten Merkmalsträger am Merkmalsbetrag. Die lineare Verbindung der Punkte $P_i(H_i,Q_i)$ mit den kumulierten relativen Häufigkeiten H_i und den kumulierten Anteilen Q_i im H-Q-Kordinatensystem heißt Lorenzkurve ($H_0 = Q_0 = 0$ und $H_n = Q_n = 1$). Für die H_i gilt im Fall von Einzelbeobachtungen:

$$H_i = \frac{i}{n}\ .$$

b) Gini-Koeffizient

Die Größe

$$(6.20)\quad D_G = \sum_{i=1}^{n} \frac{2i-n-1}{n} q_i \qquad \left(0 \le D_G \le 1 - \frac{1}{n}\right)$$

heißt Disparitätskoeffizient von Gini (oder einfach Gini-Koeffizient).

2. Einführendes Beispiel für die Lorenzkurve

Beispiel 6.4:

In einer islamischen Familie mit 4 Kindern sind 2 männlich und 2 weiblich. Das an die Kinder zu vererbende Vermögen von 1200 Dinare soll getreu nach den Regeln des Korans vermacht werden:

> *"Und wenn die Geschwister Männer und Frauen sind, so soll ein Mann so viel erhalten wie zwei Frauen" (Sure 4, Vers 175)*

Man bestimme die sich bei Befolgung des islamischen Erbrechts ergebende Lorenzkurve und den Gini-Koeffizienten!

Lösung 6.4:

Die beiden Töchter bekommen jeweils 200 (zusammen 400) und die beiden Söhne jeweils 400 (zusammen also 800) Dinare.
Die Anteile sind $h_1 = h_2 = h_3 = h_4 = 1/4$ (wobei 1 und 2 die beiden Töchter und 3 und 4 die beiden Söhne symbolisieren) und $q_1=q_2=1/6$ (so dass $Q_2 = q_1 + q_2 = 1/3$) ferner ist $q_3 = q_4 = 1/3$ so dass $q_3 + q_4 = 2/3$. Man sieht leicht, dass $D_G = 1/6$ ist.

3. Bemerkungen zur Lorenzkurve

1. Die Reihenfolge der Messwerte (nach zunehmenden Anteilen q_i) ist wesentlich.

2. Die Größen H_i und Q_i sind nicht unabhängig voneinander. Vielmehr gilt wegen $Q_i = \Sigma q_j$ (mit $j = 1,2,...,i$) und $q_j = x_{(j)} / n\bar{x} = h_j(x_{(j)} / \bar{x})$, so dass zwischen den Anteilen q_j und h_j folgender Zusammenhang besteht:

$$(6.21)\quad \frac{q_j}{h_j} = \frac{x_{(j)}}{\bar{x}}$$

Der Bruch q_j / h_j ist die Steigung der Lorenzkurve, die somit proportional ist zu $x_{(j)}$ (denn $1/\bar{x}$ ist eine Konstante): vgl. Bem. 6.

3. Die Lorenzkurve stellt den funktionalen Zusammenhang zwischen den kumulierten Anteilen am Merkmalsbetrag Q_i und den kumulierten relativen Häufigkeiten H_i dar. Es gilt

   $$\boxed{Q_i = L(H_i).}$$

   Die Funktion L heißt Lorenzkurve.

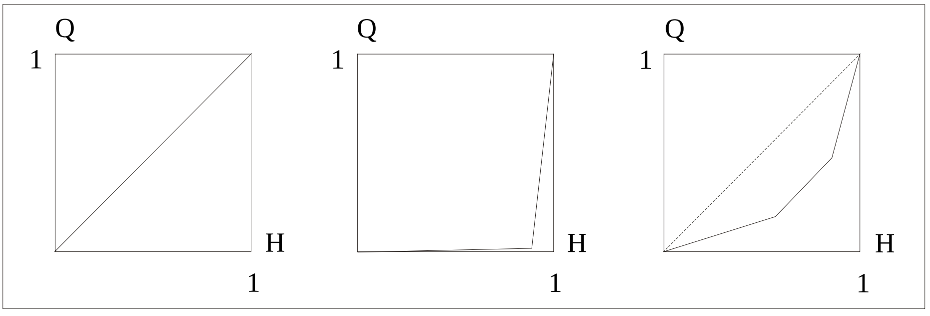
4. Über die Extremfälle der Disparität läßt sich hinsichtlich der Gestalt von L folgendes aussagen (Abb. 6.6).

   a) Bei egalitärer Verteilung ist die Funktion L eine Gerade (die 45-Grad-Linie). Es gilt wegen $q_i = 1/n$ für alle $i=1,2,...,n$ die Gleichung $Q_i=H_i$. Die Lorenzkurve ist dann mit der Geraden zwischen den Punkten $P_0(0,0)$ und $P_n(1,1)$, der Gleichverteilungsgeraden, identisch (Abb. 6.6, links).

   b) Bei vollkommener Ungleichheit ist die Lorenzkurve aus der Strecke von $P_0(0,0)$ bis $P_{n-1}(n-1/n,0)$ und dem Punkt $P_n(1,1)$ zusammengesetzt (Abb. 6.6, Mitte).

   c) In allen realistischen Fällen liegt die Lorenzkurve zwischen diesen beiden Extremfällen, als konvexer Polygonzug unterhalb der Gleichverteilungsgeraden (Abb. 6.6, rechts).

Abb. 6.6: Lorenzkurve und extreme Fälle von Disparität



5. Man kann zeigen, dass die Fläche zwischen der Lorenzkurve und der Gleichverteilungsgeraden, die auch Konzentrationsfläche[2] genannt wird

   $$F = \sum \frac{2i-n-1}{2n} q_i$$

---

[2] Der Begriff ist aber auch üblich für die Fläche oberhalb der Konzentrationskurve (vgl. Gl. 6.7).

beträgt.
Daraus folgt, dass der Gini-Koeffizient das Verhältnis zwischen F und der Dreiecksfläche unterhalb der Gleichverteilungsgeraden ist (diese Dreiecksfläche beträgt ½), d.h. es gilt

$$(6.22) \qquad D_G = \frac{F}{½} = 2F \quad .$$

Bei egalitärer Verteilung ist F=0 und somit auch D_G=0; Bei vollkommener Ungleichheit ist F = ½ - ½(1/n) und somit D_G=(n-1)/n. Es gilt für D_G also die im Axiom K6 geforderte Einschrankung.

6. Die Steigung der Lorenzkurve zwischen den Punkten $P_{i-1}(H_{i-1}, Q_{i-1})$ und $P_i(H_i,Q_i)$ beträgt wegen Gl. 6.21

$$(6.23) \qquad \frac{q_i}{h_i} = \frac{x_{(i)} / n\bar{x}}{1/n} = \frac{x_{(i)}}{\bar{x}} \qquad \text{(Steigung der Lorenzkurve)}$$

Die Steigung der Lorenzkurve ist also nichtnegativ und von 0 bis unendlich monoton steigend, d.h. sie nimmt mit zunehmendem x und H zu.

Daraus folgt unmittelbar:

- Die Steigung erreicht den Wert 1, d.h. die Lorenzkurve verläuft parallel zur Gleichverteilungsgeraden, wenn $x_{(i)} = \bar{x}$, d.h. wenn das Einkommen (der Merkmalsbetrag) der i-ten Einheit gleich dem Durchschnittseinkommen ist.
- Die Lorenzkurve verläuft also zunächst flacher (bei Beziehern unterdurchschnittlicher Einkommen), dann ab $x_{(i)} = \bar{x}$ steiler als die Gleichverteilungsgerade (bei Beziehern überdurchschnittlicher Einkommen).
- Weil die Steigung der Lorenzkurve nicht abnehmen kann, kann die Lorenzkurve die Gleichverteilungsgerade nicht schneiden; das folgt auch aus $H_i \geq Q_i$,($H_i = Q_i$ gilt außer bei egalitärer Verteilung nur für i=0 und i=n), denn

$$(6.24) \quad H_i = \frac{i}{n} \leq \frac{\sum_{j=1}^{i} x_j}{n\bar{x}} = \frac{i}{n}\frac{\bar{x}_i}{\bar{x}} = Q_i$$

wobei $\bar{x}_i$ der <u>mittlere</u> Merkmalsbetrag der ersten i Merkmalsträger ist (nicht zu verwechseln mit $x_{(i)}$, dem Merkmalsbetrag der i-ten

Einheit), der wegen der Reihenfolge der Merkmalsträger notwendig stets kleiner ist als der Mittelwert $\bar{x}$, der sich auf alle n Merkmalsträger bezieht.

Da $\bar{x}_i < x_{(i)}$, ist die Steigung der Strecke P_0P_i mit $Q_i / H_i = \bar{x}_i / \bar{x}$ kleiner als die Steigung der Tangente im Punkt P_i der Lorenzkurve, die $q_i / h_i = x_{(i)} / \bar{x}$ beträgt.

7. Zwischen dem Gini-Koeffizienten D_G und dem Rosenbluth-Index K_R besteht folgender Zusammenhang

$$(6.25) \qquad D_G = 1 - \frac{1}{nK_R} \text{ und somit } K_R = \frac{1}{n(1-D_G)}$$

Dabei ist zu beachten, dass die Reihung der Merkmalsträger bei der Konzentrationsmessung nach abnehmender und bei der Disparitätsmessung nach zunehmender Größe erfolgt.

8. Eine Lineartransformation der Merkmalswerte mit y_i=a+bx_i wirkt auf die Merkmalsanteile wie folgt ($q_i,^*$ ist ein gewogenes Mittel aus q_i und dem Wert i/n der Winkelhalbierenden [egalitäre Verteilung]),

$$q_i,^* = \left(\frac{a}{\bar{y}}\right) \cdot \frac{i}{n} + \left(\frac{b\bar{x}}{\bar{y}}\right) \cdot q_i$$

in Analogie zu Gl. 6.8, d.h. bei a,b >0 gilt $q_i \leq q_i^* \leq i/n$ (Mittelwerteigenschaft), so dass man für das Disparitätsmaß von Gini erhält

$$(6.26) \qquad D_G^* = \left(\frac{b\bar{x}}{\bar{y}}\right) D_G \, .$$

Bei proportionaler Transformation (Axiom K1) gilt a = 0 und wegen $\bar{y} = b\bar{x}$ auch $D_G^* = D_G$. Bei einer Niveauänderung im Sinne des Axioms K3 gilt b = 1 und folglich

$$(6.26a) \qquad D_G^* = \frac{\bar{x}}{a+\bar{x}} D_G \, .$$

so dass offensichtlich die Axiome K1 und K3 erfüllt sind. Bei a > 0 und b=1 rückt die Lorenzkurve in allen Punkten im Intervall 0 < H < 1 näher an die Gleichverteilungsgerade heran.

Bei der Verschiebungsprobe (Axiom K2) gilt bei einem Transfer von i zu j mit $x_{(j)} < x_{(i)}$ für die kumulierten Anteile

$$Q_j^* = Q_j + q_d \quad \text{und} \quad Q_i^* = Q_i$$

mit $q_d = d/n\bar{x}$, so dass die Lorenzkurve im Intervall $H_{j-1} < H < H_i$ näher an die Gleichverteilungsgerade heranrückt, wie es Axiom K2 bei einem negativen (egalisierenden) Transfer fordert.

10. Es ist leicht zu sehen, dass die Proportionalitätsprobe (Axiom K4) und die Ergänzungsprobe (Axiom K5) ebenfalls erfüllt sind. Die Zusammenhänge seien am Beispiel 6.5 demonstriert, gelten aber allgemein.

11. Ginis Disparitätsmaß steht auch im Zusammenhang mit Parametern der Pareto- Verteilung und der logarithmischen Normalverteilung, worauf jedoch aus Platzgründen nicht eingegangen werden kann.

12. Man kann D_G auch als einen speziellen Variationskoeffizienten auffassen. Zwischen D_G und Ginis "mittlere Differenz" S_{G*}

$$S_{G*} = \frac{1}{n^2} \sum_i \sum_k | x_i - x_k | \qquad i,k = 1,2,...,n ,$$

einem Steuungsmaß (vgl. Def. 5.5), besteht der folgende Zusammenhang:

$$(6.27) \qquad D_G = \frac{S_{G*}}{2\bar{x}} .$$

Daraus folgt übrigens auch, dass man D_G darstellen kann als

$$(6.27a) \qquad D_G = \frac{\Sigma\Sigma|q_i - q_k|}{2n} \qquad i,k = 1,...n.$$

13. Analog zu Gl. 6.27 gewinnt man auch ein Dispersionsmaß mit der durchschnittlichen Abweichung um $\bar{x}$ (d_x^* gem. Def. 5.3) als

$$(6.27b) \qquad \phi^* = \frac{d_x^*}{2\bar{x}}$$

das als Schutz-Koeffizient oder maximaler Nivellierungssatz (oder längste Lorenzkurvensehne) bekannt ist (vgl. Def. 6.13) und nach Gl. 6.44 nicht größer sein kann als der Gini-Koeffizient D_G.

14. Es ist offensichtlich, dass (unendlich viele) verschiedene Lorenzkurven trotz unterschiedlichen Verlaufs zum gleichen Flächenverhältnis D_G führen können, so dass das Maß D_G also einen Disparitätszustand nicht eindeutig abbildet.

15. Es ist bemerkenswert, dass Gini sein Disparitätsmaß D_G ursprünglich ohne Bezugnahme auf die Lorenzkurve entwickelte. Erst später erkannte er die Flächeninterpretation von D_G (Vgl. Bem. 5):

Die relativen Abstände

$$R_i = \frac{H_i - Q_i}{H_i}$$

zwischen der Lorenzkurve und Gleichverteilungsgeraden galten für Gini als Ausdruck der Disparität. Man kann zeigen, dass die Größe L^*, ein mit den kumulierten relativen Häufigkeiten H_i gewogenes arithmetisches Mittel der Größen R_i

$$L^* = \sum_{i=1}^{n} R_i \frac{H_i}{SH_i} = \frac{\Sigma(H_i - Q_i)}{SH_i} \quad (i = 1,2,...,n)$$

gegen den Disparitätskoeffizienten D_G strebt, d.h. dass gilt

$$D_G = \lim_{n \to \infty} L^* \qquad (i=1,2,...,n).$$

Aus diesem Grunde ist auch die Größe $L^* = \Sigma(H_i-Q_i)/\Sigma H_i$ bei hinreichend großer Anzahl n von Merkmalsträgern (oder von Klassen) näherungsweise der Gini-Koeffizient.

Beispiel 6.5:

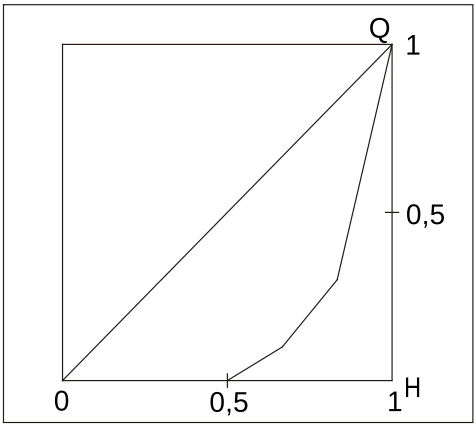
Ausgangspunkt sei eine Verteilung mit $q_1 < q_2 < q_3$ und $h_i = 1/3$ (i=1,2,3). Die *Aufspaltung jeder Einheit* in zwei gleich große Einheiten bewirkt, dass die neue Lorenzkurve die folgenden Punkte verbindet:

$P_1(0,0)$, $P_2(1/6,½q_1)$, $P_3(1/3,q_1)$, $P_4(½,q_1+½q_2)$, $P_5(2/3,q_1+q_2)$, $P_6(5/6,q_1+q_2+½q_3)$ und $P_7(1,1)$.

Die Punkte P_0, P_3, P_5 und P_7 sind identisch mit der bisherigen Lorenzkurve. Die Punkte P_2, P_4 und P_6 liegen jeweils auf der linearen Verbindung zwischen P_0 und P_3 usw. Die Lorenzkurve (und damit auch D_G) bleibt also unverändert.

Eine *Nullergänzung* durch 3 Nullträger auf der Basis der obigen Ausgangssituation bewirkt dagegen, dass die neue Lorenzkurve zwischen $0 \leq H \leq ½$ den Wert Q=0 annimmt (vgl. Abb. 6.7).

Abb. 6.7: Lorenzkurve bei Nullergänzung von n=3 Einheiten



Das Gleichheitsmaß $1\text{-}D_G = G$ betrug vor der Nullergänzung

$G_1 = 1/3\ (5q_1 + 3q_2 + q_3)$ und nach der Nullergänzung
$G_2 = 1/6\ (5q_1 + 3q_2 + q_3)$.

Es hat sich also halbiert wie es Axiom 5a bei k=2 also einer Ergänzung um m = n(k-1) = 3(2-1) = 3 Nullträgern verlangt.

bb) Berechnung bei gruppierten und klassierten Daten

Gegeben seien m Ausprägungen des Konzentrationsmerkmals bzw. m Klassen, so dass mit h_i die relativen Häufigkeiten und q_i die Anteile am Gesamtmerkmalsbetrag definiert sind (i = 1,...,m). Die kumulierten Anteile sind wieder $H_i = \Sigma h_j$ und $Q_i = \Sigma q_j$ (j=1,...,i).
Nach der Definition des Gini-Koeffizienten folgen zunächst einige weitere Bemerkungen zu den Eigenschaften der Lorenzkurve und dann zwei einführende Beispiele (Bsp. 6.6 und 6.7).

Def. 6.8: Lorenzkurve und Gini-Koeffizient bei klassierten Daten

a) Die lineare Verbindung der Punkte $P_i(H_i,Q_i)$ (i=0,1,...,m) mit $P_o(0,0)$ und $P_m(1,1)$ heißt Lorenzkurve.

b) Der Gini-Koeffizient D_G ist gegeben mit

(6.28) $D_G = 1 - \Sigma h_i(Q_i+Q_{i-1})$ oder (6.28a) $D_G = \Sigma q_i(H_i+H_{i-1}) - 1$

Diese Gleichungen gelten zunächst nur bei gruppierten Daten. Für die Berücksichtigung der Disparität innerhalb der Klassen bei klassierten Daten vgl. Bem. 4.

Bemerkungen zu Def. 6.8:

1. Die Bemerkungen im Abschnitt aa) gelten analog auch hier.

2. In Gl. 6.28 und 6.28a ist i=1 der Wert $H_{i-1} = H_{1-1} = H_0 = 0$ und entsprechend $Q_o = 0$ einzusetzen. Setzt man in Gl.6.28a für H_i den Wert i/n und entsprechend für H_{i-1} den Wert (i-1)/n ein, wie dies dem Fall der Einzelbeobachtungen entspricht, so folgt Gl. 6.20 aus Gl. 6.28a.

3. Wenn zwei und mehr aneinander angrenzende Größenklassen zusammengefaßt werden, etwa die Klassen i und i+1, so verläuft die "neue" Lorenzkurve flacher (näher an der Gleichverteilungsgeraden). Während sie vorher die Punkte P_{i-1}, P_i und P_{i+1} linear verband, stellt die Lorenzkurve jetzt eine Verbindung der Punkte P_{i-1} und P_{i+1} dar (vgl. Beispiel 6.7). Man kann leicht zeigen, dass sich die Fläche zwischen Lorenzkurve und Gleichverteilungsgerade dadurch verringert, und zwar um den Betrag $\frac{1}{2}(h_i q_{i+1} - h_{i+1} q_i)$, so dass gilt:

   > D_G verringert sich um $(h_i q_{i+1} - h_{i+1} q_i)$ wenn die Größenklassen i und i+1 zusammengefaßt werden.

   Der Ausdruck $h_i q_{i+1} - h_{i+1} q_i$ ist übrigens deshalb nichtnegativ, weil die Steigung der Lorenzkurve monoton zunimmt, d.h. weil gilt
   $q_{i+1}/h_{i+1} \geq q_i/h_i$.

4. Für Ginis Dispersionsmaß existiert eine Zerlegung nach Art der Streuungszerlegung

   $$\boxed{(6.29)\quad D_G = \left[1 - \sum h_i(Q_i+Q_{i-1})\right] + \left[\sum h_i^2 \frac{\bar{x}_i}{\bar{x}} D_G^{(i)}\right]}$$

   wobei $D_G^{(i)}$ die Disparität innerhalb der i-ten Klasse darstellt, der gesamte Ausdruck in der zweiten eckigen Klammer also die interne Disparität ist (und die erste eckige Klammer die bisher allein betrachtete externe Disparität nach Gini).

5. Wegen des unter Nr. 3 dargestellten Zusammenhangs wird bei Aggregation (Zusammenlegung) von Größenklassen die Disparität abnehmen und bei Disaggregation (Aufspaltung) entsprechend zunehmen. Jede empirische, als Polgonzug dargestellte Lorenzkurve kann nur eine Näherung für die im Abschn. 6 definierten, und den Daten "eigentlich" zugrundeliegenden stetige Lorenzkurve sein. Ein Vergleich von Disparitäten (z.B. der Einkommensverteilungen von zwei Ländern) sollte deshalb nur bei gleich vielen Größenklassen erfolgen.

6. Aus Gl. 6.28 entwickelt sich leicht eine Formel für den Spezialfall, dass die Verteilung (z.B. die Einkommensverteilung) nur aus zwei Merkmalsausprägungen $x_2 > x_1$ besteht:

Einkommen	relative Häufigkeit	Anteil am Gesamtmerkmalsbetrag	Merkmalsbetrag
gering	h	q	x_1
groß	1-h	1-q	x_2

Die Lorenzkurve hat dann nur drei Punkte $P_0(0,0)$, $P_1(h,q)$ und $P_2(1,1)$ und Ginis Dispersionsmaß ist dann der senkrechte Abstand zwischen dem Punkt $P_1(h,q)$ und der Gleichverteilungsgeraden, also die Strecke von $P_1(h,q)$ bis $P_1^*(h,h)$.

$$\boxed{(6.28b) \quad D_G = h - q\,.}$$

Diese Formel zeigt erneut, dass unterschiedliche Disparitätszustände zu gleichen Werten des Gini-Koeffizienten D_G führen können (eine häufig vorgebrachte Kritik an D_G): Haben z.B. in einem Land die 30% ärmsten Familien einen Anteil am Vermögen von nur 5% so ist die Vermögenskonzentration (genauer: Vermögensdisparität) genau so groß wie in einem anderen Land, in dem die 40% ärmsten Familien einen Anteil von 15% haben, denn $D_G = 0{,}3-0{,}05 = 0{,}4-0{,}15 = 0{,}25$.
Weiter kann man zeigen, dass in dieser Situation für den Variationskoeffizienten gilt

$$(6.28c) \quad V = \frac{D_G}{\sqrt{h(1-h)}} \leq 2D_G.$$

Für die durchschnittliche Abweichung um $\bar{x}$ erhält man schließlich

$$d_x^* = 2(x_2 - \bar{x})(1 - h),$$

so dass der Schutz-Koeffizient ϕ^* gem. Gl. 6.27b dann

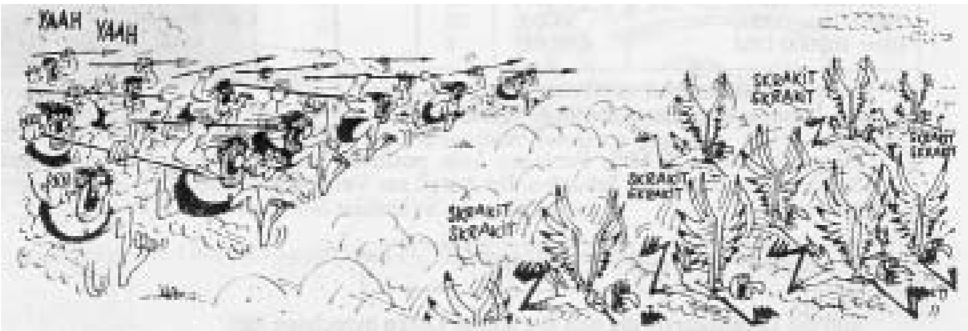
$$(6.28d) \quad \frac{d_{x,}{}^*}{2x,\bar{}} = \phi^* = \frac{2x_2(1-h)}{2\bar{x}} - \frac{2\bar{x}(1-h)}{2\,\bar{x}} = h-q = D_G$$

ist (da $x_2 / \bar{x} = 1-q$).

<u>Beispiel 6.6:</u>

Urmenschenproblem: 150 Angehörige eines primitiven Volksstammes (150 "Urmenschen") gehen auf die Jagd nach Federvieh. Ihre Beute beträgt 300 Wildgänse. Durch das an sich nur bei primitiven Völkern bekannte Gerangel um Geld, Gut und Prestige entstand eine etwas ungleiche Verteilung der Beute. Durch Eingreifen des Häuptlings konnte jedoch noch verhindert werden, dass jemand leer ausging. Es bekamen jeweils n_i Personen x_i Gänse:

x_i	1	2	3	4
n_i	60	45	30	15



Man zeichne die Lorenzkurve und berechne das Disparitätsmaß D_G von Gini. Wie sähe die Lorenzkurve aus, wenn jeder von der Beute gleichviel bekommen hätte?

<u>Lösung 6.6:</u>

Es sind 300 Gänse auf 150 "Urmenschen" zu verteilen; bei Gleichverteilung gilt also $\bar{x} = 300/150 = 2$ Gänse für jeden. Es gibt also "Unterprivilegierte" mit nur $x = 1$ Gans und "Überprivilegierte" mit $x = 3$ oder gar $x = 4$ Gänsen. Für die Anteile h_i an den $n = \Sigma n_i = 150$ Urmenschen und die Anteile q_i an den $\Sigma\, x_i n_i = 300$ Gänsen gilt dann:

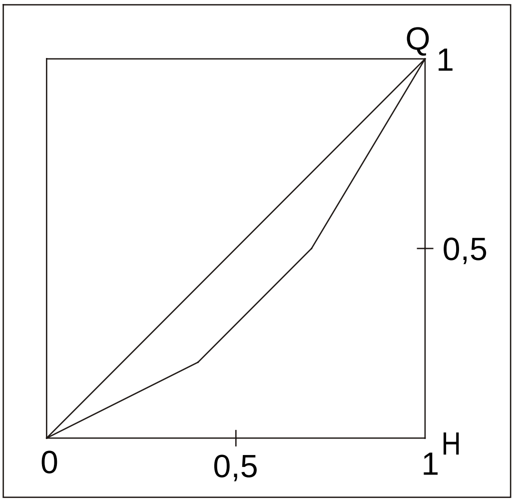
		x_i	1	2	3	4	Σ
A	Urmenschen	n_i	60	45	30	15	150
B	Gänse	$x_i n_i$	60	90	90	60	300
Anteile an A	$h_i = n_i / \Sigma\, n_i$		0,4	0,3	0,2	0,1	1,0
	$H_i = \Sigma\, h_j\ (j \leq i)$		0.4	0,7	0,9	1	
Anteile an B	$q_i = x_i n_i / \Sigma\, x_i n_i$		0,2	0,3	0,3	0,2	1,0
	$Q_i = \Sigma\, q_j\ (j \leq i)$		0,2	0,5	0,8	1	

Anteile an A = Anteile an den Urmenschen (an den Merkmalsträgern)
Anteile an B = Anteile an den Wildgänsen (an den Merkmalssumme).

Wie man sieht gilt $Q_i \leq H_i$, was ja gerade die "Ungleichverteilung" ausmacht, denn die H_1 = 40% ärmsten Urmenschen haben nicht einen Anteil von 40%, sondern von weniger, nämlich nur 20% (denn $Q_1 = 0{,}2$) an der Beute. Entsprechend haben die 70% ärmsten nicht einen Anteil von 70%,

sondern nur von 50% usw. Die Lorenzkurve dieses Beispiels ist in der Abb.6.8 dargestellt.
Für Ginis Koeffizient erhält man in diesem Beispiel $D_G = 0{,}27$.

Abb.6.8: Lorenzkurve für das Beispiel 6.6



Beispiel 6.7:

Aus der Vermögenssteuerstatistik 1983 (Statistisches Jahrbuch der Bundesrepublik Deutschland 1989, S.451) erhält man folgende (stark zusammengefaßte) Daten zur Schichtung des Gesamtvermögens von unbeschränkt vermögenssteuerpflichtigen natürlichen Personen:

Gesamtvermögen	h_i	q_i	H_i	Q_i
unter 200000	0,2417	0,0481	0,2417	0,0481
200 - 500000	0,4458	0,2011	0,6875	0,2492
0,5 - 1 Mill	0,1898	0,1815	0,8773	0,4307
1 - 5 Mill	0,1088	0,2813	0,9860	0,7200
5 - 20 Mill	0,0119	0,1460	0,9980	0,8580
über 20 Mill	0,0020	0,1420	1,0	1,0

a) Zeichnen Sie die Lorenzkurve und berechnen Sie den Gini- und den Herfindahl-Koeffizienten D_G und K_H.

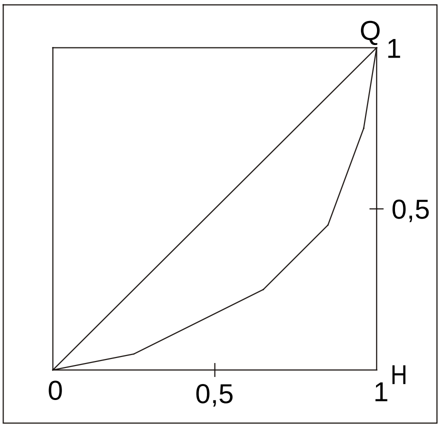
b) Wie ändert sich die Lorenzkurve und wie ändern sich und D_G und K_H wenn man die ersten beiden Größenklassen zusammenfaßt?

Lösung 6.7:

a) $D_G = 0{,}5512$ und $K_H = 0{,}3998$.

b) Die Lorenzkurve hat statt 7 Punkte (einschließlich $P_0(0,0)$ und $P_7(1,1)$) jetzt nur noch 6 Punkte, da die Verteilung nun nur noch 5 Größenklassen hat. Die bisherigen Punkte P_0 und P_2 werden linear verbunden. Gini- und Herfindahl-Index nehmen entsprechend ab: es gilt jetzt $D_G = 0{,}5241$ und $K_H = 0{,}4191$.

Abb. 6.9: Lorenzkurve für Beispiel 6.7



b) Der Variationskoeffizient als Disparitätsmaß

Der bereits als Maß der relativen Streuung eingeführte Variationskoeffizient $V = s/\bar{x}$ gilt auch als Maß der Disparität. Neben dem Variationskoeffzienten V wird auch das normierte Quadrat des Variationskoeffzienten als Disparitätsmaß benutzt.

Def. 6.9: normiertes Quadrat des Variationskoeffizienten

$$\boxed{(6.30)\quad NV = \frac{V^2}{1+V^2} = 1 - \frac{1}{nK_H}}$$

Interpretation und Eigenschaften

1. Für V gilt die folgende Beziehung zum Konzentrationsmaß von Herfindahl :

   $$\boxed{(6.31)\quad V^2 = n\Sigma q_i^2 - 1 = nK_H - 1}$$

   bei Einzelbeobachtungen und
   (6.31a) $V_2 + 1 = K_N / \bar{x}$

bei gruppierten Daten, wobei K_N das arithmetische Mittel der Momentverteilung (vgl. auch Abschn. 6a, Gl. 6.35) ist. Mit $h_i = 1/n$ und $q_i = x_i / n\bar{x}$ erhält man Gl. 6.31 aus Gl. 6.31a.

Für V gilt bei den Extremsituationen der Disparität

- egalitäre Verteilung $V^2 = n\Sigma(1/n)^2 - 1 = 0$ und $NV = 0$
- vollkommene Ungleichheit $V^2 = n - 1$ und $NV = 1 - 1/n$.

NV ist also nur der auf den Wertebereich von 0 bis (n-1)/n eines Disparitätsmaßes normierter Variationskoeffizient und hat ansonsten die gleichen Eigenschaften wie V.

2. Man kann V nach Gl. 6.31, bzw. NV allein in Abhängigkeit von n und den Merkmalsanteilen q_i darstellen, so dass Axiom K1 erfüllt ist. Beide Maße erfüllen auch K2.

3. Offensichtlich ist auch das Axiom K3 erfüllt, denn bei der Lineartransformation $y_i = a + x_i$ gilt

$$\boxed{(6.32)\quad V_y = \frac{\bar{x}}{a+\bar{x}} V_x \ ,}$$

so dass $V_y < V_x$ wenn $a > 0$ und $V_y > V_x$ wenn $a < 0$.

4. NV erfüllt auch K4 und K5. Die Disparität, gemessen an V (und NV) steigt bei Nullergänzung (Hinzukommen von Nullträgern) aber nicht in dem in Axiom K5a geforderten Ausmaß.

c) Disparität und verwandte Konzepte

1. Disparität und relative Streuung

Die Verwendung des Variationskoeffizienten V als Disparitätsmaß oder der Größe $\frac{1}{2}S_{G*} / \bar{x}$ (vgl. Gl. 6.27) bzw. $d_x^* / 2\bar{x}$ (vgl. Gl. 6.27b) als Disparitätsmaß wirft die Frage nach dem Unterschied der beiden Konzepte "Disparität" und "relative Streuung" auf.
Es gibt Überschneidungen zwischen beiden Maßzahlenklassen in dem Sinne, dass Maßzahlen möglich sind, die ganz oder zum größten Teil die Axiome beider Klassen erfüllen (abgesehen von dem Wertebereich, also Axiom K6 sowie von dem Axiom K5a). Das bedeutet jedoch nicht, dass die Forderungen, die an die jeweilige Klasse von Maßzahlen gestellt werden, auf das gleiche hinauslaufen.

So wäre beispielsweise das Verhältnis von Spannweite und Median ein brauchbares Maß der relativen Streuung. Es sei SM genannt. Offenbar erfüllt SM auch das Axiom S2 (Vergrößerung der Streuung bei Hinzutreten eines Wertes, der größer (kleiner) als der bisher größte (kleinste) Wert ist, das im besonderen Maße den Gedanken der Streuung wiedergibt. Ähnlich eng verbunden mit der anschaulichen Vorstellung ist im Falle der "Disparität" das Axiom K2 (Transfer), das jedoch von SM nicht erfüllt sein muss. Ein Transfer im Sinne von Axiom K2 wirkt sich auch meist nicht aus auf die mittlere Abweichung, die gleichwohl Grundlage eines sinnvollen Maßes der Streuung sein kann.

2. Disparität und Schiefe

Im Sprachgebrauch von Politik und Wirtschaft wird Disparität ("Ungleichheit") auch häufig im Sinne einer linkssteilen Verteilung gebraucht: die große Mehrheit verdient wenig und einige wenige "Besserverdienende" verdienen vergleichsweise viel. Schiefe und Disparität sind zwei Aspekte einer Häufigkeitsverteilung die durchaus **Gemeinsamkeiten** haben, nämlich

1. beide sind invariant gegenüber proportionalen Transformationen, d.h. sie erfüllen das Axiom K1 und
2. beide sind unabhängig von der Anzahl der Merkmalsträger in dem Sinne wie dies bei der Proportionalitätsprobe (Axiom K4) postuliert wird.

Es gibt jedoch bedeutende **Unterschiede**:

1. Besonders auffallend ist, dass die Schiefe als Gestaltparameter notwendig verschiebungsinvariant ist, d.h. dass sie auf Niveauänderungen nicht reagiert und somit Axiom K3 nicht erfüllt.
2. Nicht in gleicher Weise unmittelbar einsichtig ist, dass die Schiefe nicht Axiom K2 erfüllt, also beispielsweise sich bei einem negativen (egalisierenden) Transfer nicht notwendig verringert.
3. Bei einer Nullergänzung im Sinne des Axioms K5 soll ein Disparitätsmaß zunehmen, die Schiefe kann jedoch auch abnehmen.

Die Punkte 2 und 3 seien kurz an Beispielen veranschaulicht:

<u>zu 2:</u>

Durch einen Transfer des Betrags d von i auf j ($x_{(i)} > x_{(j)}$) im Sinne des Axioms K2 verringert sich stets die Varianz von s^2 auf s^{2*} mit

$$s^{2*} = s^2 + \frac{2d}{n} \quad (x_{(j)} - x_{(i)} + d) < s^2$$

und damit auch die Standardabweichung. Anders verhält es sich jedoch mit dem dritten zentralen Moment z_3, das zwar abnehmen kann, möglicherweise aber in einem geringeren Maße als die dritte Potenz der Standardabweichung, so dass sich die Schiefe erhöht.

Als Beispiel sei die folgende linkssteile Ausgangsverteilung gewählt: $x_{(1)} = 100$ $x_{(2)} = 200$ und $x_{(3)} = 600$.
Das dritte zentrale Moment z_3 verringert sich von 6 auf 3,75 Millionen bei einem Transfer des Betrags 50 von Einheit 3 auf Einheit 1, die dann den Betrag von 150 hat. Dadurch steigt die Momentschiefe von 0,59517 auf 0,66547, d.h. die Verteilung ist trotz eines egalisierenden Transfers im höheren Maße linkssteil.

zu 3:

Gegeben sei die folgende linkssteile Ausgangsverteilung mit den Beträgen 100, 100, 100 und 200. Tritt zu den n = 4 Personen ein Nullträger hinzu, so wird die linkssteile Verteilung (Momentschiefe 1,1547) zu einer symmetrischen Verteilung. Die Disparität ist dagegen gemessen am Disparitätsmaß D_G von Gini von 0,15 auf 0,32 gestiegen. Wie man daran sieht, kann auch eine symmetrische Verteilung eine größere Disparität aufweisen, als eine linkssteile Verteilung.

Entgegen einer verbreiteten Vorstellung ist also "Ungleichheit" nicht adäquat operationalisiert mit der "Schiefe" (wobei meist an Linkssteilheit gedacht wird) im Sinne der deskriptiven Statistik.

5. Dominanzmaße: Entdeckung oligopolistischer Strukturen

Neben dem Herfindahl-Index K_H spielt der im folgenden dargestellte Linda-Index (nach Remo Linda, der ihn 1976 vorschlug) in den Arbeiten der Monopolkommission und der amtlichen Statistik eine große Rolle. Mit dem Linda-Index, bzw. genauer dem System von Linda-Indizes L_k (k=2,3,...,n-1), will man oligopolistische Strukturen erkennen, d.h. feststellen wieviele Unternehmen (z.B. die größten vier), die sog. "Oligopolgruppe", die übrigen Unternehmen (Mitläufer, Umfeld) "dominieren". Man spricht deshalb auch von einem Dominanzmaß. Neben dem Linda-Index sind auch andere Dominanzmaße vorgeschlagen worden, auf die hier aus Platzgründen nicht eingegangen werden kann. Es sollte jedoch vorweg erläutert werden:

1. die Unterscheidung zwischen summarischen und diskreten Konzentrationsmaßen und
2. das Konzept der Dominanz.

zu 1:

Summarische Konzentrationsmaße (wie K_H) beschreiben einen Markt mit n Teilnehmern (z.B. Anbietern) mit einer einzigen Maßzahl, ohne zwischen den n Unternehmen zu differenzieren. Diskrete Maße erlauben es dagegen auch Teile des Marktes (als m < n Unternehmen) zu betrachten und so oligopolistische Strukturen zu identifizieren.

zu 2:

Auch für Dominanzmaße gibt es eine Axiomatik, auf die hier jedoch nicht eingegangen werden kann. Es ist aber z.B. unmittelbar einsichtig, dass die Vorherrschaft der Oligopolgruppe umso größer ist, je

- kleiner (je weniger Unternehmen umfassend) die Gruppe ist
- geringer die Disparität zwischen den Unternehmen der Gruppe ist
- größer ihr kumulierter Marktanteil ist und je
- größer und weniger differenziert das Umfeld ist,

denn das sind Bedingungen für eine Interessensymmetrie und damit für gemeinsames Handeln der dominierenden Unternehmen. Daraus folgt auch, dass Dominanzmaße, die auf diese Bedingungen reagieren sollten, Eigenschaften von Konzentrations- und Disparitätsmaßen miteinander kombinieren.

Def. 6.10: Linda-Indizes

Das sog. oligopolistische Gleichgewicht der i größten Unternehmen EO_i (oder $EO_{i,k}$, weil es von i und k abhängt) ist ein Verhältnis von Marktanteilen (C_i, C_k sind Konzentrationsraten [Def. 6.3])

$$(6.33)\quad EO_i = \frac{C_i}{i} : \frac{C_k - C_i}{k-i} = \frac{C_i(k-i)}{i(C_k - C_i)}$$

mit $1 \leq i \leq k-1$ und $k = 2,3,...,n-1$ bei n Einheiten (z.B. Anbietern auf dem Markt). Der erste Klammerausdruck ist der obere, der zweite der untere Mittelwert. Geometrisch veranschaulicht ist der oligopolistische Kern die Stelle, an der die Konzentrationskurve den stärksten Knick hat. Der (k-te) Linda-Index L_k ist definiert als Mittelwert der durch k dividierten Größen EO_i

$$(6.34)\quad L_k = \frac{1}{k-1}\sum_{i=1}^{k-1} EO_i / k = k(k-1)\sum_{i=1}^{k-1} EO_i \quad .$$

Bemerkungen zu Def. 6.10:

1. Bei gleichen Marktanteilen von k Unternehmen ist $C_i = i/k$ für alle i so dass $EO_i = 1$ und $L_k = 1/k$. Nach oben ist L_k nicht beschränkt, so dass gilt $1/k \le L_k < \infty$.

2. Sind die Marktanteile nicht gleich, so ist L_k eine Funktion von k mit mindestens einem Minimum. Der *Oligopolkern* (die k^* größten und dominierenden Unternehmen) ist der Wert $k = k^*$, bei dem L_k zum ersten Mal minimal ist. Das Minimum ist umso ausgeprägter, je größer die Disparität unter den k^* Kernunternehmen ist.

3. Man kann den unter Nr. 1 genannten Mangel beheben, indem man nur Unternehmen mit einem Marktanteil von mindestens δ in die Betrachtung einbezieht und L_k mit einem von δ abhängigen Faktor auf den Wertebereich $1/k \le L_k^* \le 1$ für einen modifizierten Index L_k^* normiert.

Beispiel 6.8:

Die Berechnung des Linda-Index soll demonstriert werden bei einem fiktiven Markt mit acht Unternehmen und den Marktanteilen c_i und den kumulierten Marktanteilen C_i:

	i=1	i=2	i=3	i=4	i=5	i=6	i=7	i=8
c_i	0,40	0,20	0,15	0,10	0,08	0,04	0,02	0,01
C_i	0,40	0,60	0,75	0,85	0,93	0,97	0,99	1,00

Man stelle mit den Linda-Indizes L_k die Anzahl k^* der Unternehmen fest, die eine Oligopolgruppe bilden.

Lösung 6.8:

Die Berechnung von L_k soll für k = 2 und k = 3 ausführlich gezeigt werden und für die höheren Werte für k sind die Ergebnisse in einer Tabelle zusammengefaßt.

k = 2 (man erhält $L_2 = 1$)

$EO_1 = C_1/1 : (C_2-C_1)/(2-1) = 0,4/0,2 = 2$

dann ist $EO_1/k = 2/2 = 1 = L_2$

k = 3 ($L_3 = 0,7143$)

$EO_1 = C_1:(C_3-C_1)/(3-1) = 0,4: (0,35/2) = 0,4/0,175 = 2,2857$

$EO_1/k = 2,2857/3 = 0,7619$

$EO_2 = C_2/2 : (C_3-C_2)/(3-2) = 0,6/2 : 0,15/1 = 2; EO_2/k = 2/3$

und damit ist $L_3 = (EO_1/3 + EO_2/3)/(3-1) = (0{,}7619 + 2/3)/2$
$L_3 = 0{,}7143$

Der kleinste Wert von L_k ist, wie die folgende Tabelle zeigt, erreicht bei k=5 (denn $L_5 = 0{,}55901$), so dass man sagen kann, die fünf größten Unternehmen bilden eine Oligopolgruppe.

	Werte für EO_i bei							Linda-
k	i=1	i=2	i=3	i=4	i=5	i=6	i=7	IndexL_k
4	2,67	2,4*)	2,5*)					0,65306
5	3,02	2,73	2,78	2,66				0,55901
6	3,51	3,24	3,41	3,54	4,65			0,61176
7	4,07	3,85	4,17	4,55	6,2*)	8,08		0,73613
8	4,67	4,5*)	5*)	5,67	7,97	11,11	14,14	0,94748

*) Auf Nachkommastellen wurde verzichtet, wenn nicht gerundet wurde. Die Angaben sind zu ungenau um L_k zu berechnen. Sie sollen nur eine Kontrolle ermöglichen, wenn man die Indizes "zu Fuß" (also Schritt für Schritt nach den angegebenen Formeln) berechnen möchte, was für Übungszwecke sehr zu empfehlen ist.

6. Zur Vertiefung des Verständnisses der Lorenzkurve

a) Momentverteilung und Häufigkeitsverteilung

Eine (diskrete) Häufigkeitsverteilung ist als Folge der Wertetupel (x_i,h_i) definiert. Treten die Merkmalsanteile q_i an die Stelle der relativen Häufigkeiten h_i, so spricht man von der Momentverteilung. Sie spielt nicht nur bei der Disparitätsmessung, sondern auch für die Betrachtung von Mittelwerten eine gewisse Rolle.

Def. 6.11: Momentverteilung, Scheidewert und Schwerster Wert

Die Wertetupel (x_i,q_i) heißen **Momentverteilung** (oder: Wertverteilung). Dabei sind die Größen q_i die mit

(6.2a) $q_i = \dfrac{x_{(i)}}{\Sigma x_{(i)}}$ bei Einzelwerten, bzw.

(6.2b) $q_i = \dfrac{n_i}{\Sigma x_i n_i}$ bei Häufigkeitsverteilungen

definierten Merkmalsanteile.

Der Median der Momentverteilung heißt **Scheidewert** (oder auch Medial) und der Modus der Momentverteilung ist der **Schwerste Wert** ($\bar{x}_T$).

Bemerkungen zu Def. 6.11:

1. Zwischen den Merkmalsanteilen q_i und den relativen Häufigkeiten h_i besteht allgemein die folgende Beziehung:

   (6.23) $\frac{q_i}{h_i} = \frac{x_i}{\bar{x}}$

   Bei Einzelwerten ist $x_{(i)}$ anstelle von x_i zu setzen und es gilt für alle i, dass $h_i = 1/n$. Dabei ist q_i/h_i die Steigung der Lorenzkurve zwischen den Punkten (H_{i-1}, Q_{i-1}) und (H_i, Q_i). Daraus folgt:

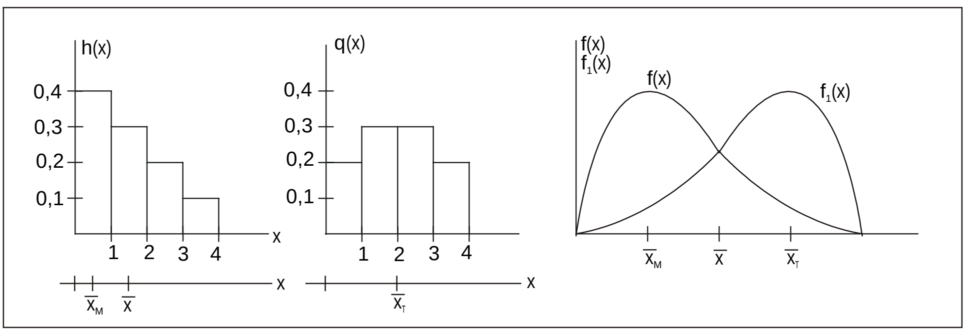
   wenn $x_i < \bar{x}$, dann $q_i < h_i$
   wenn $x_i = \bar{x}$, dann $q_i = h_i$
   wenn $x_i > \bar{x}$, dann $q_i > h_i$

   Entsprechend ist die Steigung der Lorenzkurve kleiner als 1, genau 1 oder größer als 1.

2. Für den Schwersten Wert $\bar{x}_T$ gilt im Verhältnis zum Modus (dichtester Wert) $\bar{x}_M$ deshalb $\bar{x}_T \geq \bar{x}_M$. Dies ergibt sich aus Bem. Nr. 1 wonach die Werte h_i ihr Maximum vor (links von) den Werten q_i erreicht haben müssen sowie aus dem in Abb. 6.10 dargestellten typischen Verlauf von Momentverteilung und Häufigkeitsverteilung.

3. Neben Modus und Median der Momentverteilung spielt auch das arithmetische Mittel K_N dieser Verteilung eine gewisse Rolle.

Abb. 6.10: Momentdichte $f_1(x)$ und Dichte f(x)



Es wurde als Disparitätsmaß vorgeschlagen (Niehaus 1955). Es gilt wegen Gl. 6.23 und in Verbindung mit dem quadratischen Mittel $\bar{x}_Q$ sowie dem Verschiebungssatz der Varianz (V = Variationskoeffizient)

$$(6.35) \qquad K_N = \Sigma x_i q_i = \frac{\Sigma x_i^2 h_i}{\bar{x}} = \frac{\bar{x}_Q^2}{\bar{x}} = \frac{s^2+\bar{x}^2}{\bar{x}} = \bar{x} + sV$$
$$= \bar{x}(V^2+1) \geq \bar{x} \;.$$

Das arithmetische Mittel der Momentverteilung kann also nicht kleiner als das arithmetische Mittel (der Häufigkeitsverteilung) sein. Nach Gl. 6.10 ergibt sich der folgende Zusammenhang zum Herfindahl-Index K_H

$$(6.36) \qquad K_H = \frac{K_N}{n\bar{x}} = \frac{K_N}{\Sigma x_i n_i},$$

so dass K_H auch als ein durch die Merkmalssumme normiertes arithmetisches Mittel der Momentverteilung interpretiert werden kann.

4. Der Ausdruck

$$(6.37) \qquad G_V = \frac{1}{1+V^2} = \frac{\bar{x}}{K_N} = \frac{1}{nK_N} \;,$$

also das Verhältnis der Mittelwerte von Häufigkeits- und Momentverteilung, gilt als Gleichheitsmaß (vgl. auch Gl. 6.31a). G_V und K_H bilden ein Paar gleichmäßig normierter Maße (vgl. Abschn. 6d, Gl. 6.25a) und GV ist das zum normierten Quadrat des Variationskoeffizienten NV (gem. Gl. 6.30) gehörende Gleichheitsmaß.

Beispiel 6.9:

Man bestimme die Momentverteilung und deren Lageparameter für das Beispiel 6.6.

Lösung 6.9:

In der Lösung werden bereits Angaben gemacht, auf die erst an späterer Stelle (maximaler Nivellierungssatz) hingewiesen wird.

Häufigkeitsverteilung

x_i	h_i
1	0,4
2	0,3
3	0,2
4	0,1

Momentverteilung

x_i	q_i
1	0,2
2	0,3
3	0,3
4	0,2

Abstand Lorenzkurve-Gleichverteilungsgerade

x_i	H_i-Q_i
1	0,2
2	0,2
3	0,1
4	0,0

Modus $\bar{x}_M = 1$
Median = 1,333
$\bar{x} = 2$

Schwerster Wert $\bar{x}_T = 2{,}5$
Scheidewert (Median) = 2
$K_H = 2{,}5$.

Für die Häufigkeits- und Momentverteilung gilt also der typische Verlauf der Abb. 6.10. Der Abstand zwischen Gleichverteilungsgerade und Lorenzkurve ist maximal 2 an der Stelle $x = \bar{x}$ (bei H = 0,7 und Q = 0,5, also für die Gruppe, die genau die durchschnittliche Anzahl von Gänsen erhält, nämlich 2). Dieser Abstand ist der maximale Nivellierungssatz (vgl. Def. 6.13).
Weitere Kennzahlen und Zusammenhänge: Varianz der Häufigkeitsverteilung $s^2 = 1$, quadratisches Mittel $Q = \sqrt{5}$ (weshalb auch $K_N = Q^2/\bar{x} = 5/2 = 2{,}5$), Herfindahl-Index $K_H = K_N / n\bar{x} = 2{,}5/300 = 0{,}00833$. Man kann K_H auch aus Gl. 6.10 errechnen: $V^2 = 0{,}25$, so dass $K_H = 1{,}25/150$. Das Gleichheitsmaß G_V gem. Gl. 6.37 ist 0,8.

b) Stetige Lorenzkurve

Die folgende Betrachtung mag dazu beitragen, einige Zusammenhänge zu verdeutlichen, sie ist aber für die Anwendung auf empirische Daten nicht von Bedeutung, weil in diesem Falle eine stetige Variable stets in klassierter (und damit diskreter) Form vorliegt.
An dieser Stelle werden einige Begriffe und Symbole aus der Induktiven Statistik vorausgesetzt. Leser, die hiermit nicht vertraut sind, können den Exkurs getrost überschlagen.
Für die stetige Variable X sei f(x) die Dichtefunktion und

$$\mu = \int_a^b x f(x) dx \quad (a \le x \le b)$$

das arithmetische Mittel (das erste Anfangsmoment). Dann ist die erste Momentdichte gegeben mit

$$\boxed{(6.38) \quad f_1(x) = \frac{1}{\mu} x f(x)\ .}$$

und

(6.38a) $F_1(x) = \int_a^x f_1(u) du$ ist die Momentverteilung analog zur

Verteilungsfunktion $F(x) = \int_a^x f(u) du$.

Die Zusammenhänge zwischen den Dichtefunktionen f(x) und $f_1(x)$ sowie den Verteilungsfunktionen F(x) und $F_1(x)$ sind aufschlußreich für das Verständnis der Lorenzkurve. Hierzu die folgenden Bemerkungen:

1.) Offensichtlich ist $f_1(x)$ eine Dichtefunktion, da

$$\int_a^b f_1(x)dx = \frac{\mu}{\mu} = 1$$

2.) Wegen Gl. 6.38 gilt $\mu f_1(x) = xf(x)$ nach der Produktregel der Differentiation

$$\boxed{(6.39) \qquad \mu\, f_1'(x) = x\, f'(x) + f(x),}$$

so dass das Maximum der Momentdichte $f_1(x)$, also der schwerste Wert $\bar{x}_T$ im Bereich fallender Werte der Dichtefunktion liegt, denn aus $f_1'(x) = 0$ folgt $f'(x) < 0$. Deshalb ist, wie bereits gesagt, der schwerste Wert $\bar{x}_T$ nicht kleiner als der dichteste Wert $\bar{x}_M$. In Abb. 6.10 ist der typische Verlauf der beiden Dichtefunktionen wiedergegeben.

3.) Aus (6.38) folgt

$$\boxed{(6.40) \qquad F_1(x) = \frac{1}{\mu}\int_a^b uf(u)du = \mu_1 / \mu}$$

wobei μ_1 das bis zur Stelle x erreichte arithmetische Mittel ist und $\mu_1 \leq \mu$ gilt. Daraus folgt, dass $F_1(x) \leq F(x)$, analog zur bekannten Beziehung $Q_i \leq H_i$ im diskreten Fall, gilt. Abgesehen von der Rechteckverteilung $F(x) = x(b-a)^{-1}$ gilt die Gleichheit von $F_1(x)$ und F(x) nur für die Punkte x = a und x = b.

Wir sind nun in der Lage, die Lorenzkurve für eine stetige Variable X zu definieren und diese Definition anhand eines Beispiels (Beispiel 6.10) zu erläutern.
Diese stetige Darstellung ist vor allem deshalb von Interesse, weil so die allgemeinen Eigenschaften von Disparitätsmaßen besser herausgearbeitet werden können. In allen empirischen Anwendungen ist dagegen bei der Disparitätsmessung stets von klassierten Daten oder Einzelbeobachtungen auszugehen, so dass die Lorenzkurve ein Polygonzug darstellt, den man als Annäherung an den "wahren" kontinuierlichen Kurvenverlauf auffassen kann.

Def. 6.12: Lorenzkurve im stetigen Fall

Die Funktion L(F), die im F(x), $F_1(x)$-Koordinatensystem alle Punkte mit den Koordinaten $L(F)=F_1(F)$ und F miteinander verbindet, heißt Lorenzkurve. Man erhält $F_1(F)$, indem man x in $F_1(x)$ ersetzt durch x = G(F), wobei G die inverse Verteilungsfunktion ist.

Zum Begriff der inversen Verteilungsfunktion (vgl. Def. 3.3,d):

Für einen Wert x des Merkmals X ist mit F = F(x) die kumulierte relative Häufigkeit gegeben. Man kann umgekehrt fragen, welchen x-Wert man erhält, wenn F einen bestimmten vorgegebenen Wert annehmen soll, etwa F = ½, dann ist der gesuchte x-Wert der Median $\tilde{x}_{0,5} = G(F=½)$.

Beispiel 6.10:

Mit diesem einfachen Beispiel soll die mit Definition 6.12 gegebene Vorschrift zur Herleitung der Lorenzkurve demonstriert werden. Gegeben sei die Dichtefunktion

$$f(x) = 3x^2/8 \qquad \text{für } 0 \le x \le 2.$$

Hieraus folgt $\mu = 1{,}5$ und $F(x) = 1/8\, x^3$.
Dann ist $f_1(x) = x^3/4$ und $F_1(x) = x^4/16$.
Löst man F(x) nach x auf, so erhält man $x = (8F)^{1/3}$. Dies ist die inverse Verteilungsfunktion.
Die Lorenzkurve lautet dann $L(F) = F_1(F) = (1/16)(8F)^{4/3} = F^{4/3}$.
Die Steigung der Lorenzkurve im Beispiel 6.11 ist

$$L'(F) = \frac{dL}{dF} = \frac{f_1(x)}{f(x)} = \frac{x^3}{4} \cdot \frac{8}{3x^2} = \frac{x}{1{,}5} = \frac{x}{\mu} \ge 0.$$

Ferner ist $L''(F) = [\mu f(x)]^{-1} > 0$,

d.h. die Steigung der Lorenzkurve ist nicht-negativ und monoton zunehmend.
Für das Disparitätsmaß von Gini D_G erhält man:

$$(6.41) \quad D_G = 1 - 2\int_0^1 L(F)dF$$

Man kann zeigen, dass für Ginis D_G auch gilt

$$(6.42) \quad D_G = \frac{1}{\mu}\int_a^b (F-F^2)dx$$

Im Beispiel 6.10 ist $D_G = 1/7$. Man verifiziert im obigen Beispiel übrigens leicht, dass gilt $L(F) \le F$, d.h. die Lorenzkurve ist konvex und dass für die Momentverteilung das arithmetische Mittel $\int_0^2 x f_1(x)dx = 1{,}6$ beträgt und größer ist als $\mu = 1{,}5$.

c) Schutz-Koeffizient (Maximaler Nivellierungssatz)

Im folgenden wird gezeigt, dass der senkrechte Abstand zwischen der Gleichverteilungsgeraden und der Lorenzkurve maximal ist an der Stelle x = μ und als Disparitätsmaß zu interpretieren ist. Es ist der Schutz-Koeffizient. Er wird auch maximaler Nivellierungssatz (von Lindahl) oder längste Lorenzkurvensehne genannt.

Def. 6.13: Schutz-Koeffizent

Der Schutz-Koeffizient oder maximale Nivellierungssatz von Lindahl ϕ^* ist mit

$$(6.43)\quad \phi^* = \frac{1}{\mu} F_\mu(\mu - \mu_1) = F_\mu\left(1 - \frac{\mu_1}{\mu}\right)$$

gegeben. Hierbei ist μ_1 der durchschnittliche Merkmalsbetrag derjenigen Teilgesamtheit, die einen unterdurchschnittlichen Merkmalsbetrag hat, also solange $x \le \mu$ ist.
Es ist der Anteil am Gesamtmerkmalsbetrag, der von der Teilgesamtheit mit überdurchschnittlichem Merkmalsbetrag an diejenige mit unterdurchschnittlichem Betrag umverteilt werden müßte, um zur Gleichverteilung (egalitären Verteilung) zu gelangen.

Im Beispiel 6.10 beträgt diese Größe $F(x=\mu) - F_1(x=\mu) = 27/64 - 81/256 = 0{,}1055$. Es gelten die folgenden Zusammenhänge:

1. Die senkrechte Distanz zwischen der Gleichverteilungsgeraden und der Lorenzkurve $\phi(F) = F - L(F)$ ist maximal an der Stelle $x = \mu$, an der die Steigung der Lorenzkurve 1 beträgt und sie ist dann:
   $\phi^* = \max \phi(F) = F_\mu - L(F_\mu) \quad \text{mit} \quad F_\mu = F(x=\mu)$
2. Die maximale Distanz beträgt ϕ^* gem. Gl. 6.43.

<u>Beweis</u>

1. Aus $\phi'(F) = 1 - L'(F) = 1 - x/\mu = (\mu-x)/\mu$ und $\phi''(F) = -L''(F) = -\mu f(x)^{-1} < 0$ folgt wegen $\phi'(F) = 0$, dass das Maximum von ϕ an der Stelle $x = \mu$ gegeben ist.

2. Es gilt (am Beispiel der Einkommensverteilung) für die Durchschnittseinkommen der unterdurchschnittlich (μ_1) bzw. der überdurchschnittlich (μ_2) verdienenden Merkmalsträger
   $$\mu_1 = F_\mu^{-1} \int_a^{x=\mu} x f(x)dx \quad \text{und} \quad \mu_2 = (1-F_\mu)^{-1} \int_{x=\mu}^{b} x f(x)dx \quad \text{mit } F_\mu = \int_a^{x=\mu} f(x)\,dx$$

   Wegen $\mu = F_\mu \mu_1 + (1-F_\mu)\mu_2$ gilt
   $$\phi^* = F_\mu - L(F_\mu) = F_\mu - F_1(F_\mu) = F_\mu - \mu^{-1} \int_a^{x=\mu} x f(x)\,dx = F_\mu - \mu^{-1} F_\mu \mu_1.$$

3. Die Interpretation von ϕ^* als "Nivellierungssatz" wird aus der folgenden Überlegung deutlich:
   Wird der Anteil ϕ^* von den überdurchschnittlich verdienenden Einheiten, deren bisheriger Einkommensanteil $(1 - F_\mu)\mu_2 / \mu$ beträgt, an die

unterdurchschnittlich verdienenden Einheiten übertragen, deren Anteil bisher $F_\mu \mu_1/\mu$ beträgt, so wird die
egalitäre Verteilung entstehen. Denn dann gilt für die Einkommensanteile der bisher überdurchschnittlich Verdienenden

$$\mu^{-1}(1-F_\mu)\mu_2 - F_\mu(1 - \mu_1/\mu) = 1-F_\mu$$

und der bisher unterdurchschnittlich Verdienenden

$$\mu^{-1}F_\mu\mu_1 + F_\mu(1 - \mu_1/\mu) = F_\mu.$$

Die Anteile an dem Merkmalsbetrag entsprechen jetzt genau den Anteilen der beiden Gruppen an der Gesamtzahl der Merkmalsträger, d.h. es ist F_1=F (egalitäre Verteilung). Zu diesen Zusammenhängen vgl. auch Bsp. 6.12.

d) Gleichmäßig normierte Maße

Def. 6.14: gleichmäßig normierte Maße

Maße der Gleichheit G (vgl. Def. 6.2) und der Konzentration K, die sich bezüglich der Axiome K4 und K5 gleichmäßig ihrer unteren Grenze nähern und für die gilt

$$\boxed{(6.25a)\quad K = \frac{1}{n(1-D)} = \frac{1}{nG} \quad \text{und damit} \quad G = \frac{1}{nK}}$$

heißen gleichmäßig normierte Maße (nach Jöhnk).

Bemerkungen zu Def. 6.14:

1. Zwei Paare gleichmäßig normierter Maße sind beispielsweise
   1. Gini-Koeffizient D_G und Rosenbluth-Index K_R
   2. normierter Variationskoeffizient NV = $V^2 / (1+V^2)$ und Herfindahl-Index K_H denn es gilt

Disparitätsmaß	Gleichheitsmaß	Konzentrationsmaß	vgl. Gleichung
Gini D_G	$1-D_G$	$K_R = [n(1- D_G)]^{-1}$	Gl. 6.25
norm.Var. NV	$G_V = (V^2+1)^{-1}$	$K_H = (V^2+1)/n$	Gl. 6.10, 6.37

2. Die Gleichung KnG = 1 läßt sich umformen in

$$\log(K) + \log(n) + \log(G) = 0.$$

Dies erlaubt es, zwei Konzentrationszustände (K_1,K_2) in bezug auf den Anzahleffekt ($n_1 < n_2$ oder $n_1 > n_2$) und den Disparitätseffekt Unterschiedlichkeit von D_1 und D_2,
bzw. G_1 und G_2) zu vergleichen:

$$(6.44)\qquad \log(K_2/K_1) = \log(n_1/n_2) + \log(G_1/G_2)\ .$$

Der erste Summand ist der Anzahleffekt (AE, absolute Komponente), der zweite der Verteilungs- oder Disparitätseffekt (DE, relative Komponente). Diese Zerlegung wird an dem folgenden Beispiel 6.11 demonstriert.

3. Aus der Zerlegung gem. Gl. 6.44 folgt, dass Veränderung der (absoluten) Konzentration sowohl von einem Anzahl-, als auch von einem Disparitätseffekt herrühren kann. Eindeutige Aussagen sind nur möglich, wenn sich n und G gleichsinnig verändern, d.h. wenn:

   n steigt und G steigt (D sinkt): dann sinkt K
   n sinkt und G sinkt (D steigt): dann steigt K .

Beispiel 6.11:

Ausgangsverteilung $n_1 = 2$ nach Nullergänzung $n_2 = 3$

Verteilung 1

Einheit	Anteil
1	0,8
2	0,2

Verteilung 2

Einheit	Anteil
1	0,8
2	0,2
3	0

a) Zeigen Sie die Wirkung auf das Konzentrationsmaß $K = K_R$ (Rosenbluth) und das Gleichheitsmaß $G = 1 - D_G$ sowie die Zerlegung des Konzentrationseffektes in den Anzahl- und Disparitätseffekt.

b) Das Beispiel ist dahingehend zu variieren, dass eine Verteilung Nr. 3 mit $n_3 = 4$ Einheiten aus der Ausgangsverteilung (Verteilung 1) durch Halbierung jeder Einheit entsteht.

Lösung 6.11:

Das Beispiel ist so konstruiert, dass der Disparitätseffekt (Teil a, Axiom K5) und der Anzahleffekt (Teil b, Axiom K4) jeweils in reiner Form zum Tragen kommen, was sich dann auch an der Zerlegungsformel (Gl. 6.44) zeigen muss.

a) $K_1 = 1/1{,}4 = 0{,}7143$ $G_1 = 0{,}7$ (da $D_G = 0{,}3$).
(für den Herfindahl-Index erhielte man $K_H = 0{,}68$ und für G_V den Wert $G_V = 1/1{,}36 = 0{,}7353$)
Die Nullergänzung verändert die Konzentration nicht (Axiom K5) $K_2 = K_1$, während die Disparität zunehmen muss. Es gilt $G_2 = 0{,}4667 = (2/3)G_1$.
Für die obige Zerlegungsformel (Gl. 6.42) erhält man

$$\log(1) = \log(2/3) + \log(3/2) = -0{,}1761 + 0{,}1761 = 0.$$

Anzahl- und Disparitätseffekt sind betragsmäßig gleich und heben sich auf.

b) Die neue Verteilung hat folgende Gestalt:

Einheit	1	2	3	4
Anteil	0,4	0,4	0,1	0,1

Es gilt dann $K_3 = 1/2{,}8 = ½K_1$ und $G_3 = G_1$. Dann ist

$\log(1/2) = \log(4/2) + \log(1) = -0{,}30103$.

Fazit: Die Halbierung der Konzentration ist ausschließlich auf den Anzahleffekt zurückzuführen, was ja auch bei dem gem. Axiom K4 (reiner Anzahleffekt) konstruierten Beispiel zu erwarten war.

Beispiel 6.12:

Gegeben sei das stetige Konzentrationsmerkmal X (z.B. das Einkommen), dessen Dichte lautet

$$f(x) = \begin{cases} 0{,}4 - 0{,}08x & \text{für } 0 \leq x \leq 5 \\ 0 & \text{sonst} \end{cases}$$

Man berechne die Momentdichte (und deren Maßzahlen, wie z.B. Mittelwerte), die Lorenzkurve und die anderen in diesem Exkurs dargestellten Größen.

Lösung 6.12:

- Verteilung von x:
  $f(x) = 0{,}4 - 0{,}08x$ (Dreiecksverteilung) $\mu = 5/3 = 1{,}667$
  der Median ist an der Stelle $x = 1{,}4645$ denn $F(x = 1{,}4645) = ½$,
  $F(x) = 0{,}4x - 0{,}04x^2 = F$,
  inverse Verteilungsfunktion $x = G(F) = 5 - 5\sqrt{(1-F)}$.
- Momentdichte $f_1(x)$ und Momentverteilung $F_1(x)$
  $f_1(x) = (x/\mu)f(x) = 0{,}24x - 0{,}048x^2$ $F_1(x) = 0{,}12x^2 - 0{,}016x^3$
  arithmetisches Mittel der Momentdichte $K_N = \int_0^5 x f_1(x)dx = 2{,}5$;
  Median der Momentdichte (= Scheidewert) 1,675 [Achtung: die Bestimmung des Scheidewerts verlangt wegen $F_1(x)$ die Lösung einer kubischen Gleichung]; der Modus der Momentdichte (d.h. der Schwerste Wert $\bar{x}_T$) ist $\bar{x}_T = 2{,}5$.
- Lorenzkurve
  $L(F) = F_1(F)$ wobei in $F_1(x)$ für x die inverse Verteilungsfunktion einzusetzen ist. Man erhält dann: $L(F) = 2(1-F)^{3/2} + 3F - 2$.
  Die Ableitung dieser Funktion ergibt die Steigung der Lorenzkurve. Es gilt:
  $dL(F)/dF = 3 - 3(1-F)^{½}$. Man erkennt sofort, dass dieser Ausdruck gleich ist der Relation x/μ, denn aus der inversen Verteilungsfunktion folgt
  $x = 5 - 5(1-F)^{½}$ und für μ gilt $\mu = 5/3$.
- Ginis-Disparitätsmaß D_G
  Man errechnet D_G aus Gl. 6.41. Für das Integral über L(F) in den Grenzen von 0 bis 1 erhält man (Integrationskonstante C) den Ausdruck:
  $-(4/5)(1-F)^{5/2} + 1{,}5F^2 - 2F + C$. Mit den Grenzen F=0 und F=1 ergibt sich dann für D_G der Zahlenwert $D_G = 0{,}4$.
  Das Disparitätsmaß G_V nimmt als das Verhältnis der Mittelwerte von Dichte f(x) und Momentdichte $f_1(x)$ den Wert $G_V = 2/3$ an.
- Abstand zwischen Gleichverteilungsgeraden und Lorenzkurve F - L(F)

Er beträgt $F - L(F) = 2(1-F) - 2(1-F)^{3/2}$. Dieser Abstand ist maximal an der Stelle F = 5/9 und er beträgt dort 8/27. Man erhält den Wert 5/9 auch, wenn man in F(x) für x den Wert x = μ = 5/3 einsetzt, d.h. $F_\mu = 5/9$. Das verifiziert den Satz, dass der Abstand zwischen der Gleichverteilungsgeraden und der Lorenzkurve maximal an der Stelle x = μ ist.

- Maximaler Nivellierungssatz (Schutzkoeffizient) und Interpretation hiervon:

  Der maximale Nivellierungssatz nach Lindahl ist damit 8/27. Das Einkommen der unterdurchschnittlich verdienenden Einheiten μ_1 und der überdurchschnittlich verdienenden Einheiten μ_2 beträgt:

  $$\mu_1 = \frac{1}{F_\mu}\int_0^\mu x f(x)dx = (9/5)(35/81)=7/9= 0{,}7778.$$ Entsprechend ist

  $$\mu_2 = \frac{1}{1 - F_\mu}\int_\mu^5 x f(x)dx = (9/4)(100/81) = 25/9 = 2{,}7778.$$

  Mit μ = 15/9 = 5/3 = 1,6667 ist also $\mu_1 < \mu < \mu_2$.

  Die Anteile an der Gesamtheit der Merkmalsträger betragen für die
  unterdurchschnittlich Verdienenden $F_\mu = 5/9$
  überdurchschnittlich Verdienenden $1 - F_\mu = 4/9$.

  Die Anteile am Gesamtmerkmalsbetrag sind für die
  unterdurchschnittlich Verdienenden $F_{1\mu} = F_1(x=\mu) = 7/27$
  überdurchschnittlich Verdienenden $1 - F_{1\mu} = 20/27$.

  Der maximale Nivellierungssatz ist 8/27. Wird der Anteil 8/27 den unterdurchschnittlich Verdienenden zugeschlagen, so ist ihr Anteil dann
  $F_{1\mu} + 5/9 = 7/27 + 8/27 = 15/27 = F_1(x=\mu) = 5/9$.

  Entsprechend verbleibt bei den überdurchschnittlich Verdienenden ein Anteil von 20/27-8/27=4/9, wie es ihrem Anteil an der Bevölkerung entspricht.

- Noch zur Interpretation des Gini-Koeffizienten

  Würde man nur die beiden Gruppen unter- und überdurchschnittlich Verdienende unterscheiden und mit diesen beiden Klassen eine Lorenzkurve durch lineare Verbindungen der Punkte (0,0),(5/9,7/27) und (1,1) zeichnen, so würde D_G nach Gl. 6.29a den Wert 5/9 - 7/27 = 8/27 (= maximaler Nivellierungssatz) annehmen. Man sieht, dass die Lorenzkurve als Polygonzug nur eine Näherung an die "wahre" stetige Lorenzkurve sein kann und dass im ersten Fall D_G kleiner ist als im zweiten (denn 8/27 = 0,2963 < 0,4), so dass generell gilt:

  $$\boxed{(6.44)\ \ \phi^* \le D_G \ .}$$