4 Konzentrationsmessung

4.0 Vorbemerkungen

Konzentration: Ausmaß der Ballung von großen Anteilen an der gesamten Merkmalssumme auf wenige Einheiten (Frage zum Beispiel: Welchen Anteil am Gesamtvermögen haben die Reichsten?)

Nullkonzentration: Alle haben exakt denselben Anteil, d.h. es tritt nur eine Merkmalsausprägung auf.

Grundlegende Unterscheidung:

- geringe **Anzahl** von Einheiten: *absolute* Konzentration; v.a. in der Ökonomie ((Angebots-) Oligopol)
- geringer **Anteil** von Einheiten: *relative* Konzentration; v.a. Einkommens- und Bodenverteilung, Armutsmessung

Jedes absolute Konzentrationsmaß sagt auch etwas über relative Konzentration aus und umgekehrt.

2 Arten der Beschreibung:

- durch eine **Funktion**: v.a. Lorenz-Kurve (für relative Konzentrationen)
- durch **Maßzahlen**: Gini-Koeffizient (für relative Konzentration), mehrere gebräuchliche Maße für absolute Konzentrationen

Durchgängige Annahmen

- X sei ein verhältnisskaliertes Merkmal (mit Urliste $x_1, \ldots, x_n$)

- Die Merkmalsausprägungen $a_1, \ldots, a_k$ seien der Größe nach geordnet

$$a_1 < \ldots < a_k.$$

  $h_1, \ldots, h_k$ und $f_1, \ldots, f_k$ seien die zugehörigen absoluten bzw. relativen Häufigkeiten. Die empirische Verteilungsfunktion werde mit $F_X(\cdot)$ oder $F(\cdot)$ bezeichnet.

- X nimmt nur positive Werte an, $x_i \geq 0$, für alle $i = 1, \ldots, n$ und $\sum_{i=1}^{n} x_i > 0$ (d.h mindestens ein Wert ist von Null verschieden)

- Betrachtet werden die der Größe nach geordneten Daten: $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$

  Achtung: Die Klammern im Index werden in der Literatur oft weggelassen, wenn vorausgesetzt wird, dass die Daten schon geordnet sind (z.B. in Fahrmeir et al., 2009)

Warum überhaupt Konzentrationsmaße?

- Wenn alle gleiches Vermögen haben, dann tritt nur eine einzige Merkmalsausprägung a_1 auf:
  - Dieser Wert ist zugleich arithmetisches Mittel, Median und Modus
  - Die Varianz bzw. der Variationskoeffizient ist Null

- Beispiel: 8 Personen verdienen netto jeweils 2400 Euro und zwei 6400 Euro („8 Mittelverdiener, 2 Vielverdiener“)

j	a_j	h_j	f_j	$a_j f_j$	$h_j \cdot (a_j - \bar{x})^2$
1	2400	8	0.8	1920	$8 \cdot (-800)^2 = 5120000$
2	6400	2	0.2	1280	$2 \cdot (3200)^2 = 20480000$
Summe		10	1.0	$\bar{x}$=3200	$\tilde{s}$=1600

- Wir betrachten nun den Fall, wo zwei Personen 0 Euro und 8 Personen jeweils 4000 Euro haben („2 Arme, 8 Gutverdiener“)

  Transformation: Y= 6400-X, oder anders ausgedrückt: betrachte y = g(x):= 6400-x

$$x_i = 2400 \quad \Longrightarrow \quad g(x_i) = y_i = 4000$$
$$x_i = 6400 \quad \Longrightarrow \quad g(x_i) = y_i = 0$$

Damit gilt (Transformationsregeln):

$$\bar{y} = 6400 - \bar{x} = 3200 = \bar{x}$$
$$\tilde{s}_Y = |-1|\tilde{s}_X = \tilde{s}_X$$

Y hat gleiche Varianz und gleichen Variationskoeffizienten wie X, aber das „Streuverhalten“ ist in diesen Situationen völlig unterschiedlich.

$\Longrightarrow$ Varianz bzw. Variationskoeffizienten genügt nicht als (alleiniges) Maß für (relative) Konzentration

Bem. 4.1. (Zum Ungenügen der Varianz und des Variationskoeffizienten als (alleiniges) Maß für (relative) Konzentration)

- Die üblichen Streuungsmaße messen die Variabilität um das Zentrum (arithmetisches Mittel, Median) symmetrisch

- Überschreiten und Unterschreiten des Zentrums wird gleich gewichtet

- Streuungsmaße geben keine Auskunft darüber, wie sich die Gesamtsumme (etwa das Vermögen) unter den einzelnen Einheiten verteilt, d.h. die Konzentration, also z.B. die Frage welchen Anteil am Gesamtvermögen haben die Reichsten?

4.1 Relative Konzentration

4.1.1 Die Lorenz-Kurve

Definition 4.2.

Die stückweise lineare Kurve durch die Punkte

$$(0,0),\ \ (u_1,v_1),\ \ (u_2,v_2),\ \ \ldots,\ \ (u_n,v_n)=(1,1)$$

mit

$$u_j := \frac{j}{n} \qquad \text{und} \qquad v_j := \frac{\sum_{i=1}^{j} x_{(i)}}{\sum_{i=1}^{n} x_i} = \frac{\sum_{i=1}^{j} x_{(i)}}{\sum_{i=1}^{n} x_{(i)}} \qquad \text{für } j=1,\ldots,n$$

heißt Lorenz-Kurve.

- u_j ist der Anteil der j kleinsten Merkmalsträger

- v_j der anteilige Beitrag dieser Einheiten zur Gesamtsumme

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_{(i)}$$

(„kumulierte relative Merkmalssumme“)

- Punkte oberhalb der Winkelhalbierenden kommen nicht vor!

- Die Lorenz-Kurve muss monoton wachsend sein.

Bem. 4.3. [Zur Interpretation der Lorenz-Kurve]

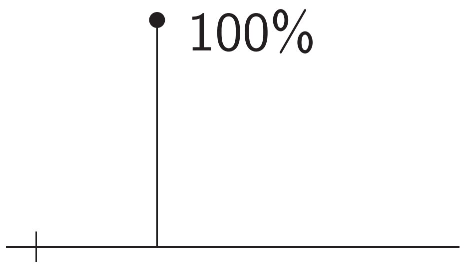
- Wie liegen die Punkte bei minimaler Konzentration (Konzentration=0)?

10%	kleinste	haben	10%
20%	"	"	20% usw.

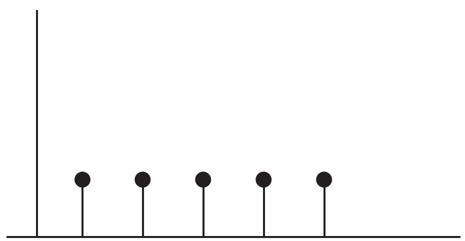
  also $u_j = v_j$ für alle j, d.h. auf der Winkelhalbierenden.

  Konzentration von 0: „gleiche Verteilung" in dem Sinne, dass jede Einheit denselben Anteil an der Gesamtsumme hat. Dies bedeutet $x_1 = x_2 = \ldots = x_n$ und damit $h_{j*} = n$ für eine bestimmte Ausprägung $a_{j*} = x_1$.

  Das ist zu unterscheiden von der so genannten (Häufigkeits-)Gleichverteilung (siehe v.a. Statistik II). Dort ist $h_1 = h_2 = \ldots = h_k$, d.h. jede Ausprägung a_j gleich häufig.



gleiche Werte, Konzentration 0

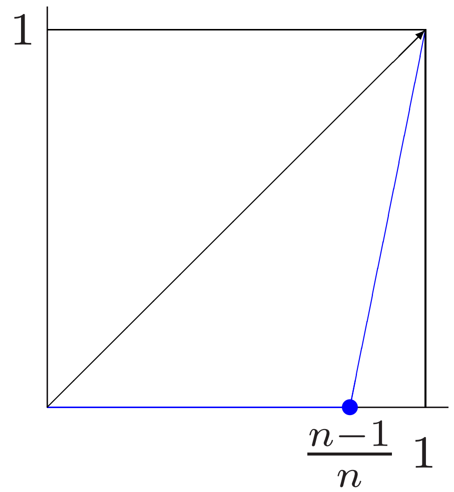


(Häufigkeits-)Gleichverteilung

- Extremfall: vollständige Konzentration

  $n-1$ Personen haben gar nichts

  n-te Person hat alles



- Je weiter die Kurve von der Winkelhalbierenden entfernt ist, umso stärker ist die Konzentration.

  $u_j = 0.9 \quad v_j = 0.0825$ Die 90% Ärmsten haben

  $u_j = 0.9 \quad v_j = 0.20$

Bem. 4.4.

a) Insbesondere bei größeren Datensätzen vereinfacht sich die Berechnung wesentlich, wenn man die relativen/absoluten Häufigkeiten $f_1, \ldots, f_k$ bzw. $h_1, \ldots, h_k$ der der Größe nach geordneten Merkmalsausprägungen $a_1 < a_2 < \ldots < a_k$ benutzt. Dann ist für $j = 1, \ldots, k$

$$u_j = \sum_{l=1}^{j} \frac{h_l}{n} = \sum_{l=1}^{j} f_l = F(a_j) \tag{4.1}$$

und

$$v_j = \frac{\sum\limits_{l=1}^{j} h_l \cdot a_l}{\sum\limits_{l=1}^{k} h_l \cdot a_l} = \frac{\sum\limits_{l=1}^{j} f_l \cdot a_l}{\sum\limits_{l=1}^{k} f_l \cdot a_l} \tag{4.2}$$

b) Ist bei klassierten Daten mit den Klassen $[c_0, c_1), [c_1, c_2), \ldots, [c_{k-1}, c_k]$ die Merkmalsverteilung in den Klassen nicht bekannt - und will man als Ergebnis dennoch eine einzelne Zahl -, so nimmt man als Approximation an, dass alle Ausprägungen in dieser Klasse auf die Klassenmitte $m_l = \frac{c_{l-1}+c_l}{2}$ fallen. Damit erhält man mit f_l und h_l, $l = 1, \ldots, k$ als relative bzw. absolute Klassenhäufigkeiten und $a_l = m_l,\ l = 1, \ldots, k:$

$$v_j = \frac{\sum_{l=1}^{j} h_l \cdot m_l}{\sum_{l=1}^{k} h_l \cdot m_l} = \frac{\sum_{l=1}^{j} f_l m_l}{\sum_{l=1}^{k} f_l m_l} \qquad (4.3)$$

c) Während normalerweise bei Lorenz-Kurven nur die Punkte $(0,0), (u_1, v_1), \ldots$ interpretierbar sind, sind bei klassierten Daten auch die linearen Zwischenstücke interpretierbar.

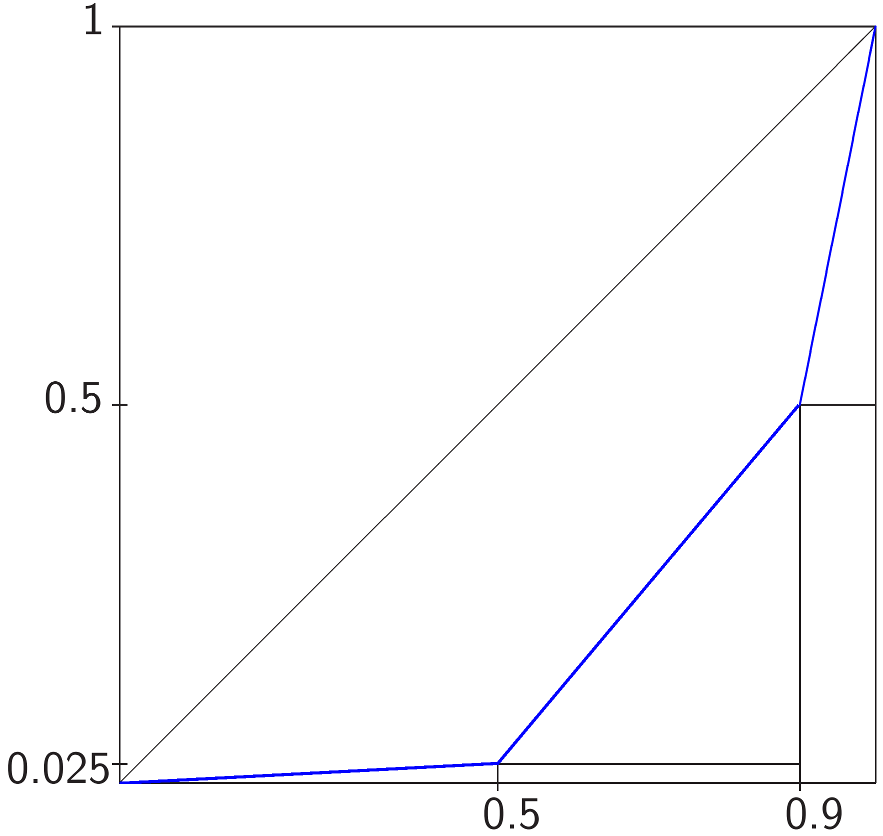
Bsp. 4.5.

Gegeben sei ein Land mit folgender klassierter Verteilung:

Klasse	1	2	3
$[c_{j-1}, c_j)$	$[0, 10)$	$[10, 227.5)$	$[227.5, \infty)$
	arm	mittel	reich
f_j	$f_1 = 0.5$	$f_2 = 0.4$	$f_3 = 0.1$

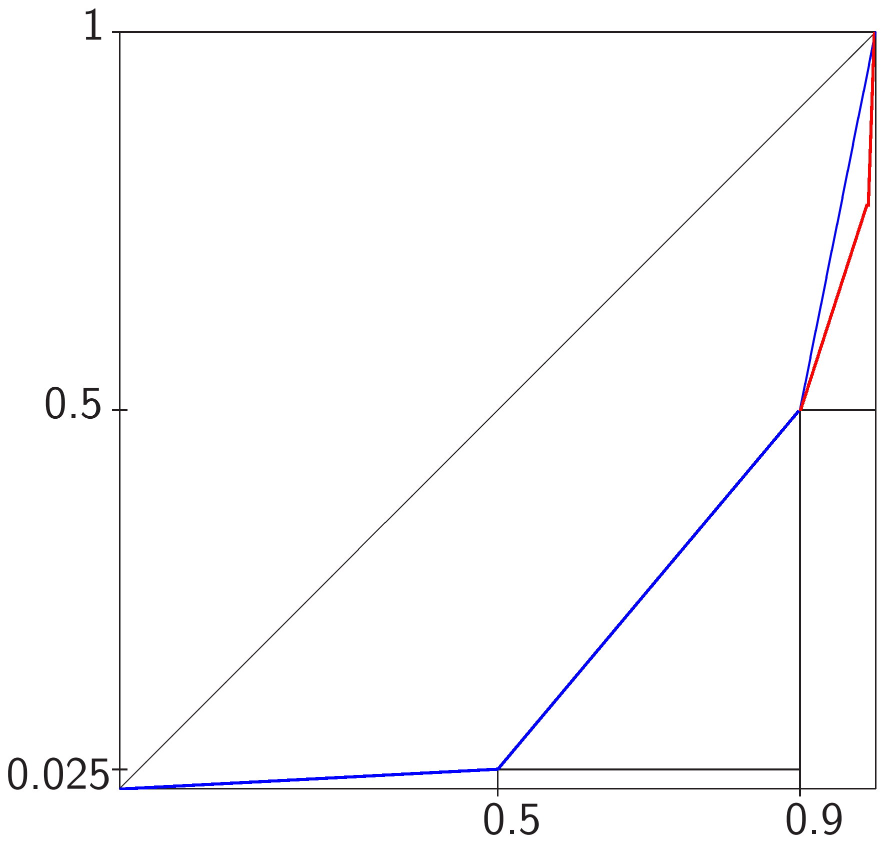
Man bestimme die Lorenz-Kurve (mit $m_3 := 500$).

j	m_j	f_j	$u_j = F(a_j)$	$f_j m_j$	$\sum_{l=1}^{j} f_l m_l$	v_j
1	5.00	0.5	0.5	2.5	2.5	
2	118.75	0.4				
3	500.00	0.1				
Summe		1.0				


1
0.5
0.025
0.5
0.9

letzte Gruppe nochmals aufspalten

10% Reiche
- 1% sehr reich und $v_{4,neu} = 1$
- 9% mäßig reich, so dass $v_{3,neu} = 0.77$



4.1.2 Der Gini-Koeffizient

Definition 4.6. [Gini-Koeffizient]

Gegeben sei die geordnete Urliste $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ eines verhältnisskalierten Merkmals X. Dann heißt

$$G := \frac{2\cdot \sum_{i=1}^{n} i\cdot x_{(i)}}{n\sum_{i=1}^{n} x_i} - \frac{n+1}{n}$$

Gini-Koeffizient und

$$G^* := \frac{n}{n-1}\cdot G$$

normierter Gini-Koeffizient (Lorenz-Münzner-Koeffizient).

Bem. 4.7.

- Man kann zeigen (Herleitung über Trapezformel $\rightarrow$ Toutenburg & Heumann, 2009):

$$G = \frac{\text{Fläche zwischen Winkelhalbierender und Lorenz-Kurve}}{\text{Fläche des Dreiecks aus Winkelhalbierender, Abszisse u. Gerade } \{\text{x}=1\}}$$

$$= 2 \cdot \text{Fläche zwischen Winkelhalbierender und Lorenz-Kurve}$$

- Es gilt: $G = 0$ bei minimaler Konzentration und $G = \dfrac{n-1}{n}$ bei maximaler Konzentration

- Damit ist: $G^* = 0$ bei minimaler Konzentration und $G^* = 1$ bei maximaler Konzentration. (Ist n sehr groß, so ist $\frac{n-1}{n} \approx 1$, also $G^* \approx G$.)

- Betrachte die geordneten Ausprägungen $a_1 < a_2 < \ldots < a_k$ mit den Häufigkeiten $h_1, h_2, \ldots, h_k$. Dann gilt

$$G = \frac{\sum_{l=1}^{k}(u_{l-1}+u_l)f_l \cdot a_l}{\sum_{l=1}^{k} f_l \cdot a_l} - 1 = \frac{\sum_{l=1}^{k}(u_{l-1}+u_l)h_l \cdot a_l}{\sum_{l=1}^{k} h_l \cdot a_l} - 1 =$$

$$= 1 - \sum_{l=1}^{k} f_l(v_{l-1}+v_l)$$

mit $u_j = \frac{1}{n}\sum_{l=1}^{j} h_l, \quad u_0 := 0, \quad v_0 := 0.$

Bsp. 4.8. [Konzentrationsmessung]

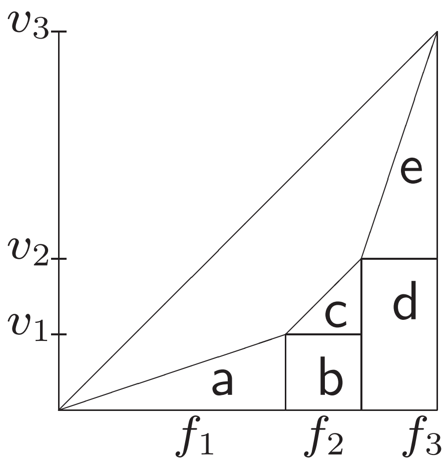
Fortsetzung von Bsp. 4.5

Man veranschauliche und berechne in der Situation von Bsp. 4.5 den Gini-Koeffizienten.

j	f_j	u_j	v_j
0	-	0	0
1	0.5	0.5	0.025
2	0.4	0.9	0.5
3	0.1	1.0	1.0

$$
\begin{aligned}
G &= 1 - \sum_{l=1}^{k} f_l(v_{l-1} + v_l) = \\
&= 1 - (f_1 \cdot v_1 + f_2 \cdot (v_1 + v_2) + f_3 \cdot (v_2 + v_3)) = \\
&= 1 - (0.5 \cdot 2.5\% + 0.4 \cdot 52.5\% + 0.1 \cdot 150\%) = \\
&= 0.6275
\end{aligned}
$$

Graphische Darstellung



$$
\begin{aligned}
\underline{\text{Gini Koeffizient}} \quad &= \quad 1 - 2 \cdot (a + b + c + d + e) = \\
&= \quad 1 - 2\left(\frac{f_1 \cdot v_1}{2} + f_2 \cdot v_1 + \frac{f_2 \cdot (v_2 - v_1)}{2} + f_3 \cdot v_2 + \frac{f_3 \cdot (v_3 - v_2)}{2}\right) = \\
&= \quad 1 - 2\left(\frac{f_1 \cdot v_1}{2} + \frac{f_2 \cdot v_1}{2} + \frac{f_2 \cdot v_2}{2} + \frac{f_3 \cdot v_2}{2} + \frac{f_3 \cdot v_3}{2}\right) = \\
&= \quad 1 - (f_1 v_1 + f_2 v_1 + f_2 v_2 + f_3 v_2 + f_3 v_3)
\end{aligned}
$$

Ausgewählte Gini-Koeffizienten der landwirtschaftlich genutzten Fläche

Paraguay	$G = 0.913$	61.6%	haben $< 10ha$
Italien	$G = 0.768$	87.7%	“
Portugal	$G = 0.716$	91.5%	“
Österreich	$G = 0.446$	53.2%	“
Schweiz	$G = 0.232$	52.3%	“

Übergang von der Lorenz-Kurve zum Gini-Koeffizienten:

- Informationsverlust, insbesondere können verschiedene Lorenz-Kurven denselben Gini-Koeffizienten haben

+ Totale Ordnung: Gini-Koeffizienten als Zahlen lassen sich immer ordnen, Lorenz-Kurven im Allgemeinen nicht.

4.1.3 Quantilsbezogene relative Konzentrationsmessung

Oft stehen die Daten in einer anderen Form zur Verfügung:

Gegeben sind dann Quantile (typischerweise Quartile, Quintile oder Dezile, allgemein $\frac{1}{q} \cdot 100\%$ Quantile) und die Anteile $z_1, z_2, \ldots, z_q$ des Merkmals, die auf die jeweiligen Quantile entfallen.

Wie kann man in diesem Fall die Lorenz-Kurve berechnen?

Bsp. 4.9. [Einkommensverteilung Brasilien]

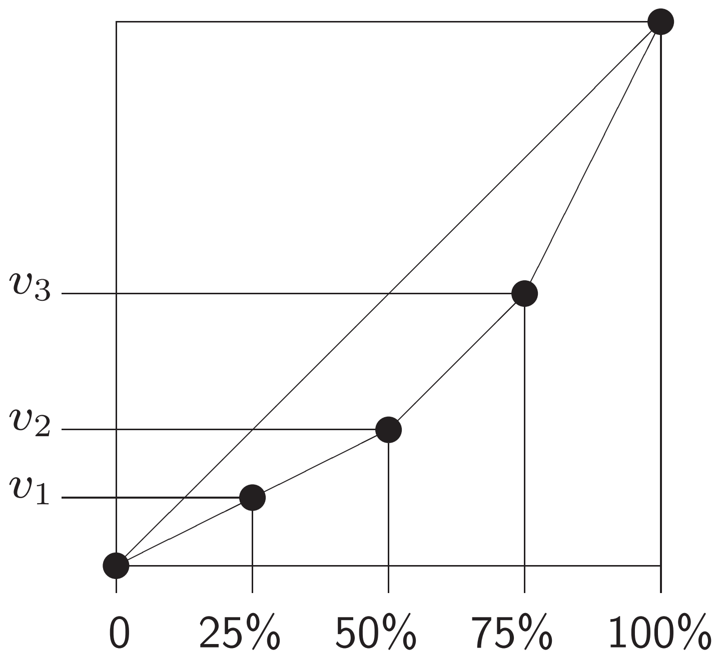
Quintil	1	2	3	4	5
	2.5 z_1	4.9 z_2	9.2 z_3	18.3 z_4	65.2 z_5

Quelle: World Resources Institute:
`http://pdf.wri.org/wrr98_chapter07.pdf` (Aufruf 2011)

Aus $z_1, z_2, \ldots$ kann man $v_1, v_2, \ldots$ und damit die Lorenz-Kurve bestimmen

$$\begin{aligned} v_1 &= z_1 \\ v_2 &= z_2 + v_1 = z_2 + z_1 \\ &\vdots \\ v_j &= \sum_{l \leq j} z_l \end{aligned}$$

Lorenz-Kurve bei gegebenen Quartilen und den Anteilen dieser Quartile:



Diese Lorenz-Kurve basiert auf vergröberter Information $(u_j, v_j), j = 1, \dots k$ und ist nur einer Annäherung an die „wahre" Lorenz-Kurve $(\tilde{u}_l, \tilde{v}_l), l = 0, 1, \dots, n$.

Beide Kurven stimmen in den Punkten $u_j = \tilde{u}_{j \cdot \alpha \cdot n}$ überein, d.h.

$$v_j = \tilde{v}_{j \cdot \alpha \cdot n}.$$

Man erhält für $\tilde{v}_{\alpha \cdot n} = v_1$ bzw. allgemein für $\tilde{v}_{j \cdot \alpha \cdot n} = v_j$

$$v_1 = \frac{\sum\limits_{i=1}^{\alpha \cdot n} x_{(i)}}{\sum\limits_{i=1}^{n} x_{(i)}} \qquad \text{bzw.} \qquad v_j = \frac{\sum\limits_{i=1}^{j \cdot \alpha \cdot n} x_{(i)}}{\sum\limits_{i=1}^{n} x_{(i)}}$$

Bildet man die folgenden Differenzen:

$$\begin{aligned} z_1 &:= v_1 - 0 \\ z_2 &:= v_2 - v_1 \\ &\vdots \\ z_j &:= v_j - v_{j-1}, \end{aligned}$$

so erhält man genau den Anteil, der ins q-te Quantil fällt, denn:

$$z_1 = \frac{\sum_{i=1}^{\alpha \cdot n} x_{(i)}}{\sum_{i=1}^{n} x_{(i)}} \qquad \text{und} \qquad z_2 = \frac{\sum_{i=1}^{2\alpha \cdot n} x_{(i)} - \sum_{i=1}^{\alpha \cdot n} x_{(i)}}{\sum_{i=1}^{n} x_{(i)}} = \frac{\sum_{i=\alpha \cdot n+1}^{2\alpha \cdot n} x_{(i)}}{\sum_{i=1}^{n} x_{(i)}}$$

usw.

Bem. 4.10.

Für den Gini-Koeffizienten G der „vergröberten Lorenz-Kurve“ $(u_j, v_j), j = 0, 1, \ldots, n$ kann man einfach auf die Formeln aus Kapitel 4.1.2 zurückgreifen.

- Dies liefert (wegen der Konvexität der Lorenz-Kurve) eine untere Schranke für den Gini-Koeffizienten $\tilde{G}$ der „wahren Lorenz-Kurve“. Es gilt $\tilde{G} = G$, wenn in den jeweiligen Quantilen alle Einkommen gleich sind.

- Man hat dann Häufigkeitsdaten mit den Ausprägungen $a_1, a_2, \ldots, a_q$ vorliegen: a_l ist der Wert im l-ten Quantil.

- Dann ist

$$z_l = \frac{\displaystyle\sum_{i=(l-1)\cdot\alpha\cdot n+1}^{l\cdot\alpha\cdot n} x_{(i)}}{\displaystyle\sum_{i=1}^{n} x_{(i)}} = \frac{\alpha \cdot n \cdot a_l}{\sum_{r=1}^{q} \alpha \cdot n \cdot a_r} = \frac{a_l}{\sum_{r=1}^{q} a_r}$$

und damit (wegen $f_l = \alpha =$ const.)

$$\begin{aligned} G &= \frac{\sum\limits_{l=1}^{k}(u_{l-1}+u_l)f_l\cdot a_l}{\sum\limits_{l=1}^{k} f_l\cdot a_l} - 1 \\ &= \sum_{l=1}^{q}(u_{l-1}+u_l)\cdot\frac{a_l}{\sum\limits_{r=1}^{q} a_r} - 1 \\ &= \left(\sum_{l=1}^{q}(u_{l-1}+u_l)\cdot z_l\right) - 1 \end{aligned}$$

Beispiel Brasilien:

Quintil j	z_j	u_j	v_j	
1	2.5	0.2	$v_1 = z_1 =$	2.5%
2	4.9	0.4	$v_2 = z_1 + z_2 = 2.5\% + 4.9\% =$	7.4%
3	9.2	0.6	$v_3 = z_1 + z_2 + z_3 =$	16.6%
4	18.3	0.8	$v_4 = z_1 + z_2 + z_3 + z_4 =$	34.9%
5	65.2	1.0	$v_5 =$	100%

Gini-Koeffizienten:

Brasilien	0.5562
Algerien	0.3572
China	0.3460
Niger	0.3300
Deutschland	0.3034
Finnland	0.2460

4.1.4 Einige weitere quantilsbasierte Maße

Insbesondere basierend auf dieser „natürlichen, äquidistanten Einteilung“ lassen sich weitere relative Konzentrationsmaße definieren:

Robin-Hood-Index (maximaler Nivellierungssatz, Schutzkoeffizient) (Wagschal (1999, S.135ff))

- Wie viel müsste den Reichen weggenommen werden, um zu einer Konzentration von 0 zu kommen?
- Ermittle für jedes Quantil den Abstand seines Anteils zu α!
- Aufaddieren der positiven Abstände liefert den Robin-Hood-Index. Dieser Anteil müsste verteilt werden, um zu einer gleichen Verteilung zu kommen!
- Grafische Bestimmung des Robin-Hood-Indexes $\rightarrow$Wagschal (1999, S.132)
- Man kann zeigen, dies ergibt genau den maximalen vertikalen Abstand zur Winkelhalbierenden.

Quantilverhältnisse

- Dezilverhältnisse
- z.B. Dezilverhältnis 90:10$=\dfrac{x_{0.9}}{x_{0.1}}$, falls $x_{0.1} > 0$
- beim Einkommensvergleich: um welchen Faktor ist der untere Wert der 10% Reichsten größer als der obere Wert der 10% Ärmsten
- Minimale Konzentration: alles in einem Punkt $x_{0.1} = x_{0.9}$
  $\Rightarrow$ Dezilverhältnis $= 1$
- Vorsicht bei extremer Konzentration!

Bsp. 4.11.

Dezilverhältnisse 90:10 des Einkommens von Vollbeschäftigten im internationalen Vergleich $\rightarrow$ Wagschal (1999, S.138)

Norwegen	1.98	Italien	2.80
Schweden	2.13	Neuseeland	3.04
Dänemark	2.17	Japan	3.04
Belgien	2.25	Frankreich	3.26
Finnland	2.29	Großbritannien	3.33
Deutschland	2.32	Österreich	3.58
Niederlande	2.59	Kanada	4.02
Schweiz	2.71	Portugal	4.05
Australien	2.79	USA	4.16

4.2 Absolute Konzentration

4.2.1 Vorbemerkungen

- Alle bisherigen Maße betrachteten die relative Konzentration: Sowohl die u_j als auch die v_j waren Anteile.
- Die absolute Konzentration bezieht die absolute Zahl der Merkmalsträger mit ein.
- Es ist jeweils inhaltlich zu entscheiden, welche der beiden Arten von Konzentration von Interesse ist.
- Relative Konzentrationsmessung ist heranzuziehen: wenn
  - die Merkmalsumme auf sehr viele Einheiten verteilt wird (z.B. Boden/Einkommen)
  - oder bei der Frage: Herrscht im Markt ein Übergewicht?
- Absolute Konzentrationsmaße werden v.a. dann verwendet, wenn die Merkmalsumme nur auf wenige Einheiten aufgeteilt wird, z.B. Betriebssysteme, Engergieversorger, Ölkonzerne.

4.2.2 Einige Maßzahlen der absoluten Konzentration

Definition 4.12.

Sei $0 \leq x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ die geordnete Urliste eines verhältnisskalierten Merkmals mit $\sum_{i=1}^{n} x_i > 0$.

Mit

$$p_{(i)} := \frac{x_{(i)}}{\sum_{j=1}^{n} x_j}$$

heißt

$$CR_g := \sum_{i=n-g+1}^{n} p_{(i)}$$

Konzentrationsrate (vom Grade g).

- Maximale Konzentration
  $CR_1 = 1$: Die größte Beobachtung hat alles.

- Minimale absolute Konzentration

$$x_{(1)} = x_{(2)} = \ldots = x_{(n)} =: x$$

$$p_{(i)} = \frac{x_{(i)}}{\sum_{j=1}^{n} x_j} = \frac{x}{n \cdot x}$$

  also $CR_g = \sum_{i=n-g+1}^{n} p_{(i)} = \frac{g}{n}$, abhängig von n

  Für $n = 2$ ist $CR_1 = \frac{1}{2}$, für $n \to \infty$ geht CR_1 gegen 0.

- Wahl von g? Typischerweise „klein“: 2 oder 3

Bsp. 4.13.

Zweitstimmenanteile polit. Parteien bei Bundestagswahlen 1949-2005

	1949	1953	1965	1972	1983	1994	2002	2005	2009
CDU/CSU	31,0%	45,2%	47,6%	44,9%	48,8%	41,5%	38,5%	35,2%	33,8%
SPD	29,2%	28,8%	39,3%	45,8%	38,2%	36,4%	38,5%	34,2%	23,0%
FDP	11,9%	9,5%	9,5%	8,4%	7,0%	6,9%	7,4%	9,8%	14,6%
Grüne	-	-	-	-	5,6%	7,3%	8,6%	8,1%	10,7%
PDS/Die Linke	-	-	-	-	-	4,4%	4,0%	8,7%	11,9%
Sonstige	27,9%	16,5%	3,6%	0,9%	0,4%	3,5%	3,0%	4,0%	7,0%

Definition 4.14.

In der Situation von Def. 4.12 heißt

$$H := \sum_{i=1}^{n} p_{(i)}^2$$

Herfindahl-Index. Die Größe $1 - H$ wird auch Rae-Index genannt.
In der Politikwissenschaft wird $\frac{1}{H}$ auch als „Anzahl der effektiven Parteien“ bezeichnet.

Bem. 4.16.

- $\sum p_{(i)}$ gibt Idee, da konstant 1. Durch das Quadrieren werden große Anteile relativ gesehen noch größer.

- Da über alle $p_{(i)}$ summiert wird, müssen für den natürlichen Herfindahl-Index die $p_{(i)}$ nicht geordnet sein.

- H liegt zwischen $\frac{1}{n}$ und 1.

  maximale Konzentration:

$$H = 1, \quad \text{da } p_{(n)} = 1, \text{und alle restlichen } p_{(i)} = 0$$

  minimale Konzentration $p_{(i)} = \frac{1}{n}$:

$$H = \sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 = n \cdot \frac{1}{n^2} = \frac{1}{n};$$

  H hängt tatsächlich von Anzahl Marktteilnehmer ab!

- In der Tat wird die absolute Konzentration gemessen:
  Hat man q Einheiten, die jeweils den gleichen Anteil ausmachen, so gilt:

$$
\begin{array}{ll}
q = 1 & H = 1 \text{ maximale Konzentration} \\
q = 2 & H = (\frac{1}{2})^2 + (\frac{1}{2})^2 = \frac{1}{2} \\
q = 3 & H = (\frac{1}{3})^2 + (\frac{1}{3})^2 + (\frac{1}{3})^2 = \frac{1}{3} \\
\text{usw.} &
\end{array}
$$

  während hier immer $G \equiv 0$ gilt.

.

Bsp. 4.18. [Herfindahl- und Rae-Index des deutschen Parteienwesens]

Bsp. 4.19. [„Durchschnittliche Fraktionalisierung von Parteiensystemen“]

(Waagschal (1999, S. 145))

Land	durchschnittl. Rae-Index 1945-93
USA	0.53
Österreich	0.60
Großbritannien	0.62
Deutschland	0.64
Schweiz	0.71
Italien	0.75
Frankreich	0.79
Niederlande	0.79
Finnland	0.82