

Friedrich-Alexander-Universität
Erlangen-Nürnberg
Wirtschafts-und Sozialwissenschaftliche Fakultät

Diskussionspapier
26 / 1999

Systematik der Schiefemessung für ordinalskalierte
Merkmale

Ingo Klein



Lehrstuhl für Statistik und Ökonometrie
Lehrstuhl für Statistik und empirische Wirtschaftsforschung
Lange Gasse 20 · D-90403 Nürnberg

Systematik der Schiefemessung für ordinalskalierte Merkmale

Ingo Klein
Lehrstuhl für Statistik und Ökonometrie
Universität Erlangen–Nürnberg
Lange Gasse 20
D–90403 Nürnberg
Germany
E-mail: ingo.klein@wiso.uni-erlangen.de

Abstract

We derive almost all known measures of skewness from differences of probability or differences of quantiles. Because ordinal variables are measured non–uniquely with respect to strictly increasing transformations functions of differences of quantiles cannot be used to describe the skewness of these variables. Therefore, we construct a class of very simple measures of skewness as functions of differences of probability. The concepts are illustrated by the binomial– and the beta–distribution.

1 Einleitung

Schiefemaße werden in der Literatur ausschließlich für quantitative Merkmale angegeben. Dies zeigt sich bereits daran, daß diese Maßzahlen von Differenzen von Quantilen, Differenzen von Median und Mittelwert und Differenzen von Teilmittelwerten ausgehen. Die gängigen Charakterisierungen von Schiefemaßen setzen sogar stetige Verteilungsfunktionen voraus (siehe z.B. die Übersichten in Oja (1981) und in Handl (1986)).

Betrachtet man aber die Häufigkeitsverteilung der Zensuren, die im Grundstudium im Sommersemester 1998 in der Vordiplomsklausur des Lehrstuhls für Statistik und Ökonometrie erzielt wurden, so erkennt man eine deutliche Schiefe der Verteilung, obwohl es sich bei dem Merkmal "Zensur" nicht um ein quantitatives und schon gar nicht um ein Merkmal mit stetiger Verteilungsfunktion handelt.

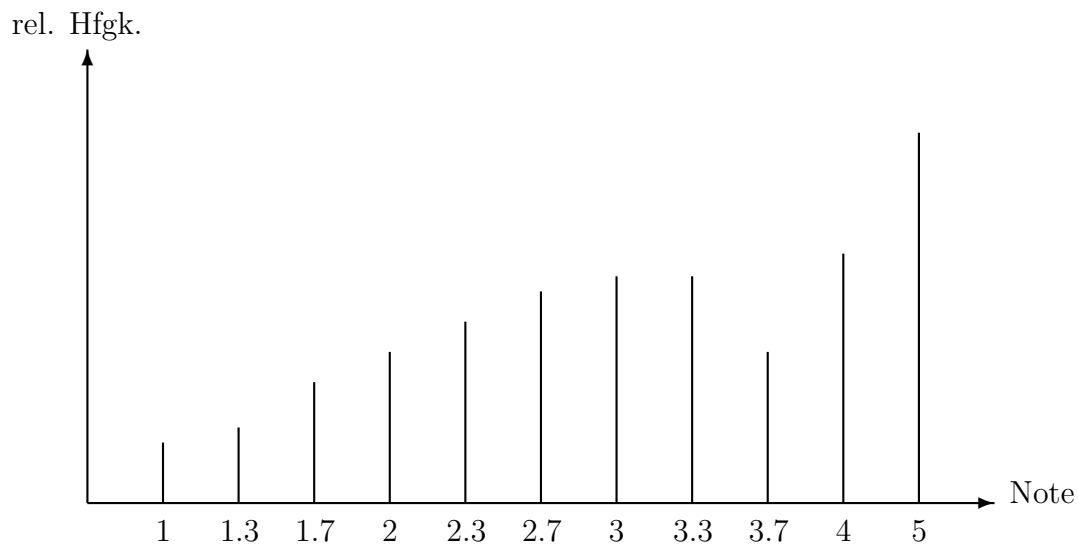


Abbildung 1: Relative Häufigkeitsverteilung der Noten der Grundausbildungsklausur in Statistik vom Sommersemester 1998 des Lehrstuhls für Statistik und Ökonometrie der Universität Erlangen–Nürnberg

Die Abbildung 1 betont die Ordinalskala der Zensuren, in dem die einzelnen Zensuren auf der Abszisse bewußt nicht entsprechend ihres numerischen Wertes sondern äquidistant im Sinne einer klassifizierenden Reihenfolge abgetragen wurden. Rein graphisch ist das Ausmaß der Schiefe für die Verteilung von ordinalskalierten Merkmalen nur schwer zu beurteilen, da die zulässigen Transformationen der Ordinalskala die Werte auf der Abszisse nahezu beliebig verschieben können, solange nur die Reihenfolge der Werte aufrechterhalten bleibt. Konstant bleiben jedoch die

Ordinatenwerte, die als Basis für eine Klasse von Schiefemaßen für ordinalskalierte Merkmale dienen sollen. Voraussetzung für die Anwendbarkeit dieser Schiefemaße ist jedoch, daß lediglich eine endliche Anzahl potentieller Merkmalsausprägungen existiert, was aber im sozialwissenschaftlichen Kontext kaum eine Einschränkung bedeutet, da zumeist zwischen einer endlichen Anzahl von Alternativen zu entscheiden ist. Wenn eine Maßzahl für ordinalskalierte Merkmale entwickelt wird, kann sie auch für quantitative Merkmale kalkuliert werden. Die vorgeschlagenen Schiefemaße werden deshalb zunächst auf endlich–diskrete Merkmale angewendet und am Beispiel der Binomialverteilung illustriert.

Einheitlicher Ausgangspunkt der vorgeschlagenen Schiefemaße für ordinalskalierte Merkmale ist eine Charakterisierung von Symmetrie und Schiefe durch Wahrscheinlichkeits– oder Häufigkeitsdifferenzen, die auch bereits van Zwet (1979) betrachtet hat, um die Relation zwischen Mittelwert, Median und Modus zu beschreiben. Diese Wahrscheinlichkeitsdifferenzen lassen sich aber im Falle quantitativer Merkmale mit stetiger Verteilungsfunktion äquivalent durch Quantilsdifferenzen ersetzen. Dies soll auch für die neuen Schiefemaße geschehen und an der Betaverteilung exemplifiziert werden. Anschließend wird gezeigt, daß nahezu sämtliche für diesen Fall in der Literatur diskutierten Schiefemaße ebenfalls von solchen Quantilsdifferenzen ausgehen. Schließlich werden die unterschiedlichen Schiefemaße miteinander verglichen.

Verteilungsmaßzahlen dienen der Deskription von Verteilungsfunktionen und werden in der Literatur (vgl. z.B. Bickel & Lehmann (1975)) zumeist als Funktionale, d.h. Abbildungen einer Teilmenge aller Verteilungsfunktionen in die Menge der reellen Zahlen, formalisiert. Dabei bleibt offen, ob diese Verteilungsfunktionen Wahrscheinlichkeitsverteilungen von Zufallsvariablen oder Häufigkeitsverteilungen von Merkmalen beschreiben. Wir werden deshalb im folgenden für beide Fälle eine einheitliche Symbolik verwenden und mit F_X eine Verteilungsfunktion der Zufallsvariablen X bezeichnen, von der auch alternativ als Merkmal gesprochen wird. Damit ist klar, daß als Ausprägungen ordinalskalierter Merkmale der Einfachheit halber nur Zahlenwerte zugelassen werden. Der Begriff Masse bezeichnet gleichermaßen Häufigkeits– wie Wahrscheinlichkeitsmasse. $P(\cdot)$ kann sowohl ein Häufigkeits– als auch Wahrscheinlichkeitsmaß sein.

Da wir es mit ordinalen Merkmalen zu tun haben, sollen deren Menge potentieller Ausprägungen abzählbar–endlich ($= k$) und diese potentiellen Ausprägungen x_i , $i = 1, 2, \dots, k$ entsprechend $x_1 < x_2 < \dots < x_k$ geordnet sein.

2 Schiefe für ordinalskalierte Merkmale

Auch die Schiefemessung für ordinalskalierte Merkmale muß zunächst von den Extrempunkten ausgehen, die geeignet quantifiziert werden sollen. Dabei steht der Fall der symmetrischen Verteilung im Mittelpunkt, für den ein geeignetes Schiefemaß den Wert Null annehmen soll. Anschließend sind diejenigen Verteilungssituationen zu qualifizieren, für die ein Schiefemaß ein u.U. mögliches Minimum bzw. Maximum

annehmen soll.

2.1 Ausgangspunkt: Symmetrie

Bickel & Lehmann (1975) gehen von einer Symmetriedefinition aus, die einen sog. Symmetriepunkt μ benötigt. Sie sagen, daß F_X eine symmetrische Verteilung ist, wenn ein $\mu \in \mathbb{R}$ existiert, so daß $X - \mu$ und $-(X - \mu)$ identisch verteilt sind. μ heißt Symmetriepunkt. D.h. es gilt

$$F_{X-\mu}(x) = F_{-(X-\mu)}(x) = P(-(X-\mu) \leq x) = P(X-\mu \geq -x) = 1 - P(X-\mu < -x)$$

Diese Definition benötigt implizit einen Abstand einer Zufallsvariablen von dem Symmetriepunkt μ und ist damit für den Fall der Ordinalskala nicht geeignet.

Wenn X ein Merkmal repräsentiert mit abzählbar–endlich vielen potentiellen Merkmalsausprägungen $-\infty < x_1 < x_2 < \dots < x_k < \infty$, kann eine Definition der Symmetrie erfolgen, die ohne das Konzept des Symmetriepunktes auskommt. Es muß für eine symmetrische Verteilung F_X gelten:

$$F_X(x_i) = P(X \leq x_i) = P(X \geq x_{k+1-i}) = 1 - F_X(x_{k-i})$$

für $i = 1, 2, \dots, k - 1$. Es ist einzuräumen, daß nun statt der Existenz eines Symmetriepunktes die Endlichkeit gefordert werden muß.

Entscheidend für Symmetrie und Schiefe ist also die Differenz

$$\Delta_{1i} = F_X(x_i) - (1 - F_X(x_{k-i})).$$

F_X heißt dann

1. rechtsschief (linkssteil), wenn

$$\Delta_{1i} = F_X(x_i) - (1 - F_X(x_{k-i})) \geq 0 \text{ für } i = 1, 2, \dots, k - 1$$

und

$$\Delta_{1i} = F_X(x_i) - (1 - F_X(x_{k-i})) > 0 \text{ für mindestens ein } i \in \{1, 2, \dots, k - 1\}$$

gilt (d.h. F_X baut sich schneller als die Gegenwahrscheinlichkeit $1 - F_X$ auf) und

2. linksschief (rechtssteil), wenn

$$\Delta_{1i} = F_X(x_i) - (1 - F_X(x_{k-i})) \leq 0 \text{ für } i = 1, 2, \dots, k - 1$$

und

$$\Delta_{1i} = F_X(x_i) - (1 - F_X(x_{k-i})) < 0 \text{ für mindestens ein } i \in \{1, 2, \dots, k - 1\}$$

gilt (d.h. F_X baut sich langsamer als die Gegenwahrscheinlichkeit $1 - F_X$ auf).

2.2 Extremfälle

Es bleibt die Frage, wie eine Situation extremer Links- bzw. Rechtsschiefe zu charakterisieren ist. Die Linksschiefe muß zunehmen, wenn die kleinste Merkmalsausprägung an Masse verliert und die größte Merkmalsausprägung an Masse gewinnt. Im Extremfall konzentriert sich die gesamte Häufigkeitsmasse auf diese größte Ausprägung. D.h. $F_X(x_k) = 1$ und $F(x_i) = 0$ für $i = 1, 2, \dots, k - 1$. Die Abbildung 2 charakterisiert die Häufigkeitsverteilung der Zensuren in diesem extremen Fall.

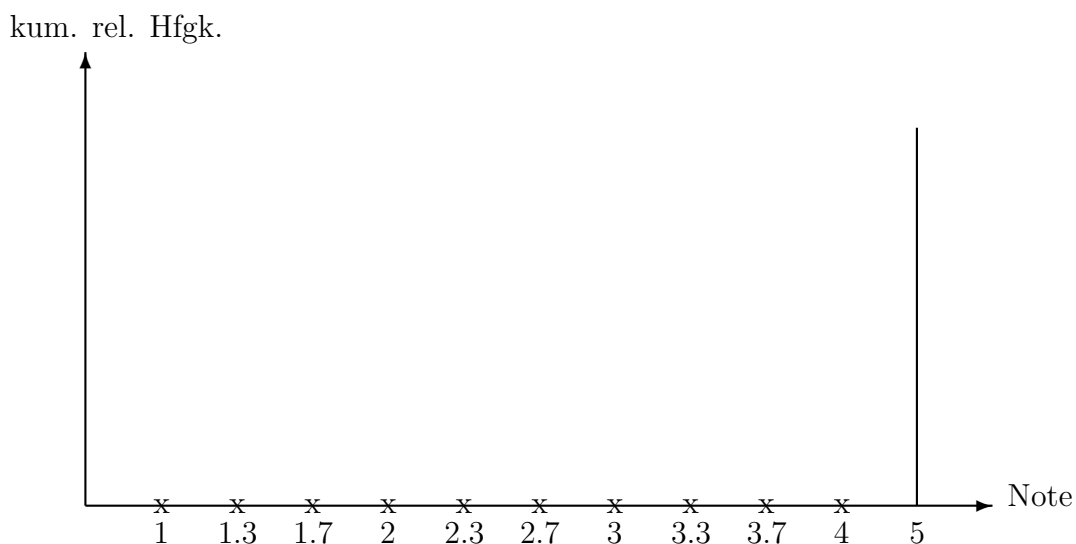


Abbildung 2: Extreme linksschiefe (rechtssteile) kumulierte relative Häufigkeitsverteilung von Zensuren

Umgekehrt ist dann klar, daß eine Konzentration der gesamten Masse auf die kleinste Ausprägung dem Fall der extremen Rechtsschiefe entspricht (siehe Abbildung 3). D.h. $F_X(x_1) = 1$, woraus $F_X(x_i) = 1$ für $i = 1, 2, \dots, k - 1$ folgt.

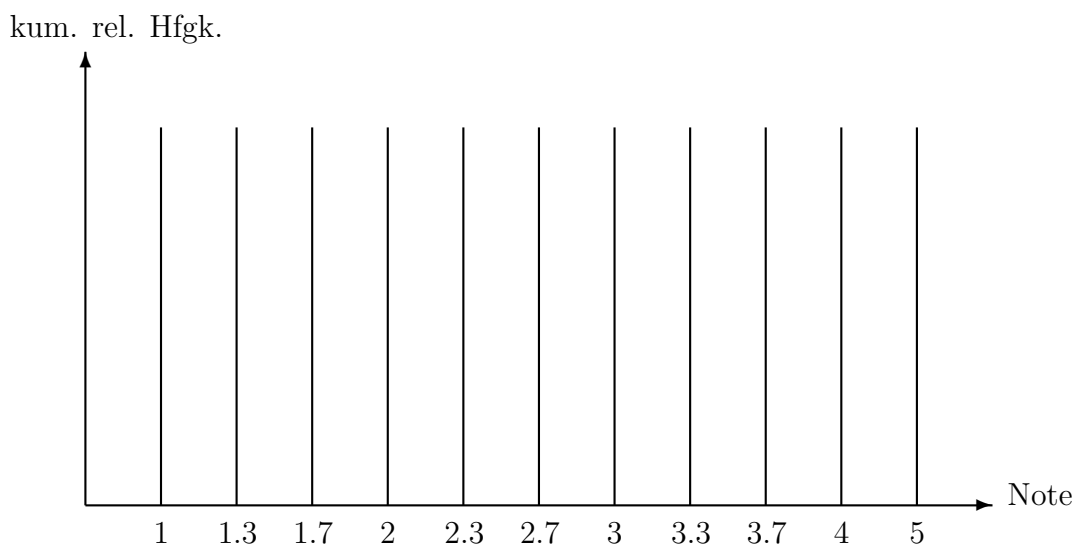


Abbildung 3: Extreme rechtsschiefe (linkssteile) kumulierte relative Häufigkeitsverteilung der Zensuren

Zu fordern ist lediglich, daß ein Schiefemaß in jeweils genau einer dieser beiden extremen Situationen ein Minimum bzw. ein Maximum annimmt. Welche Situation das Minimum und welche das Maximum liefert, ist per Konvention festzulegen.

2.3 Funktionale Abhängigkeit

Über die konkrete funktionale Form eines Schiefemaßes ist ohne sehr weitreichende zusätzliche Forderungen keine Aussage möglich. Im Falle ordinalskaliertener Merkmale kann aber unter sehr schwachen Bedingungen gezeigt werden, daß ein Schiefemaß nur von Werten der Verteilungsfunktion F_X abhängen kann.

Die erste Bedingung betrifft die sog. Vergleichsinvarianz bezüglich zulässiger Skalentransformationen. Betrachtet man eine Ordinalskala und formuliert eine Maßzahl als Funktional S auf einer geeignet zu wählenden Teilmenge \mathcal{F} aller Verteilungsfunktionen, so lautet die Forderung der Vergleichsinvarianz

$$S(F_X) = S(F_Y) \iff S(F_{g(X)}) = S(F_{g(Y)})$$

für alle streng monoton zunehmenden Transformationen g und $F_X, F_Y \in \mathcal{F}$. Diese Vergleichsinvarianz ist von Pfanzagl (1971) in den Mittelpunkt eines meßtheoretischen (empirischen) Sinnhaftigkeitskonzeptes von Maßzahlen gestellt worden. Für diese Sinnhaftigkeit muß zumindest ein skalenunabhängiger Paarvergleich von Verteilungen möglich sein. Die wissenschaftstheoretische Einbettung dieses Konzeptes wird in Klein (1984) diskutiert.

Klein ((1994), Korollar 6.2.2, S. 179) gezeigt, daß diese Vergleichsinvarianz notwendig die Absolutinvarianz

$$S(F_X) = S(F_{g(X)})$$

für alle streng monoton zunehmenden Transformationen g nach sich zieht, wenn es mindestens eine streng monoton zunehmende Transformation gibt, die den Wert von S nicht verändert. Es ist nun sinnvoll von Schiefemaßen zu fordern, daß sie durch Translationen von Verteilungen nicht bewegt werden.

Damit kann ein bezüglich aller streng monotoner Transformationen vergleichsinvariantes Schiefemaß nur noch von sog. Maximalinvarianten abhängen und dies sind nach Klein ((1994), Satz 6.7.1, S. 202) im Falle der Ordinalskala die Werte der Verteilungsfunktion selbst.

Diese Argumentation greift zum Beispiel auch für Streuungsmaße und führt dazu, daß das von Vogel & Dobbener (1982) vorgeschlagene Streuungsmaß für ordinalskalierte Merkmale ebenfalls eine Funktion von Summenhäufigkeiten ist. Der Median als Lagemaß wird von jeder streng monotonen Transformation bewegt, so daß für diesen die vorstehende Argumentation nicht greift, und er von den Merkmalsausprägungen explizit abhängt.

2.4 Schiefemaße für ordinalskalierte Merkmale

Da der Fall der Rechtsschiefe sich nur durch einen Wechsel des Ungleichheitszeichens von dem der Linksschiefe unterscheidet, soll im folgenden nur noch die Rechtsschiefe diskutiert werden.

Ausgehend von der symmetriepunktfreien Symmetriedefinition kann dann ein Schiefemaß für Verteilungen F_X mit abzählbar-endlichem Träger $x_1 < x_2 < \dots < x_k$ mittels

$$\begin{aligned} SO_I(F_X) &= \sum_{i \in I} \Delta_{1i} \\ &= \sum_{i \in I} F_X(x_i) - (1 - F_X(x_{k-i})) \\ &= -n(I) + \sum_{i \in I} (F_X(x_i) - F_X(x_{k-i})) \end{aligned}$$

vorgenommen werden, wobei $I \subseteq \{1, 2, \dots, k-1\}$ und $n(I)$ die Anzahl der Elemente von I angeben. I determiniert das konkrete Schiefemaß. Wenn $I = \{i\}$ für ein $i \in \{1, 2, \dots, k\}$ ist, dann werden zwei einzelne Wahrscheinlichkeiten miteinander verglichen, d.h.

$$SO_{\{i\}}(F_X) = \Delta_{1i} = -1 + F_X(x_i) + F_X(x_{k-i}).$$

Ist $I \subseteq \{1, 2, \dots, k-1\}$, dann werden Teilsummen von Wahrscheinlichkeiten verglichen, z.B. für $I = \{1, 2\}$

$$SO_{\{1,2\}}(F_X) = \sum_{i=1}^2 \Delta_{1i} = -2 + F_X(x_1) + F_X(x_2) + F_X(x_{k-2}) + F_X(x_{k-1})$$

und schließlich wird für $I = \{1, 2, \dots, k-1\}$ die gesamte Summe möglicher Wahrscheinlichkeitsdifferenzen betrachtet, d.h.

$$SO_{\{1,2,\dots,k-1\}}(F_X) = \sum_{i=1}^{k-1} \Delta_{1i} = -(k-1) + 2 \sum_{i=1}^{k-1} F_X(x_i).$$

Die vorstehenden Schiefemaße besitzen die folgenden leicht zu verifizierenden Eigenschaften:

1. Als Wertebereich gilt

$$-n(I) \leq SO_I(F) \leq n(I).$$

2. Das Minimum von $SO_{\{1,2,\dots,k-1\}}(F_X)$ ist $-(k-1)$ und wird im Falle $F_X(x_i) = 0$ für $i = 1, 2, \dots, k-1$ angenommen, was einer Einpunktverteilung im Punkt x_k entspricht. Dies ist die bereits beschriebene Situation einer extremen Rechtssteilheit bzw. Linksschiefe. Ist jedoch $I \subset \{1, 2, \dots, k-1\}$, so ergibt sich das Minimum $-n(I)$ im Falle der extremen Rechtssteilheit, wenn $k-1 \in I$. Ansonsten stellt das jeweilige Minimum nur eine notwendige Bedingung für die extreme Rechtssteilheit dar.
3. Das Maximum wird im Falle $I = \{1, 2, \dots, k-1\}$ für $F_X(x_i) = 1$, $i = 1, 2, \dots, k-1$ angenommen, was einer Einpunktverteilung im Punkt x_1 entspricht. Dies ist die Situation einer extremen Linkssteilheit bzw. Rechtsschiefe. Wiederum stellt sich das selbe Ergebnis auch für $I \subset \{1, 2, \dots, k-1\}$ ein, wenn $1 \in I$.
4. $SO_I(F_X) = 0$ ist eine notwendige, aber nicht hinreichende Bedingung für Symmetrie.
5. $SO_I(F_X)$ hängt ausschließlich von Verteilungsfunktionswerten ab und ist damit absolutinvariant bezüglich streng monoton zunehmender Transformationen, was nach der Darlegungen des vorherigen Abschnittes für die Schiefemessung ordinalskaliertter Merkmale auch zu fordern ist, wenn Schiefemaße invariant bezüglich Lageveränderungen sein sollen.

Bedauerlicherweise sind diese Schiefemaße abhängig von $n(I)$. Diese Abhängigkeit läßt sich durch Einführung des auf das Intervall $[-1, 1]$ normierten Schiefemaßes

$$SO_I^*(F_X) = SO_I(F_X)/n(I)$$

beheben.

Wir betrachten zunächst neuerlich das Beispiel der Zensurenverteilung.

Beispiel 2.1 Die folgende Tabelle stellt die Klausurergebnisse der Vordiplomklausuren der Statistischen Grundausbildung des Lehrstuhls für Statistik und Ökonometrie der Universität Erlangen–Nürnberg für drei verschiedene Semester mit sehr unterschiedlichen Teilnehmerzahlen zusammen. Die Schiefe ist in allen drei Situationen ähnlich.

Note	1. Klausur	2. Klausur	3. Klausur
1.0	4	8	6
1.3	0	10	1
1.7	5	16	5
2.0	11	20	9
2.3	4	24	9
2.7	5	28	3
3.0	6	30	5
3.3	7	30	6
3.7	7	20	9
4.0	17	33	9
5.0	17	49	16
$SO_{\{1,2,\dots,10\}}$	-3.205	-2.500	-2.154
$SO^*_{\{1,2,\dots,10\}}$	-0.320	-0.250	-0.215

Tabelle 1: Notenverteilung und Schiefemaße für drei ausgewählte Vordiplomklausuren im Fach Statistische Grundausbildung bei Verwendung einer elfgliedrigen Zensurenskala

Faßt man die Zensuren statt in 11 Kategorien in die üblichen 5 ganzzahligen Kategorien zusammen, so ergibt sich ein modifiziertes Bild der Schiefe. Diese reduziert sich deutlich und ist bei der dritten Klausur kaum noch wahrnehmbar.

Note	1. Klausur	2. Klausur	3. Klausur
1.0	4	18	7
2.0	15	60	23
3.0	18	88	14
4.0	24	53	18
5.0	17	49	16
$SO_{\{1,2,\dots,10\}}$	-0.897	-0.410	-0.333
$SO^*_{\{1,2,\dots,10\}}$	-0.224	-0.102	-0.083

Tabelle 2: Notenverteilung und Schiefemaße für drei ausgewählte Vordiplomklausuren im Fach Statistische Grundausbildung bei Verwendung einer fünfgliedrigen Zensurenskala

Die Schiefemaße SO_I können selbstverständlich auch für die Verteilung quantitativer Merkmale kalkuliert werden. Dies soll zunächst an diskreten quantitativen Merkmalen illustriert werden.

3 Anwendung für quantitative Merkmale (SO-Maße)

3.1 Diskreter Fall

Da jedes quantitative Merkmal X auch ein ordinalskaliertes ist, kann das Schiefemaß $SO_I(F_X)$ für diesen Fall berechnet und interpretiert werden. Die Anwendung ist offensichtlich, wenn die Anzahl der möglichen Realisationen von X endlich ist, d.h. wir betrachten wiederum $x_1 < x_2 < \dots < x_k$, wobei die Differenzen $x_{i+1} - x_i$, $i = 1, 2, \dots, k-1$ sinnvoll interpretierbar sind. Sind diese Differenzen alle gleich $c > 0$, d.h. x_1, \dots, x_k äquidistant, dann läßt sich eine enge Beziehung zwischen dem Schiefemaß $SO_{\{1,2,\dots,k-1\}}$ und dem Erwartungswert herleiten.

Sei F_X die Verteilungsfunktion von X , dann ist bekanntlich (vgl. z.B. Mood, Graybill & Boes (1974), S. 65)

$$E(X) = \int_0^\infty (1 - F_X(x))dx - \int_{-\infty}^0 F_X(x)dx.$$

Betrachten wir den einfachsten Fall äquidistanter Realisationen mit $x_1 = 0$ und $c = 1$, dann ist

$$SO_{\{1,2,\dots,k-1\}}(F_X) = (k-1) - 2E(X).$$

Für k ungerade, zeigt dieses Schiefemaß Symmetrie an, wenn $E(X) = (k-1)/2$ und damit gleich dem Median von F_X ist. Eine rechtsschiefe Verteilung wird angezeigt, wenn der Mittelwert kleiner als die mittlere Realisation $(k-1)/2$ ist.

Dieses Schiefemaß soll nun am Beispiel der Binomialverteilung untersucht werden.

Beispiel 3.1 Binomialverteilung

Für die Binomialverteilung mit den Parametern n und p sind $c = 1$, $E(X) = np$ und $k = n + 1$. Das Schiefemaß $SO_{\{1,2,\dots,k-1\}}$ vereinfacht sich dann zu:

$$SO_{\{1,2,\dots,n\}} = n - 2np = n(1 - 2p).$$

Im symmetrischen Falle ($p = 1/2$) ist $SO_{\{1,2,\dots,n\}} = 0$. Umgekehrt folgt offensichtlich aus $SO_{\{1,2,\dots,n\}} = 0$ die Symmetriebedingung $p = 1/2$. Das normierte Schiefemaß ist

$$SO_{\{1,2,\dots,n\}}^* = 1 - 2p$$

und damit unabhängig von n .

Es soll nun dieses Schiefemaß mit der bekanntesten Alternative, dem dritten standardisierten Moment, verglichen werden. Dieses lautet für die Binomialverteilung bekanntlich (vgl. Mood, Graybill, Boes (1974), S. 539)

$$\beta_3 = (1 - 2p) / \sqrt{np(1-p)}.$$

und unterscheidet sich von dem normierten Schiefemaß $SO_{\{1,2,\dots,n\}}^*$ nur durch den Nenner. β_3 ist jedoch nicht beschränkt und geht mit $p \rightarrow 1$ gegen $-\infty$ bzw. mit $p \rightarrow 0$ gegen ∞ .

Die Schiefemaße SO_I sind nicht anwendbar für die geometrische, die negative Binomial- und die Poisson-Verteilung, da deren Menge möglicher Realisationen nicht nach oben beschränkt ist. Es bleiben mithin vor allem der triviale Fall der diskreten Gleichverteilung und der Hypergeometrischen Verteilung übrig, wenn man von abgeschnittenen Verteilungen und dem großen Bereich der empirischen Verteilungen für diskrete Merkmale absieht.

3.2 Stetiger Fall

Die für ordinalskalierte und quantitative Merkmale eingeführten Schiefemaße SO_I lassen sich auch auf den Fall stetiger Zufallsvariablen mit einem endlichen Intervall $(a, b) = \{x | 0 < F_X(x) < 1\}$ übertragen. Eine stetige Zufallsvariable X sei durch eine stetige Verteilungsfunktion F_X charakterisiert.

Betrachtet man die Differenzfunktion

$$\Delta_{1x}(x) = F_X(a+x) - (1 - F_X(b-x)) = F_X(a+x) + F_X(b-x) - 1 \text{ für } x \in (a, b),$$

dann ist F_X symmetrisch, wenn $\Delta_{1x} = 0$ für $x \in (a, b)$ ist. Im Falle $\Delta_{1x} \geq 0$ für $x \in (a, b)$, wobei für mindestens ein x das echte Größerzeichen gilt, liegt eine rechtsschiefe (linkssteile) Verteilung vor. Unter Verwendung der inversen Verteilungsfunktion

$$F_X^{-1}(u) = \inf\{x | F_X(x) \geq u\}$$

kann vollkommen äquivalent eine Schiefecharakterisierung mit der Differenzfunktion

$$\Delta_{1u}(u) = (a+b) - F_X^{-1}(u) - F_X^{-1}(1-u) \text{ für } u \in (0, 1)$$

vorgenommen werden. So liegt Rechtsschiefe oder Symmetrie vor, wenn $\Delta_{1u}(u) \geq 0$ für $u \in (0, 1)$ ist. Dies ist aber äquivalent zu

$$b - F_X^{-1}(1-u) \geq F_X^{-1}(u) - a \text{ für } u \in (0, 1).$$

D.h. die unteren Quantile liegen bei einer rechtsschiefen Verteilung dichter bei der Untergrenze a als die korrespondierenden oberen Quantile bei der Obergrenze b . Zieht man in der vorstehenden Ungleichung auf beiden Seiten $(b-a)/2$ ab, so erhält man als weitere alternative Formulierung der Rechtsschiefe die Bedingung

$$(a+b)/2 - F_X^{-1}(u) \geq F_X^{-1}(1-u) - (a+b)/2 \text{ für } u \in (0, 1).$$

Wiederum lassen sich drei Klassen von Schiefemaßen betrachten, die nun aber unmittelbar auf den Quantilsdifferenzen ansetzen können. Da diese Quantilsdifferenzen nur translationsinvariant und nicht invariant gegenüber streng monoton zunehmenden linearen Transformationen sind, empfiehlt sich eine Standardisierung der Quantilsdifferenzen mit einem geeigneten Skalenfunktional $S(F_X)$, das nach Bickel und Lehmann (1976) die Eigenschaft

$$S(F_{aX+b}) = S(F_X)$$

für $a > 0$ besitzen muß, um zu Schiefemaßen zu gelangen, die entsprechend der Forderung von Oja (1981) die Eigenschaft der Invarianz bezüglich streng monoton zunehmenden affinen Transformationen besitzen. Es lassen sich zumindest die folgenden drei Klassen von Schiefemaßen unterscheiden:

1. Schiefemaße auf der Basis einzelner Quantile: Es soll als Skalenfunktional der Quantilsabstand $S(F_X) = F_X^{-1}(1-u) - F_X^{-1}(u)$ für $u \in (0, 0.5)$ betrachtet werden, womit

$$SO_u(F_X) = \frac{\Delta_{1u}(u)}{S(F_X)} = \frac{a + b - F_X^{-1}(1-u) - F_X^{-1}(u)}{F_X^{-1}(1-u) - F_X^{-1}(u)} \quad \text{für ein } u \in (0, 0.5)$$

ein mögliches Schiefemaß ist.

2. Schiefemaße auf der Basis von Teilmittelwerten: Nun empfiehlt sich ein Skalenfunktional, daß ebenfalls auf Teilmittelwerten beruht. Ein solches ist z.B. die Differenz von Teilmittelwerten

$$S(F_X) = \mu(1-\beta, 1-\alpha) - \mu(\alpha, \beta)$$

für $0 < \alpha < \beta < 0.5$, wobei ein Teilmittelwert durch

$$\mu(\alpha', \beta') = \frac{1}{\beta' - \alpha'} \int_{\alpha'}^{\beta'} F_X^{-1}(u) du$$

für $0 < \alpha' < \beta' < 1$ definiert ist. Damit läßt sich das Schiefemaß

$$\begin{aligned} SO_{(\alpha, \beta)}(F_X) &= \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} \Delta_{1u}(u) du / S(F_X) \\ &= \frac{(a + b) - \mu(\alpha, \beta) - \mu(1 - \beta, 1 - \alpha)}{\mu(1 - \beta, 1 - \alpha) - \mu(\alpha, \beta)} \end{aligned}$$

für $0 < \alpha < \beta < 0.5$ einführen. Insbesondere ist dann für eine rechtsschiefe Verteilung

$$1/2(\mu(\alpha, \beta) + \mu(1 - \beta, 1 - \alpha)) \leq (a + b)/2$$

für alle $0 < \alpha < \beta < 1$.

3. Schiefemaß unter Verwendung des Mittelwertes $E(X)$: Jetzt könnte man an die Standardabweichung als geeignetes Skalenfunktional denken, so daß sich das Schiefemaß

$$\begin{aligned} SO_{(a, b)}(F_X) &= \int_0^1 \Delta_{1u} du / \sqrt{Var(X)} \\ &= (b + a - 2E(X)) / \sqrt{Var(X)} \end{aligned}$$

ergibt.

Diese Maße unterscheiden sich nur in dem Bereich, in dem die jeweilige Differenzfunktion zur Beurteilung der Schiefe herangezogen wird und sind so konstruiert, daß Symmetrie mit dem Wert 0 einhergeht.

Beispiel 3.2 *Beta-Verteilung*

Eine spezielle Verteilungsfamilie, die über dem kompakten Intervall $[0, 1]$ definiert ist, ist durch die Dichte

$$f_X(x) = \frac{1}{B(p, q)} x^{p-1} (1-x)^{q-1}$$

für $x \in [0, 1]$, $p > 0$, $q > 0$ gegeben. Es handelt sich um die parametrische Familie der Beta-Verteilungen, wobei

$$B(p, q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx$$

die vollständige Betafunktion angibt.

Wegen $b = 1$ und $E(X) = p/(p+q)$ und $Var(X) = pq/((p+q+1)(p+q)^2)$ (siehe Mood, Graybill & Boes (1974), S. 540) ist

$$SO_{(0,1)}(F_X) = \left(1 - 2\frac{p}{p+q}\right) / \sqrt{Var(X)} = (q-p) \sqrt{\frac{p+q+1}{pq}}.$$

Dann ist $SO(F_X) = 0$ genau dann, wenn $p = q$, d.h. die zugehörige Beta-Verteilung symmetrisch ist.

Das dritte standardisierte Moment lautet

$$\beta_3(F_X) = \frac{2(q-p)}{p+q+2} \sqrt{\frac{p+q+1}{pq}}$$

und ist dem Maß $SO_{(0,1)}$ sehr ähnlich.

4 Vergleich mit den gängigen Schiefekonzepten

SK

4.1 Ein allgemeines Konstruktionsprinzip

Es gibt mittlerweile eine umfassende Literatur zur Schiefemessung (siehe z.B. die Referenzen in MacGillivray (1986), Handl (1986), Büning (1991)). Es wird dabei fast ausschließlich die Schiefe von Zufallsvariablen mit stetiger Verteilungsfunktion betrachtet, wie wir es auch im vorangegangenen Abschnitt getan haben. Wir versuchen zunächst, einen gemeinsamen Zugang zu den gängigen Schiefemaßen aufzuzeigen, der ähnlich strukturiert ist wie die Vorgehensweise bei der Herleitung der Maße SO .

Ausgangspunkt sei wiederum eine Definition von Symmetrie, Rechts- und Linkssteilheit, die jetzt aber Bezug auf ein Lokalisationsfunktional $L(F_X)$ als potentiellen Symmetriepunkt nimmt. Ein Lokalisationsfunktional muß nach der Definition von Bickel & Lehmann (1975) mit dem Symmetriepunkt der symmetrischen Verteilung übereinstimmen. Zumeist wird der Median als Lokalisationsfunktional verwendet (siehe z.B. Handl (1986), S. 116f.). Es wird sich aber noch als sinnvoll erweisen, mittlere Quantilsabstände

$$0.5(F_X^{-1}(1-u) - F_X^{-1}(u)) \text{ für } u \in (0, 0.5)$$

oder Teilmittelwerte der Form

$$\mu(\delta, 1-\delta) \text{ für } \delta \in (0, 0.5)$$

als Lokalisationsfunktionale zu betrachten.

Die Ansätze zur Schiefemessung gehen implizit von der folgenden Wahrscheinlichkeitsdifferenzfunktion

$$\Delta_{2x}(x) = 1 - F_X(L(F_X) - x) - F_X(x + L(F_X)) \text{ für } x \in \mathbb{R}$$

aus und definieren die stetige Verteilungsfunktion F_X als symmetrisch, wenn $\Delta_{2x}(x) = 0$ für alle $x \in \mathbb{R}$ gilt und als rechtssteil (linksschief), wenn

$$\Delta_{2x}(x) \geq 0 \text{ für alle } x \in \mathbb{R} \text{ und } \Delta_{2x}(x) > 0 \text{ für mindestens ein } x \in \mathbb{R}$$

ist.

Betrachtet man die zugehörige Quantilsdifferenzenfunktion

$$\Delta_{2u}(u) = F_X^{-1}(u) + F_X^{-1}(1-u) - 2L(F_X) \text{ für } u \in (0, 1),$$

so ergibt sich als äquivalente Charakterisierung von Schiefe und Symmetrie, daß F_X symmetrisch ist, wenn $\Delta_{2u}(u) = 0$ für alle $u \in (0, 1)$ und rechtssteil (linksschief), wenn

$$\Delta_{2u}(u) \geq 0 \text{ für alle } u \in (0, 1) \text{ und } \Delta_{2u}(u) > 0 \text{ für mindestens ein } u \in (0, 1)$$

gilt.

Diese Quantilsdifferenzenfunktion vergleicht jeweils die Abstände des (unteren) u -Quantils und des (oberen) $(1-u)$ -Quantils zum Lokalisationsfunktional miteinander, wobei Rechtsschiefe zu vermuten ist, wenn das untere Quantil vom Lokalisationsfunktional höchstens so weit entfernt ist wie das obere Quantil. Diese Quantilsdifferenzfunktion ist wiederum nur invariant gegenüber Translationen. Die Invarianz gegenüber streng monoton zunehmenden linearen Transformationen läßt sich wiederum erreichen, wenn durch ein geeignetes Skalenfunktional $S(F_X)$ dividiert wird.

Ausgehend von einer geeignet standardisierten Quantilsdifferenzenfunktion lassen sich nahezu alle bekannten Schiefemaße herleiten:

1. Schiefemaße auf der Basis einzelner Quantile: Betrachtet man Δ_{2u} für ein fixiertes u als Schiefemaß, so erhält man für den Median als Lokalisationsfunktional und den Quantilsabstand als Skalenfunktional eine Klasse von Schiefemaßen, die von Groeneveld & Meeden (1984) vorgeschlagen wurden und für $u = 0.25$ das von Bowley (1920) diskutierte Maß

$$SK_{0.25} = \frac{F^{-1}(0.25) + F^{-1}(0.75) - 2F^{-1}(0.5)}{F^{-1}(0.75) - F^{-1}(0.25)}$$

als Spezialfall enthält.

2. Schiefemaße auf der Basis von Teilmittelwerten: Integriert man $\Delta_{u2}/(\beta - \alpha)$ nur über ein Teilintervall (α, β) , so erhält man

$$SK_{(\alpha, \beta)} = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} \Delta_{2u}(u) du / S(F_X) = \frac{\mu(\alpha, \beta) + \mu(1 - \beta, 1 - \alpha) - 2L(F_X)}{S(F_X)}$$

für ein Lokalisationsfunktional $L(F_X)$ und ein Skalenfunktional $S(F_X)$. Verwendet man als Lokalisationsfunktional den Teilmittelwert $\mu(\delta, 1 - \delta)$, so ergibt sich der Zähler eines Maßes, das sehr viel Ähnlichkeit mit den von Hogg (1982) vorgeschlagenen Schiefemaßen besitzt.

3. Schiefemaße auf der Basis von Mittelwerten: Integriert man Δ_{u2} über das gesamte Intervall $(0, 1)$, verwendet den Median als Lokalisationsfunktional und die Standardabweichung als Skalenfunktional, so ergibt sich

$$SK_{(0,1)} \int_0^1 \Delta_{2u}(u) du / \sqrt{Var(X)} = 2(E(X) - F_X^{-1}(0.5)) / \sqrt{Var(X)}$$

das Schiefemaß von Pearson (vgl. Oja (1981), S. 164).

5 Schiefemaße auf Quotientenbasis (SKQ - und SOQ -Maße)

Alle bislang betrachteten Schiefemaße waren derart konstruiert, daß der Wert 0 den Fall der Symmetrie repräsentiert. Nun gibt es eine Reihe von Vorschläge für Schiefemaße (siehe z.B. Handl (1986)), die im Falle der Symmetrie den Wert 1 annehmen. Für das Schiefemaß SK_u läßt sich sehr einfach eine Quotientendarstellung einführen, wie am Beispiel des Maßes $SK_{(\alpha, \beta)}$ gezeigt werden soll. Es ist

$$SK_{(\alpha, \beta)} \begin{cases} < 0 \\ = 0 \\ > 0 \end{cases} \iff \frac{\mu(1 - \beta, 1 - \alpha) - L(F_X)}{L(F_X) - \mu(\alpha, \beta)} \begin{cases} < 1 \\ = 1 \\ > 1 \end{cases}$$

so daß alternativ das Schiefemaß

$$SKQ_{(\alpha,\beta)} = \frac{\mu(1-\beta, 1-\alpha) - L(F_X)}{L(F_X) - \mu(\alpha, \beta)}$$

betrachtet werden kann. Für das Lokalisationsfunktional

$$L(F_X) = \mu(\delta, 1 - \delta)$$

ergeben sich die von Hogg, Fisher & Randles (1975) betrachteten Schiefemaße. Sie setzen konkret $\alpha = 0$, $\beta = 0.05$ und $\delta = 0.25$. Für sämtliche betrachteten Schiefemaße der Klasse *SO* und *SK* lassen sich entsprechende Quotientendarstellungen *SKQ* und *SOQ* finden, da jeweils Abstände zu Lokalisationsmaßen (SK-Maße) bzw. zur Intervallmitte $(a + b)/2$ (SO-Maße) betrachtet werden.

6 Zusammenfassung und Ausblick

Wir haben ein allgemeines Bildungsgesetz aufgeklärt, das nahezu sämtlichen in der Literatur vorgeschlagenen Schiefemaßen mehr oder weniger explizit zugrundeliegt. Es basiert auf Differenzfunktionen, wobei Differenzen von Wahrscheinlichkeiten oder Differenzen von Quantilen zu betrachten sind. Werden statt Differenzen Quotienten eingeführt, so ergeben sich die restlichen Maßzahlen. Für ordinalskalierte Merkmale sind Differenzen von Quantilen kein geeigneter Ansatzpunkt zur Messung von Verteilungseigenschaften (außer der Lage) darstellen, da diese durch zulässige streng monoton zunehmenden Skalentransformationen beliebig verändert werden können. Deshalb liegt es nahe, für diese über die Differenzen von Wahrscheinlichkeiten zu Schiefemaßen zu gelangen. Diese Art der Vorgehensweise läßt sich durch eine in Klein (1994) allgemein gezeigte Invarianzeigenschaft via Maximalinvarianten leicht begründen.

Da einige aus der Literatur bekannte Schiefemaße hier in leicht modifizierter Form angegeben wurden, bleibt es weiteren Arbeiten überlassen, nachzuweisen, ob und unter welchen Bedingungen diese Maßzahlen z.B. die von Oja (1981) oder ähnlich Handl (1986) formulierten axiomatischen Anforderungen erfüllen.

7 Literatur

1. Bickel, P.J. & Lehmann, E.L. (1975). Descriptive statistics for nonparametric models. I. Introduction. II. Location. *Annals of Statistics* **3**, 1038–1069.
2. Bickel, P.J. & Lehmann, E.L. (1976). Descriptive statistics for nonparametric models. III. Dispersion. *Annals of Statistics* **4**, 1139–1158.
3. Bowley, A.L. (1920). *Elements of Statistics*. Charles Scribner's Sons, New York.

4. Büning, H. (1991). *Adaptive und robuste Tests*. deGruyter, Berlin.
5. Dabrowska, D. (1985). Descriptive Parameters of Location, Dispersion and Stochastic Dependence. *Statistics* **16**, 63–88.
6. Groeneveld, R.A. & Meeden, G. (1984). Measuring skewness and kurtosis. *The Statistician* **33**, 391–399.
7. Handl, A. (1986). *Maßzahlen zur Klassifizierung von Verteilungen bei der Konstruktion adaptiver verteilungsfreier Tests im unverbundenen Zwei-Stichproben-Fall*. Unveröffentlichte Dissertation, FU Berlin.
8. Hogg, R.V. (1974). Adaptive Robust Procedures. A Partial Review and Some Suggestions for Further Applications and Theory. *Journal of the American Statistical Association* **69**, 909–927.
9. Hogg, R.V., Fisher, D.M. & Randles, R.H. (1975). A two sample adaptive distribution-free test. *Journal of the American Statistical Association* **70**, 656–661.
10. Hogg, R.V. (1982). On Adaptive Statistical Inference. *Communications in Statistics, Theory and Methods* **11**, 2351–2542.
11. Klein, I. (1984). *Das Problem der Auswahl geeigneter Maßzahlen in der deskriptiven Statistik*. Physica, Würzburg.
12. Klein, I. (1994). *Mögliche Skalentypen, invariante Relationen und wissenschaftliche Gesetze*. Vandenhoeck & Ruprecht, Göttingen.
13. MacGillivray, H.L. (1986). Skewness and Asymmetry: Measures and Orderings. *Annals of Statistics* **14**, 994–1011.
14. Mood, A.S., Graybill, F.A. & Boes, D.C. (1974). *Introduction to the theory of statistics*. McGraw–Hill, Auckland.
15. Oja, H. (1981). On Location, Scale, Skewness and Kurtosis of Univariate Distributions. *Scandinavian Journal of Statistics* **8**, 154–168.
16. Pfanzagl, J. (1971). *Theory of Measurement*. Physica, Würzburg.
17. van Zwet, W.R. (1979). Mean, Median, Mode II. *Statistica Neerlandica* **33**, 1-5.
18. Vogel, F. & Dobbener, R. (1982). Ein Streuungsmaß für komparative Merkmale. *Jahrbücher für Nationalökonomie und Statistik* **197**, 145-157.