Prof. Dr. Peter Michael von der Lippe

Grenzwertsätze, Gesetze der Großen Zahl(en)

Der folgende Text ist **gedacht als Begleitlektüre zu meiner Vorlesung "Induktive Statistik"** in einem Punkt, nämlich dem **Kapitel 7 der Vorlesung** und meines Buches[1], das erfahrungsgemäß besonders viele Verständnisprobleme bereitet. Was in der Vorlesung gesagt wird (wurde) ist deshalb hier auch in ausformulierter Form nachzulesen. Der Text ist wie folgt gegliedert:

1. Nichtstochastische Begriffe von "Grenzwert"

a) Konzept der Epsilon-Umgebung

Aus der Schulmathematik ist bekannt, dass unterschieden wird

Konvergenz einer Zahlenfolge, oder "algebraische Folge" (Z) und

Konvergenz einer Funktionenfolge (F).

Die Unterscheidung von Z und F taucht auch in der Statistik auf in Form von Gesetzen der Großen Zahl(en) ($\Rightarrow$ Z) und Grenzwertsätzen (besser: Grenzverteilungssätzen, $\Rightarrow$ F).

Ein einfaches Beispiel für eine **Zahlenfolge** (**Z**) in Abhängigkeit von n ist die Folge

$$Z(n) = 0{,}1 + \frac{1}{n} ,$$

die "gegen 0,1 strebt" (vgl. Bild 1 auf der nächsten Seite).

Die übliche anschauliche Vorstellung dabei ist:

> je größer n ist, desto näher kommt der Wert der Zahl 0,1. Oder: mit n "gegen unendlich" wird die Folge (Z) der Konstanten 0,1 immer ähnlicher.

Es ist wichtig, sich klar zu machen, dass diese verbreitete "Alltagsvorstellung" zumindest unexakt wenn nicht mißverständlich ist. Sie soll im folgenden "verfeinert" werden. Die im Kasten wiedergegebene Sprechweise ist laienhaft und problematisch, weil die verwendeten Begriffe nicht exakt sind.

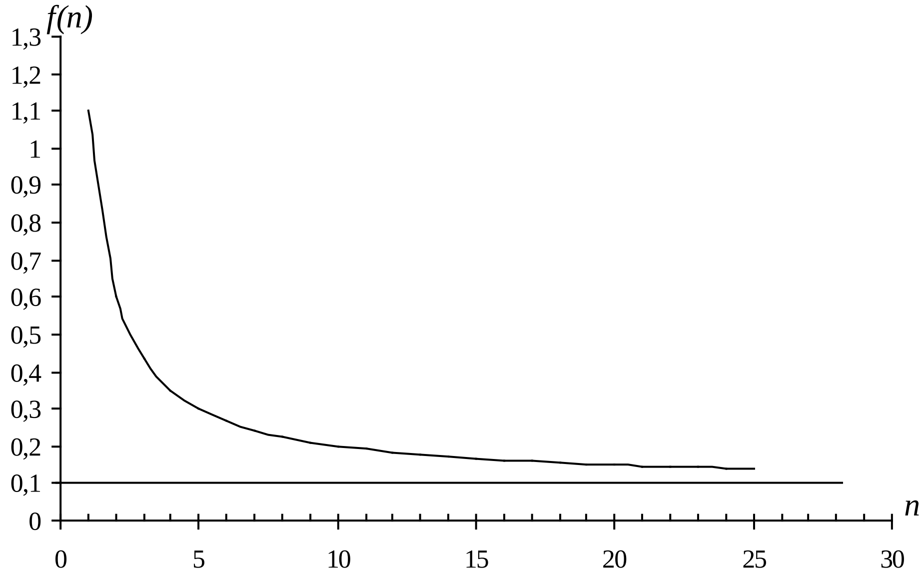
---

[1] Gemeint ist das Buch von der Lippe, Induktive Statistik, Formeln, Aufgaben, Klausurtraining, Oldenbourg Verlag, München u. Wien 1999, im folgenden zitiert als "v.d.L., Induktive".

Was heißt "je größer" oder immer "ähnlicher werden" oder gegen etwas "streben"? Was heißt vor allem "unendlich"?

Ein Begriff, der das vage Konzept "gegen unendlich" vermeidet, konnte erst gefunden werden, als die Idee der ε-Umgebung (Epsilon-Umgebung) eingeführt wurde.

Bild 1



Der Begriff

(1) $\lim\limits_{n\to\infty} Z(n) = \lim\limits_{n\to\infty}(0{,}1 + 1/n) = 0{,}1$

kann dann auch wie folgt ausgedrückt werden:

(1a) $\bigwedge\limits_{\varepsilon>0} \bigvee\limits_{N_\varepsilon} \bigwedge\limits_{n>N_\varepsilon} \left|\{0{,}1 + 1/n\} - 0{,}1\right| < \varepsilon.$

oder in Worten: es gibt (Symbol $\vee$) eine Zahl N_ε (deren Größe von ε abhängt, daher N_ε) so dass für alle[2] n (Symbol $\wedge$ oder $\forall n$), die größer sind als N_ε (also $n > N_\varepsilon$) die Werte der Folge jeweils *in* der ε-Umgebung liegen (davor, also für $n \le N_\varepsilon$ werden sie außerhalb der ε-Umgebung liegen). Sie liegen damit quasi in einem Band ober- und unterhalb von 0,1. Das "Liegen" in einem Band um 0,1 wird ausgedrückt durch $|Z_n - 0{,}1| < \varepsilon$, wobei ε beliebig groß oder klein sein kann, es muss nur eine reelle positive Zahl sein. Das Zeichen $\bigwedge\limits_{\varepsilon>0}$ bedeutet einfach "für *jedes* (beliebige) positive ε" gibt es ein N_ε, so dass ... gilt.

Für unser Beispiel sei angenommen ε = 0,05, dann liegt die ε-Umgebung um 0,1 von 0,05 bis 0,15 und ab einem Wert von N_ε = 20 liegen alle Werte der Folge innerhalb den beiden Grenzen der ε-Umgebung, nämlich 0,15 und 0,05 (sie liegen hier in diesem speziellen Fall sogar zwischen 0,15 und 0,1, weil sich die Folge "monoton" dem Grenzwert nähert, durch ein monotones Abnehmen).

---

[2] Genau genommen muß es heißen: innerhalb der ε–Umgebung liegen *"fast alle"* Werte der Folge Z(n) in dem Sinne von alle bis auf *endlich* viele.

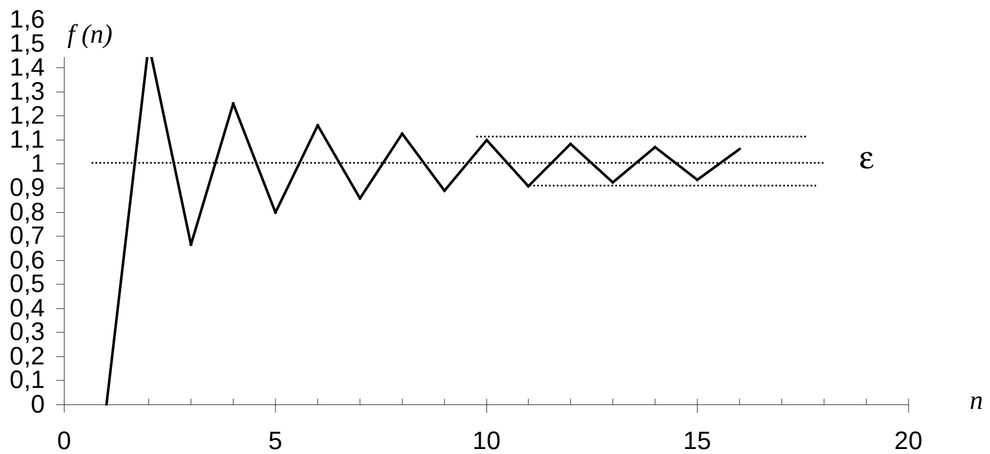
Man erhält die folgenden Zahlen für die Werte Z_{21}, Z_{22}, Z_{23}, Z_{24} der Zahlenfolge in Bild 1:

n	21	22	23	24
0,1 + 1/n	0,14762	0,14545	0,14348	0,14167

Eine Folge muss sich nicht notwendig *monoton* ihrem Grenzwert nähern, sie kann dies auch *oszillierend* (schwingend) tun. Ein Beispiel hierfür ist die Zahlenfolge

(1b) $Z_n = f(n) = 1+(-1)^n \frac{1}{n}$ oder $\frac{n+(-1)^n}{n}$ (vgl. das folgende Bild 2).

Bild 2



Ein einfaches Beispiel für eine **Funktionenfolge** (**F**) im Unterschied zu einer Zahlenfolge (Z) ist die Funktion $F(n) = \left(1+\frac{x}{n}\right)^n$

Sie hat folgende Gestalt:

n	F(n)
1	1+x
2	$\left(1+\frac{1}{2}x\right)^2 = \frac{1}{4}x^2 + x + 1$
3	$\left(1+\frac{1}{3}x\right)^3$

usw. Bei n = 1 ist dies eine Gerade. Dann kommt eine Parabel (n = 2) und damit eine Krümmung, die in einem gewissen Sinne immer stärker wird. Wohin "strebt" nun diese Folge von Funktionen F(1), F(2), F(3), ... mit wachsendem n (wenn n immer größer wird)?

Man kann zeigen, dass die Folge gegen eine feste (konstante, nicht mehr von n abhängige) Funktion strebt, die quasi bei $n \to \infty$ gilt. Es ist die e-Funktion, so dass wegen

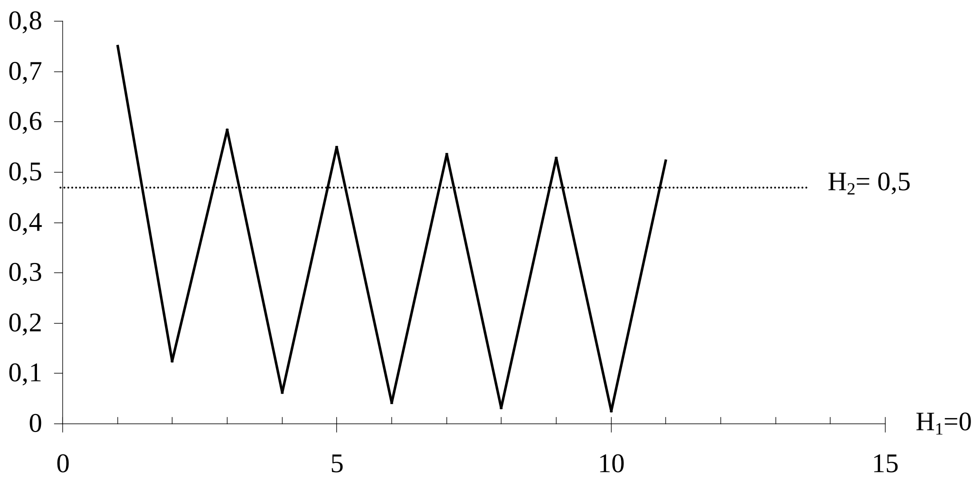
$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x$$

e^x die Grenzfunktion (asymptotische Funktion) der Funktionenfolge $F(n) = \left(1+\frac{x}{n}\right)^n$ ist.

b) Grenzwert und Häufungspunkt

Für einen Grenzwert ist es notwendig, dass alle (unendlich viele) Werte ab einem bestimmten N_ε in der ε-Umgebung um einen Punkt (den Grenzwert) liegen. Es darf nur *einen* solchen Punkt geben. Betrachtet man z.B. die Folge $\frac{1}{4n}\left[n-(-1)^n n+1\right]$ (vgl. Bild 3), so streben bei *gerad*-zahligem n (also 2, 4, 6,...) die Werte gegen 0 (weil $\frac{1}{4n} \to 0$) und bei *ungerad*-zahligem n alle Werte gegen $\frac{2n}{4n}+\frac{1}{4n}=\frac{1}{2}+\frac{1}{4n}$ also gegen $\frac{1}{2}$.

Bild 3



Die Folge hat somit zwei Häufungspunkte $H_1 = 0$ und $H_2 = 1/2$ und deshalb keinen Grenzwert. Es mag verwirrend klingen, dass es zweimal unendlich gibt: unendlich viele Werte die sich 0 nähern (besser "nahe bei 0 sind", im Sinne von "in einer ε-Umgebung um 0") und auch unendlich viele Werte, die sich ½ nähern.

Weil die Folge zwei Häufungspunkte hat, hat sie keinen Grenzwert (sie konvergiert nicht), denn ein Grenzwert ist praktisch "der einzige Häufungspunkt"[3]. Eine Folge kann unendlich viele Häufungspunkte haben, aber nur höchstens einen Grenzwert.

Die Zahlenfolge

$$Z(n)\begin{cases} n, \text{wenn n ungerade} \\ 1/n, \text{wenn n geradzahlig} \end{cases}$$

hat die Werte $1, \frac{1}{2}, 3, \frac{1}{4}, 5, \frac{1}{6}$ usw.[4] und keinen Grenzwert, weil es zwar unendlich viele Werte gibt die gegen Null streben ($\frac{1}{2}, \frac{1}{4}, \frac{1}{6}, ...$), also gegen einen festen Wert, aber auch unendlich

[3] Existiert ein Grenzwert (*ein* und nur ein Häufungspunkt), so sagt man, die Folge "konvergiert". Ein wichtiges Thema in der Mathematik ist es, Kriterien herauszufinden, die darüber entscheiden, ob eine Folge konvergiert oder divergiert, und wenn sie konvergiert, wie schnell sie konvergiert.

[4] Es ist keine Schwierigkeit, das entsprechende Bild (das auch in der Vorlesung gezeigt wird) zu zeichnen.

viele Werte die davon wegstreben und immer größer werden, also nicht gegen einen festen Wert "streben". Das sind die ungeradzahligen Werte 1, 3, 5, ...

Man kann sich auch vorstellen, dass der Grenzwert nicht erst allmählich, sondern zumindest zum Teil schon sofort erreicht wird. Die Zahlenfolge[5]

$$Z(n) = \frac{n-(-1)^n+1}{4n} \text{ oder äquivalent } Z(n) = \begin{cases} \frac{1}{4} & \text{für gerades n} \\ \frac{n+2}{4n} & \text{für ungerades n} \end{cases}$$

ist eine Modifikation von Bild 3. Der Grenzwert $\frac{1}{4}$ wird für geradzahlige Werte von n sofort (schon ab n = 2) erreicht, für ungerades n aber erst mit wachsendem n, weil $\frac{n+2}{4n} = \frac{1}{4} + \frac{1}{2n}$ gegen $\frac{1}{4}$ strebt.[6]

2. Stochastische Begriffe

Den beiden Begriffen Z und F entsprechen in der Statistik (genauer Stochastik, d.h. in der Wahrscheinlichkeitsrechnung) zwei Konvergenzbegriffe

- stochastische Konvergenz (Gesetze der großen Zahl[7], Z) einer Zufallsvariable (bzw. einer Folge von Zufallsgrößen, etwa dem Ergebnis einer wiederholten Durchführung eines Zufallsexperiments) statt einer Zahlenfolge (Zahlen im Sinne von nichtzufälligen algebraischen Größen) und
- Konvergenz von Verteilungen (Grenzwertsätze).

Man betrachtet hier also eine Folge von Zufallsvariablen bzw. eine Folge von Wahrscheinlichkeitsverteilungen.

a) Stochastische Konvergenz

Es ist allgemein bekannt, dass bei einer echten (idealen) Münze die relative Häufigkeit der Wappenwürfe gegen die Wahrscheinlichkeit ½ streben wird. In diesem Sinne ist die Wahrscheinlichkeit eines Ereignisses die relative Häufigkeit, die man erhielte, wenn man das Zufallsexperiment (z.B. Werfen einer Münze) unendlich oft unter gleichen Bedingungen wiederholt. Man könnte nun meinen, dass mit dem "Gesetz der großen Zahl" gemeint ist, dass die relative Häufigkeit mit wachsendem n immer ähnlicher der Zahl ½ wird. Angenommen, man werfe n mal und erhalte x mal Wappen

n	10	100	1000	...
x	6	58	493	...

die relativen Häufigkeiten wären dann 0,6, 0,58, 0,493 usw., und es ist plausibel anzunehmen, dass sie immer mehr dem Wert 0,5 nahe kommen. Das ist aber mit "stochastischer Konvergenz" *nicht* gemeint.

Es ist wichtige, auch hier, sich von plausibel klingenden Alltagsvorstellungen freizumachen:

---

[5] Für das Bild, wie diese Zahlenfolge aussieht gilt Fußnote 4 entsprechend.

[6] Wir verlassen damit den Abschnitt, der sich mit einer Wiederholung der, bzw. einer Anknüpfung an die Schulmathematik beschäftigt. Bewußt nicht behandelt wurden Dinge, die für unsere Zwecke nicht bedeutsam sind, wie z.B. das Konzept der "beschränkten" Folge, Begriffe wie Supremum und Infimum, Konvergenzkriterien usw.

[7] Man findet auch die Formen wie: "Gesetze der **G**roßen Zahl", oder "der großen Zahl**en**"

Stochastische Konvergenz (z.B. der relativen Häufigkeit $h_n = x/n$ gegen die Wahrscheinlichkeit $\pi = ½$) bedeutet **nicht**, dass "h_n gegen π strebt", in dem Sinne, dass für $n \geq N_\varepsilon$ alle Werte von h_n in einer ε-Umgebung um π liegen[8] **sondern**, dass die Wahrscheinlichkeit dafür, dass h_n in einer ε-Umgebung um π liegt gegen 1 strebt. Der Begriff des Grenzwerts kann wegen der Zufälligkeit von X *nicht auf h_n* angewendet werden, *sondern* er bezieht sich *auf eine Wahrscheinlichkeitsaussage über h_n*.

Unter "Stochastischer Konvergenz" versteht man, dass die Wahrscheinlichkeit dafür, dass die relative Häufigkeit (im Falle des Münzwurfs, also bei $\pi = ½$) in einer ε-Umgebung um ½ liegt (und bleibt) gegen 1 strebt, oder in Symbolen

(2) $$\lim_{n\to\infty} P\left(\left|\frac{x}{n}-\frac{1}{2}\right| < \varepsilon\right) = 1$$

oder allgemein bei stochastischer Konvergenz einer von n abhängigen Zufallsvariablen X_n gegen die Konstante c gilt

(2a) $\lim_{n\to\infty} P(|X_n - c| < \varepsilon) = 1$ oder plim $X_n = c$, oder n Worten:

der Wahrscheinlichkeitslimes (sprich auch: "plim") von X_n ist die Konstante c.

So gilt etwa im Fall des Münzwurfs (nur als Beispiel) mit der Zufallsvariable $X_n = X/n$ und ihrer Realisation $x_n = h_n = x/n$ und $c = \pi = 1/2$[9].

Man kann dieses Wirken des Gesetzes der großen Zahl (in diesem konkreten Falle ist es der Satz von Bernoulli [siehe unten]) auch leicht zahlenmäßig zeigen.

Die Anzahl und damit auch der Anteil der Wappenwürfe bei n Versuchen ist binomialverteilt mit den Parametern n und $\pi = 1/2$. Für die Wahrscheinlichkeit von genau x "Erfolgen" (Wappen) bei n Versuchen (oder: n Wiederholungen des Zufallsexperiments) erhält man danach[10] die Formel $\binom{n}{x}\pi^x(1-\pi)^{n-x}$. Etwa bei n = 3 erhält man für x = 1 die Wahrscheinlichkeit $\binom{3}{1}\frac{1}{2}^1 \cdot \frac{1}{2}^2 = 3 \cdot \frac{1}{2}^3 = \frac{3}{8} = 0{,}375$.

Man kann also leicht ausrechnen, mit welcher Wahrscheinlichkeit die relative Häufigkeit Werte zwischen 0,25 und 0,75 (einschließlich) annimmt, d.h. in der folgenden ε-Umgebung (mit $\varepsilon = 0{,}25$) liegt: $\left|\frac{x}{n} - 0{,}5\right| \leq 0{,}25$.

[8] Es ist ja gerade der "Witz" der Zufälligkeit einer Zufallsvariable X, dass man nie ausschließen kann, dass $h_n = x/n$ nicht auch außerhalb dieser Umgebung liegen könnte, selbst dann wenn n sehr groß ist. Man kann also keinen Wert N_ε angeben ab dem *alle* Werte der Folge h_n ($n \geq N_\varepsilon$) *mit Sicherheit* innerhalb der ε-Umgebung liegen werden. Man kann nur sagen, dass es nicht sonderlich wahrscheinlich sein dürfte, dass bei einem großen Wert von n (bei $n \to \infty$) so ungewöhnlich hohe oder niedrige Werte von h_n, die oberhalb oder unterhalb der Grenzen der ε-Umgebung um π liegen, auftreten.

[9] Gl. 2 und 2a sind äquivalente Schreibeweisen der gleichen Sache, so wie es auch Gl. 1 und 1a sind.

[10] d.h. wegen der Binomialverteilung.

Bei n = 3 sind es die relativen Häufigkeiten 1/3 und 2/3, die in dem Intervall $0,25 \leq X/n \leq 0,75$ liegen und die mit Wahrscheinlichkeit von jeweils 3/8 - zusammen also mit 6/8 = 0,75 – auftreten, so dass die gesuchte Wahrscheinlichkeit 0,75 beträgt.

Bei n = 4 sind es die relativen Häufigkeiten ¼, ½ und ¾ die relevant sind, und die mit Wahrscheinlichkeit von jeweils ¼ und 3/8 - zusammen also[11] mit 7/8 = 0,875 - auftreten, d.h. es gilt:

x/n	0/4 = 0	1/4 = 0,25	2/4 = 0,5	3/4 = 0,75	4/4 = 1
P(x/n)	1/16	4/16 = 1/4	6/16 = 3/8	4/16 = 1/4	1/16

Bei n = 4 gilt also $P\left(\left|\frac{x}{n}-\frac{1}{2}\right| \leq \frac{1}{4}\right) = 0{,}875$ bei. Die entsprechende Wahrscheinlichkeit ist dann bei n = 5 Münzwürfen

x/n	0/5 = 0	1/5 = 0,2	2/5 = 0,4	3/5 = 0,6	4/5 = 0,8	5/5 = 1
P(x/n)	1/32	5/32	10/32	10/32	5/32	1/32

20/32 = 0,625 und damit allerdings kleiner als 0,875. Aber schon bei n = 6

x/n	0	0,167	0,333	0,5	0,667	0,833	1
P(x/n)	1/64	6/64	15/64	20/64	15/64	6/64	1/64

ist die gesuchte Wahrscheinlichkeit mit 50/64 = 0,78125 wieder etwas größer als 0,625 und bei n = 7 erhält man aufgrund der Tabelle 1 in v.d.L., Induktive, S. 117 für die gesuchte Wahrscheinlichkeit $P(2/7 \leq X/n \leq 5/7) = 0{,}875$ da ja 2/7 = 0,2857 > 0,25 und 5/7 = 0, 7143 < 0,75. Die Summe der Wahrscheinlichkeiten für x = 2/7, 3/7 usw. also 0,1641, 0,2734 usw., ist insgesamt 0,875 (oder 112/128). Bei n = 8 Münzwürfen erhält man für die gesuchte Intervall wahrscheinlichkeit aufgrund der genannten Tabelle den Wert: $1-2\cdot(0{,}0039+0{,}0313)=$ 0,9296 oder genauer 238/256 = 0,9297. Sieht man von den Fällen n = 3 und n = 4 ab, so strebt die Wahrscheinlichkeit $P_n = P(0{,}25 \leq X/n \leq 0{,}75)$ monoton gegen 1, denn es gilt

n	2	3	4	5	6	7	8
P_n	0,5	0,75	0,875	0,625	0,78125	0,875	0,9297

d.h. der feste Wert $\pi = 1/2$ ist der Wahrscheinlichkeitslimes der relativen Häufigkeit x/n, wie es auch nach dieser Fassung des **(schwachen) Gesetzes der Großen Zahl** (d.h. nach dem Theorem von Bernoulli) sein soll.

Der Unterschied zwischen $\lim_{n\to\infty} P\left(\left|\frac{x}{n}-\frac{1}{2}\right| < \varepsilon\right) = 1$, dem "schwachen" Gesetz und

(2b) $$P\left(\lim_{n\to\infty} \frac{x}{n} = \frac{1}{2}\right) = 1,$$

dem **"starken" Gesetz** ist mehr von rein theoretischem Interesse. Nach Gl. 2b gibt es ein N_ε, so dass ab $n \geq N_\varepsilon$ alle Elemente der Folge x/n (zufälliger Werte) sich wie die Werte einer (algebraischen, deterministischen) Zahlenfolge verhalten und fast sicher in der ε-Umgebung um ½ liegen.[12]

[11] angedeutet durch Schattierung.

[12] Fälle, in denen das schwache, nicht aber zugleich auch das starke Gesetz der Großen Zahl gilt, sind selten und sehr konstruiert. Vgl. das Lehrbuch von M. Fisz für ein Beispiel.

b) Konvergenz von Verteilungen

Für die Statistik ist es von großer Bedeutung, ob es eine Grenzverteilung (asymptotische Verteilung) gibt und wenn ja unter welchen Voraussetzungen das der Fall ist und von welcher Gestalt diese Grenzverteilung ist.

Das Konzept der Grenzverteilung ist in der Vorlesung an einer Stelle bereits eingeführt worden, nämlich im Falle des Grenzübergangs von der Binomial- zur Poissonverteilung.[13]

Eine Folge von Binomialverteilungen $B(n_i,\pi_i)$, also $B(n_1,\pi_1)$, $B(n_2,\pi_3)$, ... "strebt gegen" die Poissonverteilung mit dem Parameter $\lambda = n_i\pi_i$ (1 = 1, 2, ...), wenn n zunimmt ($n_1 < n_2 < ...$) und π beständig abnimmt ($\pi_1 > \pi_2 > ...$) und das Produkt konstant ist, $n_1\pi_1 = n_2\pi_2 = ... = \lambda$.

Im folgenden geht es uns nicht nur darum, zu zeigen,

- dass sehr häufig in der Statistik die Normalverteilung als "Grenzverteilung" auftritt und warum das so ist, sondern auch darum, dass man
- zumindest bei großem n (Stichprobenumfang) viel aussagen kann über die Gestalt der Verteilung aller möglichen Stichprobenergebnisse (d.h. über die asymptotische Stichprobenverteilung); natürlich nur *wenn* es sich überhaupt um "Stichproben" (also eine Zufallsauswahl) handelt.[13]

Es ist klar, dass man bei einer konkreten Stichprobe, vor allem wenn ihr Umfang n klein ist, wenig darüber aussagen kann, was eventuell "herauskommen" wird. Es liegt ja im Wesen einer *Zufall*s-auswahl, dass das Ergebnis schwanken kann, und dass man nicht im vorhinein wissen kann, was man als Ergebnis erhalten wird. Trotzdem kann man aber aufgrund von sog. Grenzwertsätzen (eigentlich Sätze über die Grenzverteilung) sagen, mit welcher *Wahrscheinlichkeit* man bestimmte Ergebnisse erhalten wird, *wenn* der Stichprobenumfang hinreichend groß ist (theoretisch "asymptotisch", d.h. bei $n \to\infty$, praktisch jedoch schon meist [Faustregel] ab n = 30 oder n = 50).

Das ist praktisch gesehen von unschätzbarem Wert und es ist zugleich der theoretische Hintergrund für alles was in der Vorlesung folgt, d.h. für die "Inferenz" (Schätzen und Testen) auf der Basis von Stichproben.

Man kann danach insbesondere etwas aussagen über die Wahrscheinlichkeitsverteilung bezogen auf *alle* insgesamt $\binom{N}{n} = \frac{N!}{n!(N-N)!}$ möglichen (denkbaren) Stichproben vom Umfang n, die man aus einer Grundgesamtheit vom Umfang N ziehen kann, bzw. könnte (oder wie es in der Statistik heißt: man kann Aussagen über die "Stichprobenverteilung" machen).

Man kann z.B. sagen (Zentraler Grenzwertsatz):

Summen $\left(\sum x\right)$ und Durchschnitte $\left(\bar{x} = \sum x / n\right)$ sind unter sehr allgemeinen, praktisch immer gültigen Voraussetzungen asymptotisch (bei $n \to \infty$) normalverteilt.

[13] Neben der Poissonverteilung als Grenzverteilung ist natürlich schon an früher Stelle der Vorlesung (Kap. 5 und 6) erkennbar, dass z.B. sehr oft die Normalverteilung als Grenzverteilung auftreten dürfte. Intuitiv sehr einleuchtend ist das z.B. im Fall der Binomialverteilung mit $\pi = ½$.

[13] Man kann nicht genug betonen, dass man die Wahrscheinlichkeitsrechnung nur anwenden kann, wenn man eine *Zufalls*auswahl hat. Ohne Zufall keine Wahrscheinlichkeitsrechnung. Diese simple Weisheit scheint immer gerne vergessen zu werden. Man mag nichtzufälligen Auswahlverfahren ohne Ende die verschiedensten positive Eigenschaften nachsagen, bloß man kann dies leider nicht auf der Basis wahrscheinlichkeitstheoretischer Überlegungen tun.

Die überragende Rolle der Normalverteilung in der Statistik kommt daher, dass sie in sehr vielen Fällen, wie z.B. hier als "Grenzverteilung" auftritt.

Dass Summen und Durchschnitte asymptotisch normalverteilt sind, lässt sich ebenfalls leicht plausibel machen. Beim einmaligen (n = 1) Werfen eines Würfels ist die Augensumme (hier identisch mit dem Durchschnitt, weil n = 1) wie folgt verteilt:

$x = \bar{x}$	1	2	3	4	5	6
f(x)	1/6	1/6	1/6	1/6	1/6	1/6

Man erhält das Bild einer (diskreten) *Gleich*verteilung (mit jeweils gleichen Höhen von 1/6).

Beim zweimaligen Werfen (n=2) gilt entsprechend:

$\bar{x}$	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6
$36 \cdot f(\bar{x})$	1	2	3	4	5	6	5	4	3	2	1

Man erhält eine (diskrete) *Dreiecks*verteilung (genauer Simpson-Verteilung) mit zunächst zunehmenden Höhen (von 1/36 bis 6/36) und dann wieder abnehmenden Höhen (von 6/36 bis 1/36). Man kann sich leicht vorstellen, dass sich das Bild mit wachsendem n immer mehr dem der Normalverteilung annähert (vgl. Abb. 7.3 in v.d.L., Induktive Seite 75f).[14]

Interessant ist dabei vor allem auch, dass man für das arithmetische Mittel $\bar{X}$ (was ja eine Zufallsvariable ist) eine Normalverteilung erhält, egal wie die Zufallsvariable X selbst in der Grundgesamtheit verteilt ist (im Falle der durchschnittlichen Augenzahl beim Würfeln ist es z.B. die obige Gleichverteilung).

Die wichtigsten Varianten des Gesetzes der Großen Zahl und der Grenzwertsätze und deren Voraussetzungen sind dargestellt in v.d.L., Induktive Seite 73 – 76. Sie können hier nicht im einzelnen wiederholt werden.

Die *folgende Übersicht (nächste Seite)* mag jedoch hilfreich sein als zusammenfassende Übersicht.

Einige Randbemerkungen, die für ein vertieftes Studium nützlich sein mögen, für den Anfang jedoch auch überschlagen werden können:

1. Man unterscheidet lokale und globale Grenzwertsätze (letztere auch "Integralsätze" genannt), je nachdem ob man die Wahrscheinlichkeitsfunktion f(x) oder die Verteilungsfunktion F(x) betrachtet:

lokal (X = x)	global (X
$f_n(x) \to f(x)$ Grenzwahrscheinlichkeitsverteilung	$F_n(x) \to F(x)$ Grenzverteilungsfunktion

2. Stochastische Konvergenz von einer Zufallsgröße X_n gegen einen festen Wert c kann als Spezialfall der Konvergenz der Verteilungsfunktion aufgefaßt werden: die Verteilung der Größe $Z_n = X_n - c$ konvergiert danach gegen die Verteilungsfunktion F(z) der Einpunktverteilung (siehe hierzu auch unten Seite 12, Beginn des Abschnitts 4).

[14] Man findet ganz ähnliche graphische Veranschaulichungen des zentralen Grenzwertsatzes z.B. in vielen anderen Lehrbüchern oder z.B. in einem Statistikskript von Prof. Dr. W. Neubauer (Univ. Frankfurt/M.).

<u>Übersicht[15]:</u> Grenzwertsätze und Varianten des Gesetzes der Großen Zahl(en)

homograd (Anzahl X und Anteil P = X/n [oder h_n statt P] der "Erfolge")

heterograd (Merkmalssumme $Y = \Sigma X_i$ und Mittelwert $\bar{x} = Y / n$)

stochastische Konvergenz (Konvergenz mit Wahrscheinlichkeit), Variante des Gesetzes der Großen Zahl

der Anteil P konvergiert mit Wahrscheinlichkeit gegen π nach dem Theorem von **Bernoulli** $\text{plim}(P) = \pi$

der Mittelwert $\bar{x}$ konvergiert gegen den Mittelwert der Grundgesamtheit (bzw. den Erwartungswert) μ nach dem Satz von **Ljapunoff** $\lim(\bar{X}) = \mu$

Grenzverteilung der Schätzfunktion $\hat{\theta}$ (Variante der Grenzwertsätze): $\hat{\theta}$ ist asymptotisch normalverteilt mit Erwartungswert $\mu_{\hat{\theta}}$ und Varianz $\sigma^2_{\hat{\theta}}$

Grenzwertsatz von **de Moivre-Laplace**

$\hat{\theta}$	$\mu_{\hat{\theta}}$	$\sigma^2_{\hat{\theta}}$
$X = \Sigma X_i$	$n\pi$	$n\pi(1-\pi)$
$P = X/n$	π	$\pi(1-\pi)/n$

Grenzwertsatz von **Lindeberg-Lévy**

$\hat{\theta}$	$\mu_{\hat{\theta}}$	$\sigma^2_{\hat{\theta}}$
$Y = \Sigma X_i$	$n\mu$	$n\sigma^2$
$\bar{X} = Y/n$	μ	σ^2/n

Voraussetzungen bezüglich der Verteilungen von X_1, X_2,... (in der Grundgesamtheit): unabhängig* identisch** verteilt (independently identically distributed [i.i.d.])

Zweipunktverteilung mit $E(X_i) = \pi$ und $V(X_i) = \pi(1 - \pi)$ $\forall i$

beliebig diskret oder stetig verteilt mit $E(X_i) = \mu$, $V(X_i) = \sigma^2$ $\forall i$

*) Diese Annahme kann auch gelockert werden (Normalverteilung auch Grenzverteilung bei abhängigen ZVn, vgl. Satz von Markoff).

) Verallgemeinerung (nicht identische Verteilungen): **Zentraler Grenzwertsatz von **Ljapunoff**

---

3. Das Gesetz der großen Zahl gilt nicht nur für Kennzahlen einer empirischen Häufigkeitsverteilung $h_n(x)$, wie z.B. das arithmetische Mittel $\bar{x}$ (in dem Sinne, dass gilt $\text{plim}\,\bar{x} = \mu$), sondern auch für alle Punkte[16] von $h_n(x)$ und der daraus abgeleiteten Summenhäufigkeitskurve $H_n(x)$ in dem Sinne, dass für jedes feste x bei $n \to \infty$ die empirische (Stichproben-) Verteilungsfunktion $H_n(x)$ mit Wahrscheinlichkeit gegen die theo-

[15] Eine Abwandlung von Übers. 7.1 und 7.2 von v.d.L., Induktive.

[16] bei allem Werten von x im Definitionsbereich von h(x).

retische (Grundgesamtheits-) Verteilungsfunktion F(x) konvergiert. Das ist bekannt als "Hauptsatz der (mathematischen) Statistik" von Gliwenko und Cantelli.

4. Wir machen im folgenden unterschiedliche Annahmen über die Verteilung der X_i (in der Grundgesamtheit)

die X_i sind .... verteilt	dann ist .... standardnormalverteilt nach ....
identisch *normal*verteilt mit $E(X_i) = \mu$ und $V(X_i) = \sigma^2$ (für alle i = 1,..., n)	$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ für jedes n. Das ist trivial da die Normalverteilung reproduktiv ist
identisch verteilt aber nicht notwendig *normal*verteilt	$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ wenn $n \to \infty$ nach dem Satz von Lindeberg Levy
nicht identisch verteilt, sondern mit $E(X_i) = \mu_i$ und $V(X_i) = \sigma_i^2$	$\sum_{i=1}^{n} \frac{X_i - \sum \mu_i}{\sum \sigma_i^2}$ wenn $n \to \infty$ nach dem zentralen Grenzwertsatz*

* von Ljapunow (oder Ljapunoff); am Ausdruck $\Sigma\sigma^2$ wird deutlich erkennbar, wie wichtig es ist, (paarweise) Unabhängigkeit der Zufallsvariablen X_1, X_2, ... vorauszusetzen.

3. Abschätzungen von Wahrscheinlichkeiten, Ungleichungen in der Stochastik

Von Interesse sind oft nicht nur die Grenzwerte von Wahrscheinlichkeiten, die naturgemäß meist 0 oder 1 sind. Es ist oft auch interessant zu sehen, welche Werte eine Wahrscheinlichkeit *mindestens* oder *höchstens* annimmt in Abhängigkeit von bestimmten Momenten (z.B. Erwartungswert, Varianz) der Verteilung einer Zufallsvariable X. Ein sehr allgemeiner Satz in diesem Zusammenhang ist

(3) $$P(h(x) \geq \varepsilon) \leq \frac{E(h(x))}{\varepsilon}$$

wobei h(x) eine beliebige nichtnegative Funktion der Zufallsvariable X ist, E( ) der Erwartungswert ist und der Satz aussagt, dass ein Wert von h(x), der größer/gleich ε ist *höchstens* (≤) eine Wahrscheinlichkeit von $E(\,)/\varepsilon$ hat. Von diesem Satz sind die bekannten Ungleichungen von Markoff und Tschebyscheff jeweils Spezialfälle[17]. Ist die Funktion h beispielsweise definiert als $h(x) = (x - \mu)^2$ mit $\mu = E(X)$, dann ist E(h(x)) natürlich gleich der Varianz σ^2, so dass man erhält

(4) $$P(|x - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2},$$

was eine spezielle Version der Tschebyscheffschen Ungleichung ist.

Achtung: Es ist nicht beabsichtigt, an dieser Stelle die Tschebyscheffsche Ungleichung und deren Herleitung zu erklären. Das geschieht in der Vorlesung!!

Man kann eine Abschätzung auch benutzen, um stochastische Konvergenzen zu beweisen, z.B. die Konvergenz der relativen Häufigkeit von x/n gegen die Wahrscheinlichkeit π. Gibt es nur zwei Ausprägungen einer Zufallsvariable, etwa Zahl oder Wappen, so nimmt x die Werte

[17] Heike H.D. und C. Tarcolea, Grundlagen der Statistik und Wahrscheinlichkeitsrechnung, München u. Wien (Oldenbourg), 2000, S. 346.

0 (Nichterfolg) und 1 (Erfolg – z.B. Wappen) an und man spricht von einer Zweipunkt- oder Bernoulli-Verteilung[18] mit $f(x) = \pi^x (1-\pi)^{1-x}$ für $x \in \{0,1\}$, $0 < \pi < 1$.

Ist x = 0 so ist f(x) = $1-\pi$ und für x = 1 erhält man f(x)=π. Die Zufallsvariable nimmt also die Werte 1 und 0 mit den Wahrscheinlichkeiten π und $1-\pi$ an. Betrachtet man eine Folge unabhängiger zweipunktverteilter Zufallsversuche, also z.B. das n-malige Werfen einer Münze, so ist die relative Häufigkeit x/n (relativiert) binomialverteilt mit $E(X/n) = \pi$ und der Varianz $V(X/n) = \pi(1-\pi)/n$ und es gilt dann gem. Tschebyscheffscher Ungleichung

(4a) $$P\left\{\left|\frac{x}{n} - \pi\right| < \varepsilon\right\} > 1 - \frac{\pi(1-\pi)}{n}.$$

Weil n auf der rechten Seite im Nenner steht gilt mit $n \to \infty$ für den Grenzwert dieser Wahrscheinlichkeit $\lim_{n\to\infty} P\{\ \} = 1$ (weil bei einer Wahrscheinlichkeit > 1 natürlich nicht definiert ist), was wiederum nichts anderes ist als das Gesetz der Großen Zahl in der oben zitierten Form des Theorems von Bernoulli.

4. Zusammenhänge zwischen einigen Konvergenzbegriffen in der Statistik

Man kann eine Konstante c als eine degenerierte Zufallsvariable x auffassen, so dass x nur den Wert c annimmt mit Wahrscheinlichkeit 1. Dieses Konzept einer "Einpunktverteilung" von c stellt sich formal wie folgt dar[19]: $f(x) = \begin{cases} 1 & \text{für } x = c \\ 0 & \text{sonst} \end{cases}$.

Man kann dann sagen, dass die relative Häufigkeit im Sinne des Gesetzes der Großen Zahl (gem. Gl. 2) gegen die Wahrscheinlichkeit 1 strebt, also $\lim P\left\{\left|\frac{x}{n} - \pi\right| < \varepsilon\right\} = 1$ bedeutet faktisch nichts anderes, als dass die Grenzverteilung von $\frac{x}{n}$ die Einpunktverteilung von $\pi = c$ ist, so dass das Gesetz der Großen Zahl nur ein spezieller Grenzwertsatz (Grenzverteilungssatz) ist.

Führt man den Begriff des absoluten Moments k-ter Ordnung ein, was der Erwartungswert von $|x^k|$ ist, also $E(|x^k|)$, so bedeutet Konvergenz im k-ten Mittel (abgekürzt **kM**) gegen die Zufallsvariable X oder gegen eine Konstante, etwa μ oder c, dass ein entsprechendes Moment asymptotisch verschwindet: $\lim_{n\to\infty} E\left(|X_n - \mu|^k\right) = 0$. Ein spezieller Fall hiervon ist die Konvergenz im quadratischen Mittel (k = 2 oder abgekürzt 2**M**), so dass bei hinreichend großem n die Werte X und μ mit hoher Wahrscheinlichkeit nahe beieinander liegen und deshalb die Varianz $E(X_n - \mu)^2$ gegen Null strebt. Hieraus folgt auch die oben ausschließlich betrachtete "gewöhnliche" stochastische Konvergenz, die auch "Konvergenz mit Wahrscheinlichkeit" (abgekürzt **WK**) genannt wird und die Konvergenz von Verteilungen (**VK**), wie z.B. die Konvergenz gegen die Normalverteilung. Es gilt, dass man wie folgt folgern kann:

k**M** $\Rightarrow$ 2**M** $\Rightarrow$ **WK** $\Rightarrow$ **VK**

(die Umkehrung dieser Schlußweise, also in der Richtung $\Leftarrow$ ist jedoch nicht zulässig).

[18] Das ist im Prinzip nichts anderes als die Binomialverteilung für den speziellen Fall n = 1.
[19] "Sonst" heißt hier: für andere Werte von x, also $x \neq c$.