1. Grenzwertsätze

Der wichtigste Grund für die Häufigkeit des Auftretens der Normalverteilung ergibt sich aus den ***Grenzwertsätzen***.

Grenzwertsätze sind Aussagen über eine Zufallsvariable für den Fall, dass die Anzahl der Beobachtungen *n* gegen Unendlichkeit strebt.

Zu den wichtigsten Grenzwertsätzen zählen das ***Schwache Gesetz der großen Zahlen*** und der ***zentrale Grenzwertsatz***.

<u>Annahme:</u> X sei eine *n*-dimensionale Zufallsvariable, deren Komponenten $X_1, X_2, \ldots, X_n$ ***stochastisch unabhängig*** und ***identisch verteilt*** sind.

Aus der identischen Verteilung folgt, dass alle Komponenten den gleichen Erwartungswert $E(X_i) = \mu$ und die gleiche Varianz $V(X_i) = \sigma^2$ haben.

1.1 Das Gesetz der Großen Zahlen

Erste Formulierung stammt von POISSON: „la loi des grands nombres“ (1837). Poisson meinte dabei eine große Zahl von Versuchen bei einem Zufallsexperiment.

Zufallsvariable X mit dem Erwartungswert μ und der Varianz σ^2 sei Realisation eines Zufallsexperimentes.

Das Experiment wird *n*-mal wiederholt:

$E(\overline{X}_n) = \mu$ und $V(\overline{X}_n) = \frac{\sigma^2}{n}$

Aus der letzten Gleichung ist ersichtlich, dass die Varianz des Stichprobenmittelwertes umso kleiner wird, je größer der Stichprobenumfang *n* ist.

Wenn *n* gegen unendlich strebt, konvergiert die Varianz gegen den Grenzwert Null:

$$\lim_{n \to \infty} V(\overline{X}_n) = 0$$

Je größer *n* und je kleiner damit die Varianz der Verteilung von $\overline{X}_n$ wird, desto näher wird das arithmetische Mittel einer Stichprobe $\overline{x}$ bei μ liegen.

<u>Schwaches Gesetz der Großen Zahlen:</u>

Seien $X_1, X_2, \ldots, X_n$ unabhängig und identisch verteilte Zufallsvariablen, deren Erwartungswerte $E(X_i) = \mu$ und Varianzen existieren, und sei $\overline{X}_n$ das arithmetische Mittel aus ihnen. Dann gilt für jedes beliebig kleine $\varepsilon > 0$

$$P(|\overline{X}_n - \mu| \geq \varepsilon) \rightarrow 0 \quad \text{für } n \rightarrow \infty$$

Die Wahrscheinlichkeit, dass die Abweichung des arithmetischen Mittels $\overline{X}_n$ vom Erwartungswert μ größer oder gleich einem vorgegebenen Wert ε ist, geht für große *n* gegen null. Mit anderen Worten:

Das ***Schwache Gesetz der Großen Zahlen*** besagt, dass sich eine Folge von Zufallsvariablen (Folge der arithmetischen Mittel $\overline{X}_n$) mit wachsendem *n* dem Erwartungswert μ nähert (gegen den Erwartungswert konvergiert):

$\boxed{p\lim_{n\to\infty} \overline{X}_n = \mu}$ Der Wahrscheinlichkeitslimes von $\overline{X}_n$ ist gleich μ.

Dies ist eine ***stochastische Konvergenz*** oder die ***Konvergenz nach Wahrscheinlichkeit***.

Zu beachten: der Wahrscheinlichkeitslimes ist vom gewöhnlichen Limes zu unterscheiden. Im Gesetz der großen Zahlen wird nicht behauptet, dass $\lim_{n\to\infty} \bar{x}_n = \mu$, also dass die Folge der beobachteten arithmetischen Mittel

$\overline{x}_n$ gegen μ konvergieren würde, sondern die Wahrscheinlichkeit für die Abweichung $\geq \varepsilon$ wird immer kleiner.

Das Starke Gesetz der großen Zahlen behauptet eine stärkere Konvergenz (fast sichere Konvergenz):

$$P(\lim \overline{X}_n = \mu) = 1$$

Beispiel: historische Lottozahlen (Zeitraum: 25 Jahre, Stichprobenumfang: 9114)

1 187	*2* 194	*3* 194	*4* 178	*5* 176	*6* 187	*7* 175
8 179	*9* 201	*10* 173	*11* 175	*12* 180	*13* 157	*14* 181
15 172	*16* 180	*17* 190	*18* 183	*19* 191	*20* 181	*21* 200
22 191	*23* 186	*24* 177	*25* 196	*26* 200	*27* 180	*28* 167
29 186	*30* 182	*31* 199	*32* 211	*33* 189	*34* 175	*35* 183
36 199	*37* 175	*38* 199	*39* 199	*40* 195	*41* 183	*42* 182
43 189	*44* 181	*45* 188	*46* 191	*47* 175	*48* 195	*49* 207

Empirisch:

$$\bar{x} = (1 \cdot 187 + 2 \cdot 194 + 3 \cdot 194 + \ldots + 49 \cdot 207) / 9114 = 25{,}2211$$

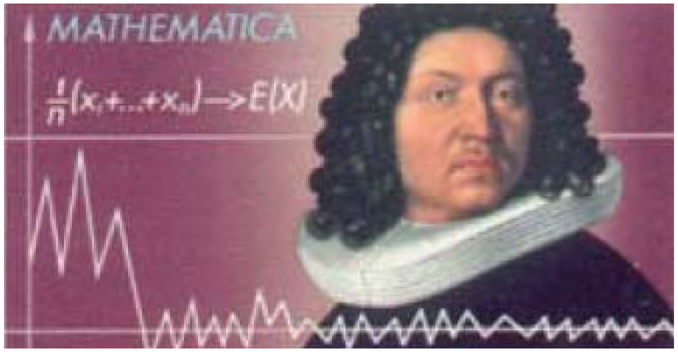
$$s_x = \sqrt{200{,}6512} = 14{,}1651$$

Theoretisch:

$\mu = (49+1)/2 = 50/2 = 25$ (Erwartungswert einer gleichförmig verteilten Variablen)

$\sigma_x = \sqrt{(49^2 - 1)/12} = \sqrt{200} = 14{,}1421$ (Standardabweichung bei Gleichverteilung)

1.2. Bernoullis Gesetz (Jacob Bernoulli, 1668):



Bernoullis Gesetz ist ein Spezialfall des allgemeinen Gesetzes der großen Zahlen.

Der Begriff der *relativen Häufigkeit* ist mit dem Begriff der Wahrscheinlichkeit eng verbunden. Die relative Häufigkeit eines Ereignisses aus unabhängigen

Wiederholungen eines Zufallsexperiments wird als eine Näherung für die Wahrscheinlichkeit des Ereignisses angesehen.

Das Zufallsexperiment sei ein Bernoulli-Experiment mit *n* unabhängigen Versuchen. Bei jedem Versuch tritt das Ereignis A mit der Wahrscheinlichkeit *p* ein.

Die Zufallsvariable sei eine Bernoulli-Variable, die wie folgt definiert ist:

$$X_i = \begin{cases} 1, & \text{falls } A \text{ beim } i\text{-ten Versuch eintritt} \\ 0, & \text{falls } \overline{A} \text{ beim } i\text{-ten Versuch eintritt} \end{cases}$$

Für jeden einzelnen Versuch *i* gilt:

$E(X_i) = p$ und $V(X_i) = pq,$ wobei $q = 1-p$

Die Anzahl der Erfolge in der Versuchsreihe ist binomialverteilt mit den Parametern *n* und *p* und der Varianz *npq*.

Wir interessieren uns jedoch nicht für die absolute Anzahl, sondern für die ***relative Häufigkeit*** der Erfolge in der Versuchsreihe:

$$H_n = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$ relative Häufigkeit des Eintretens von *A*

Die Zufallsvariable H_n liegt zwischen $0 \leq H_n \leq 1$

Für die Zufallsvariable H_n gilt:

$$\boxed{\begin{aligned} E(H_n) &= p \\ V(H_n) &= pq/n \end{aligned}}$$

Der Erwartungswert der relativen Häufigkeit H_n entspricht der Erfolgswahrscheinlichkeit *p* und die Varianz der relativen Häufigkeit H_n wird umso kleiner, je größer die Anzahl der Versuche ist. Wenn *n* gegen Unendlichkeit strebt, konvergiert die Varianz gegen den Grenzwert Null:

$$\lim V(H_n) = 0$$

Für sehr große *n* wird die relative Häufigkeit h_n einer Versuchsreihe nahe bei dem Wert *p* liegen. Je größer *n* wird, umso kleiner wird das Intervall um *p* sein, in das der Anteilswert h_n einer Versuchsreihe mit großer Wahrscheinlichkeit fällt.

Ein Bernoulli-Experiment mit der Erfolgswahrscheinlichkeit *p* werde *n*-mal unabhängig wiederholt, h_n sei dabei die relative Häufigkeit der Erfolge. Dann gilt für jedes beliebig kleine $\varepsilon > 0$:

$$P(|H_n - p| \geq \varepsilon) \to 0 \quad \text{für} \quad n \to \infty$$ **(Bernoullis Gesetz der großen Zahlen)**

Andere (äquivalente) Schreibweise:

$$\boxed{\underset{n\to\infty}{p\lim} H_n = p}$$

Bernoullis Gesetz der großen Zahlen besagt, dass die relative Häufigkeit H_n stochastisch gegen die Wahrscheinlichkeit *p* konvergiert.

Das Schwache Gesetz der großen Zahlen bildet eine Brücke zwischen dem theoretischen Konzept der Wahrscheinlichkeitsrechnung und den beobachtbaren Ergebnissen von Zufallsexperimenten.

Im Sinne der stochastischen Konvergenz ist die Wahrscheinlichkeit eines Ereignisses gleich dem Grenzwert der relativen Häufigkeit bei wiederholter unabhängiger Durchführung des Experiments.

1.3. Konvergenz der empirischen Verteilungsfunktion (Hauptsatz der Statistik)

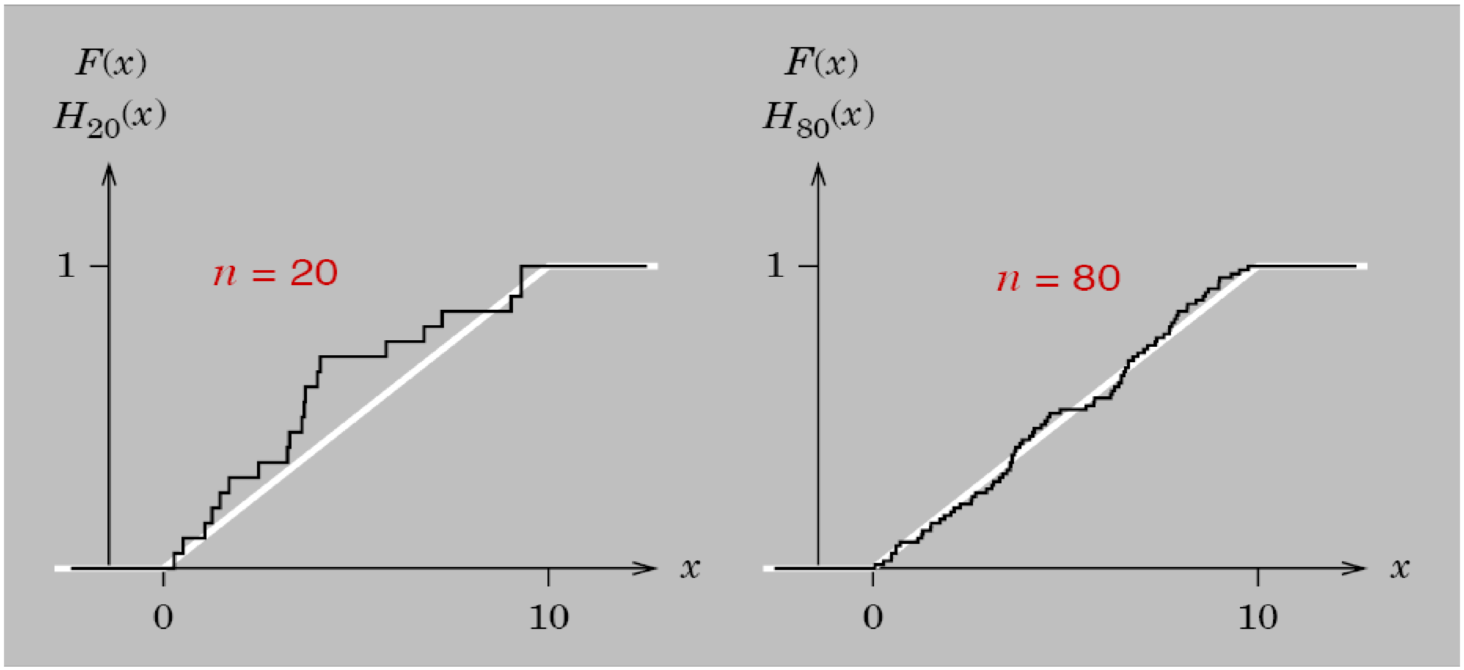
Nicht nur die Konvergenz der Mittelwerte, sondern auch die Konvergenz der ganzen Verteilungsfunktionen kann durch die Grenzwertsätze beschrieben werden.

Die empirische Verteilungsfunktion $F_n(x)$ konvergiert mit zunehmendem Stichprobenumfang n gegen die Wahrscheinlichkeitsverteilungsfunktion $F(x)$ für $n \to \infty$:

$$\boxed{p\lim_{n\to\infty} F_n(x) = F(x)}$$

Andere Schreibweisen: $p\lim_{n\to\infty}(F_n(x) - F(x)) = 0, \qquad P(\lim_{n\to\infty} F_n(x) = F(x)) = 1$

Bsp.: auf einem PC werden Zufallszahlen gezogen, gleichförmig verteilt über dem Intervall [0, 10]. Quelle: J. Schira, S. 404-405.



Konvergenz der empirischen Verteilungsfunktion gegen die Wahrscheinlichkeitsverteilungsfunktion.

1.4. Der zentrale Grenzwertsatz

Sei X_1, X_2, …, X_n eine Folge von unabhängigen und identisch verteilten Zufallsvariablen mit $\mu = E(X_i)$ *und* $\sigma^2 = V(X_i)$.

Wir betrachten die Summe dieser Zufallsvariablen $S_n = X_1 + X_2 + ... + X_n$.
Der Erwartungswert von der Summe S_n ist $n\mu$ und die Varianz ist $n\sigma^2$.

Dann strebt die Verteilungsfunktion der standardisierten Größe

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \equiv \frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}} \qquad (\overline{X}_n = S_n / n)$$

mit wachsendem n gegen die Standardnormalverteilung

$$F_n(z_n) \rightarrow F_{St}(z_n) \quad \text{für } n \rightarrow \infty.$$

Vorteil des zentralen Grenzwertsatzes: er stellt keinerlei Anforderung an die ursprüngliche Verteilung. Die Verteilungsfunktion der Summe bzw. des arithmetischen Mittels der identisch verteilten und unabhängigen Zufallsvariablen konvergiert bei $n \rightarrow \infty$ gegen die Normalverteilung.

Dies erklärt die Sonderstellung der Normalverteilung, ihre große theoretische und praktische Bedeutung.