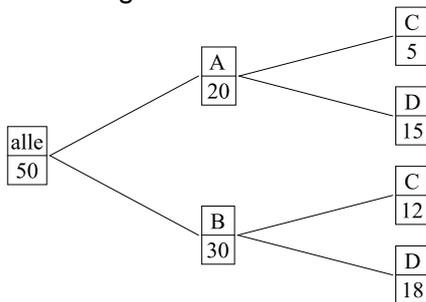


Bedingte Wahrscheinlichkeiten:

Ein Baumdiagramm der nachstehenden Art bzw. die zugehörige Vierfeldertafel gibt jeweils absolute Häufigkeiten an.



	C	D	
A	5	15	20
B	12	18	30
	17	33	50

Interpretiert man die numerischen Einträge als Beschreibungen einer Urne (in der 50 Kugeln sind, davon 20 mit Eigenschaft A, 5 mit den Eigenschaften A und C usw.), so kann man dem Baumdiagramm bzw. der Vierfeldertafel Wahrscheinlichkeiten zuordnen:

Ebenfalls mit A werde das Ereignis bezeichnet, eine Kugel mit der Eigenschaft A gezogen zu haben; analog für B, C, D.

Der linksstehende Teil des Baumdiagramms bzw. die beiden mittleren Zeilen (und damit die rechte Spalte) der Vierfeldertafel bedeuten $P(A) = \frac{20}{50}$ und $P(B) = \frac{30}{50}$: Die

Wahrscheinlichkeit, eine Kugel mit der Eigenschaft A zu ziehen, beträgt $\frac{2}{5}$ usw.

Wenn man eine A-Kugel gezogen hat, so hat sie mit Wahrscheinlichkeit $\frac{5}{20}$ die Eigenschaft

C. Natürlich ist $P(C) \neq \frac{5}{20}$, denn man muss ja berücksichtigen, dass es sich um eine A-

Kugel handelt. Man schreibt $P(C|A) = \frac{5}{20}$ und $P(D|A) = \frac{15}{20}$. Die A-Kugeln bilden den neuen „Grundraum“, und es gilt $P(C|A) + P(D|A) = 1$.

Bei $P(C|A)$ gilt: Das Ereignis, eine A-Kugel gezogen zu haben, ist das *bedingende* Ereignis. Das Ereignis, eine C-Kugel gezogen zu haben, ist das *bedingte* Ereignis.

An der Vierfeldertafel kann man gut sehen, dass $P(C|A) = \frac{5}{20}$ und $P(A|C) = \frac{5}{17}$

verschieden sind. Man muss daher sorgfältig zwischen bedingendem und bedingtem Ereignis unterscheiden.

Manche Schulbüchern schreiben nicht $P(C|A)$, sondern $P_A(C)$.

Zur stochastischen Unabhängigkeit:

Das Ereignis, eine C-Kugel gezogen zu haben, ist vom Ereignis, eine A-Kugel gezogen zu haben, *stochastisch unabhängig*, wenn die Tatsache, dass es sich um eine A-Kugel handelt, gar keine Rolle spielt, d.h. wenn $P(C|A) = P(C)$. Das ist bei dem oben angegebenen Beispiel nicht der Fall.

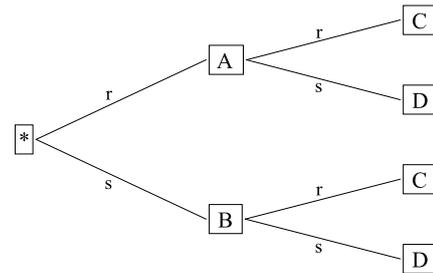
Auch die Ereignisse B und D sind *nicht* stochastisch unabhängig, da $P(D) = \frac{33}{50}$ und

$P(D|B) = \frac{18}{30}$ verschieden voneinander sind.

Beim nebenstehenden Baumdiagramm (mit $r + s = 1$) sind die Ereignisse C und D beide von A und von B stochastisch unabhängig wegen

$$P(C|A) = r \text{ und } P(C) = r^2 + s \cdot r = r.$$

Solche Baumdiagramme sind für stochastische Unabhängigkeit typisch. Bei der Binomialverteilung liegen solche Baumdiagramme vor.



Die bedingte Wahrscheinlichkeit lässt sich auch als $P(C|A) = \frac{P(C \cap A)}{P(A)}$ ausdrücken. Ist C

von A stochastisch unabhängig, so muss $\frac{P(C \cap A)}{P(A)} = P(C)$ bzw. $P(C \cap A) = P(C) \cdot P(A)$

gelten. Die stochastische Unabhängigkeit hat also die kontraintuitive Eigenschaft, in A und C *symmetrisch* zu sein.

Die *stochastische* Unabhängigkeit darf nicht mit der *kausalen* Unabhängigkeit verwechselt werden. Was soll kausale Unabhängigkeit bedeuten?

Zwei Ereignisse A und B heißen *kausal abhängig*, wenn A aus B kausal folgt oder B aus A. Sie heißen *kausal unabhängig*, wenn weder das eine noch das andere gilt, man also weder B aus A noch A aus B kausal schließen kann.

Dass diese beiden Begriffe nicht deckungsgleich sind, zeigen zwei Beispiele:

1. Wirft man mit zwei Würfeln, so sind die Ereignisse A: „Augensumme ist 8“ und B: „Erster Wurf ist eine 6“ kausal unabhängig, denn weder kann man von A auf B schließen noch umgekehrt. Andererseits hat man wegen $P(A) \cdot P(B) = \frac{5}{36} \cdot \frac{6}{36}$ und $P(A \cap B) = \frac{1}{36}$ keine stochastische Unabhängigkeit.

2. In einer Urne sind 2 rote und 2 schwarze Kugeln. Es werden nacheinander zwei Kugeln ohne Zurücklegen gezogen. Dann sind die Ereignisse A: „Die erste Kugel ist schwarz“ und

B: „Die zweite Kugel ist rot“ wegen $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$, $P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{1}{2}$ und

$P(A \cap B) = \frac{1}{3}$ nicht stochastisch unabhängig, aber man hat kausale Unabhängigkeit, da

man weder von A auf B noch von B auf A schließen kann.

Mit der oben angegebenen Definition der kausalen Unabhängigkeit gilt: Die stochastische Unabhängigkeit impliziert kausale Unabhängigkeit, jedoch nicht umgekehrt, wie man an den Beispielen hat sehen können.

Zur Varianz:

Die Varianz ist ein Streuungsmaß. Da sie eine andere Einheit hat als die in Rede stehenden Größen, nimmt man ihre Quadratwurzel, die Standardabweichung.

Da man jedoch für einfache Verteilungen die Varianz gut hilfsmittelfrei berechnen kann, tritt dieser Begriff im Kerncurriculum auf.

Zum Term der Binomialverteilung:

Beispiel: Urne mit n Kugeln, von denen s schwarz und w weiß sind. Man zieht n -mal *mit*

Zurücklege. Dabei zieht man jeweils mit Wahrscheinlichkeit $p = \frac{w}{n}$ eine weiße Kugel und mit

Wahrscheinlichkeit $q = \frac{s}{n}$ eine schwarze Kugel. Die Zufallsvariable A gebe die Anzahl der

weißen Kugeln an. Dann ist $P(A = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$. Dabei ist der Binomialkoeffizient die

Anzahl der Möglichkeiten, bei n Ziehungen k weiße zu bekommen. Der Binomialkoeffizient

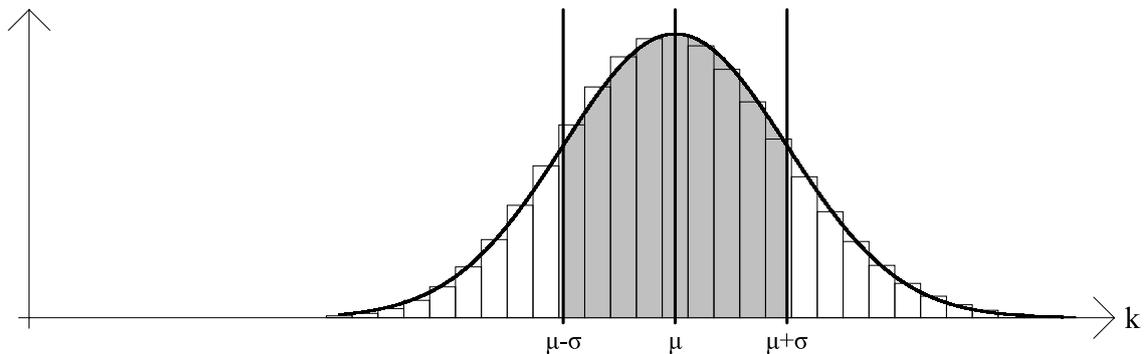
gibt also die *Anzahl* der Pfade an, die jeweils mit Wahrscheinlichkeit $p^k \cdot q^{n-k}$ eingeschlagen werden.

Prognoseintervalle:

Die Binomialverteilung wird durch die Normalverteilung approximiert (das geht einigermaßen gut, wenn $n \cdot p \cdot (1-p) > 9$ ist). Das Bild zeigt die Verhältnisse für $n = 100$ und $p = 0,25$; im

Unterricht wird man mit Schiebereglern n und p variieren.

Wahrscheinlichkeitsdichte



Der grau gekennzeichnete Flächeninhalt zwischen $\mu - \sigma$ und $\mu + \sigma$ unter dem Graphen der Normalverteilung gibt die Wahrscheinlichkeit an, mit der Werte zwischen $\mu - \sigma$ und $\mu + \sigma$ angenommen werden. Diese Wahrscheinlichkeit beträgt (unabhängig von den Werten für μ und σ stets) etwa 68 %.

Zwischen $\left\{ \begin{array}{l} \mu - \sigma \\ \mu - 2 \cdot \sigma \\ \mu - 3 \cdot \sigma \end{array} \right\}$ und $\left\{ \begin{array}{l} \mu + \sigma \\ \mu + 2 \cdot \sigma \\ \mu + 3 \cdot \sigma \end{array} \right\}$ liegen stets etwa $\left\{ \begin{array}{l} 68,3\% \\ 95,4\% \\ 99,7\% \end{array} \right\}$ aller Werte. Ebenfalls gilt:

Zwischen $\left\{ \begin{array}{l} \mu - 1,64 \cdot \sigma \\ \mu - 1,96 \cdot \sigma \\ \mu - 2,58 \cdot \sigma \end{array} \right\}$ und $\left\{ \begin{array}{l} \mu + 1,64 \cdot \sigma \\ \mu + 1,96 \cdot \sigma \\ \mu + 2,58 \cdot \sigma \end{array} \right\}$ liegen stets etwa $\left\{ \begin{array}{l} 90\% \\ 95\% \\ 99\% \end{array} \right\}$ aller Werte.

Dies sind die $k \cdot \sigma$ -Umgebungen.

Binomialverteilung und Normalverteilung mit GeoGebra:

Der Befehl

$$\text{binomial}(n, p)$$

liefert das Histogramm; es ist also sinnvoll, für n und für p Schieberegler zu haben.

Der Erwartungswert $\mu = n \cdot p$ und die Standardabweichung $\sigma = \sqrt{n \cdot p \cdot (1 - p)}$ werden im Algebra-Fenster definiert. Dann bekommt man den Graphen der Normalverteilung durch

$$\varphi = \text{Normal}[\mu, \sigma, x].$$

So lässt sich die Approximation der Binomialverteilung durch die Normalverteilung gut visualisieren.

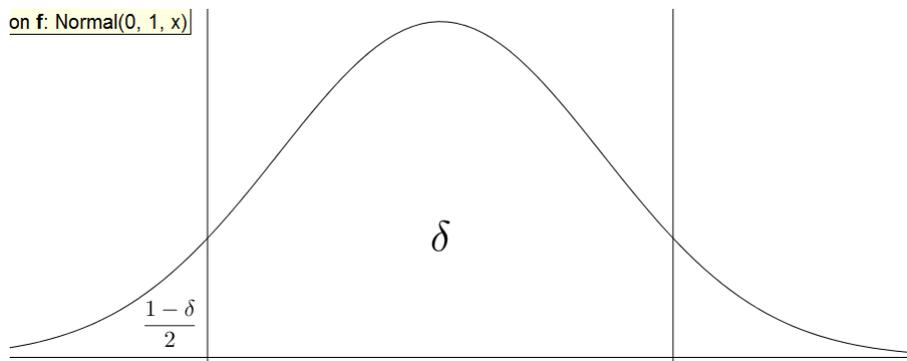
Der Flächeninhalt „unter“ der Standard-Normalverteilung zwischen den Grenzen $-k$ und k ist mit

$$\text{Integral}(\varphi, -k, k)$$

abrufbar. Ist umkehrt der Flächeninhalt δ gegeben, bekommt man k mithilfe von

$$\text{InversNormal}(0, 1, (1 + \delta) / 2)$$

aufgrund der folgenden Graphik:

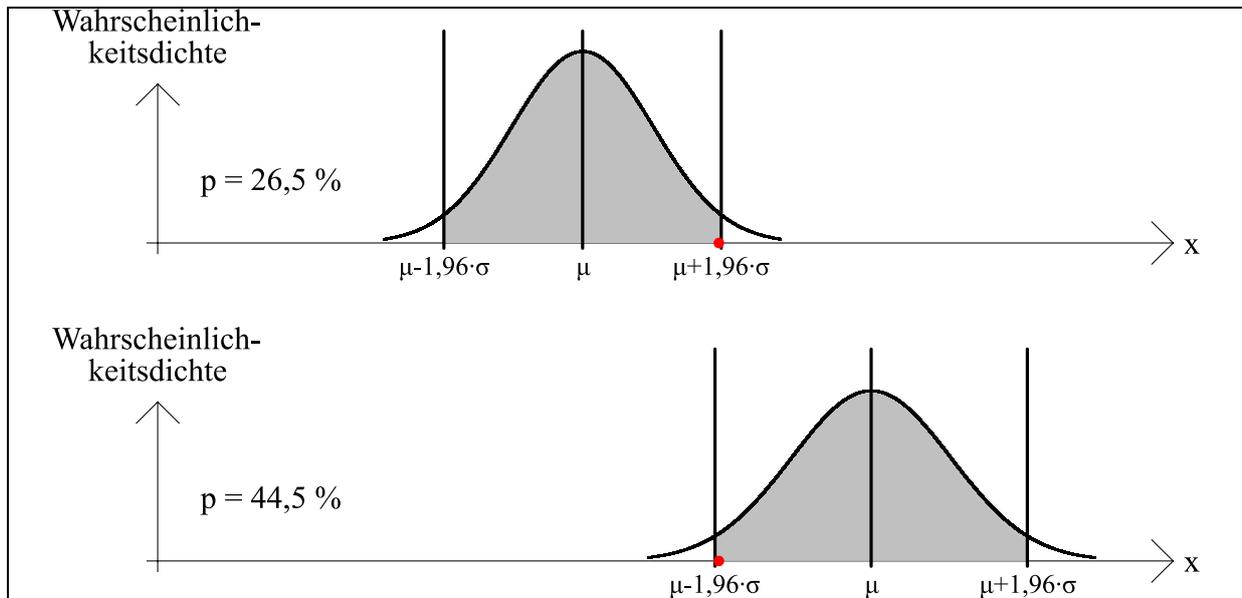


Konfidenzintervalle:

Eine Urne enthalte eine unbekannte Anzahl weißer und schwarzer Kugeln. Man hat 100-mal mit Zurücklegen gezogen und 35 weiße Kugeln erhalten. Mit welchen Anteilen von weißen Kugeln ist dies Ergebnis verträglich?

Approximiert man die Binomialverteilung durch die Normalverteilung, so weiß man, dass zwischen $\mu - 1,96 \cdot \sigma$ und $\mu + 1,96 \cdot \sigma$ stets etwa 95 % aller Werte liegen.

Nun probiere man verschiedene (etwa zu 95 % gehörige) Prognoseintervalle aus:



Der rote Punkt bezeichnet den Messwert 35. Ist $p = 26,5 \%$ bzw. $p = 44,5 \%$, so überdeckt das zu 95 % gehörige Prognoseintervall gerade noch den Messwert. Das zu 95 % gehörige Konfidenzintervall ist demnach das Intervall von 26,5 % bis 44,5 %. Es überdeckt den wahren Wert mit der Wahrscheinlichkeit 95 %.

Sicherheitswahrscheinlichkeit und Konfidenzintervalle:

Bei vorgegebener Sicherheitswahrscheinlichkeit δ überdecken die Konfidenzintervalle den wahren Wert mit der Wahrscheinlichkeit δ .

Dies lässt sich einerseits durch Simulation illustrieren. Im nebenstehenden Bild wurden 20-mal zu 0,4 gehörige binomialverteilte Zufallszahlen erzeugt und zu denen jeweils das 90 %-Konfidenzintervall berechnet.

Wenn man das häufig wiederholt, so überdecken im Schnitt 2 Konfidenzintervalle den wahren Wert 0,4 nicht.

Man sieht andererseits den obigen Satz formal folgendermaßen ein:

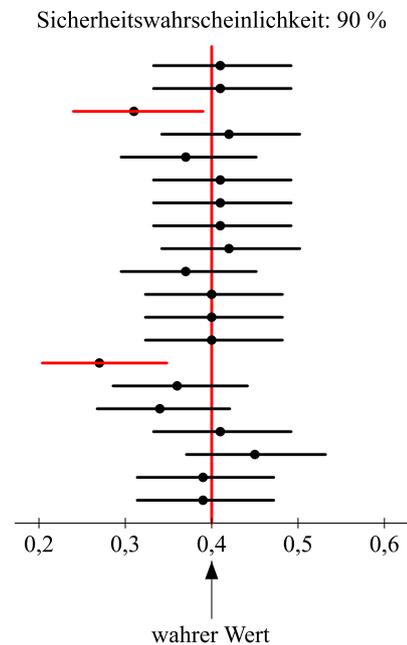
Beobachtet werde die absolute Häufigkeit A bei n Würfeln. Dann muss p so bestimmt werden, dass

$$\mu - f \cdot \sigma \leq A \leq \mu + f \cdot \sigma$$

bzw.

$$n \cdot p - f \cdot \sqrt{n \cdot p \cdot (1-p)} \leq A \leq n \cdot p + f \cdot \sqrt{n \cdot p \cdot (1-p)}$$

gilt.



Dabei ist f der zur Sicherheitswahrscheinlichkeit δ gehörige Faktor: Zwischen $\mu - f \cdot \sigma$ und $\mu + f \cdot \sigma$ liegen stets $\delta \cdot 100\%$ aller Werte. Durch die Ungleichung

$$n \cdot p - f \cdot \sqrt{n \cdot p \cdot (1-p)} \leq A \leq n \cdot p + f \cdot \sqrt{n \cdot p \cdot (1-p)}$$

sind die Wahrscheinlichkeiten p charakterisiert, die zum (von A und δ abhängigen) Konfidenzintervall gehören. Der wahre (aber unbekannte) Wert p liegt genau dann im zu A und δ gehörigen Konfidenzintervall,

wenn $n \cdot p - f \cdot \sqrt{n \cdot p \cdot (1-p)} \leq A \leq n \cdot p + f \cdot \sqrt{n \cdot p \cdot (1-p)}$ ist. Diese Bedingung ist aber mit der

Wahrscheinlichkeit δ erfüllt. Denn für die absolute Häufigkeit A gilt mit der Wahrscheinlichkeit δ , dass für sie die Ungleichung $n \cdot p - f \cdot \sqrt{n \cdot p \cdot (1-p)} \leq A \leq n \cdot p + f \cdot \sqrt{n \cdot p \cdot (1-p)}$ gilt. Daher

überdeckt das aus den Größen A und δ konstruierte Konfidenzintervall mit der Wahrscheinlichkeit δ den wahren Wert p .

Viele Konfidenzintervalle mit GeoGebra:

Die Grenzen der Konfidenzintervalle sind Zufallsgrößen, da sie von der Zufallsgröße „Messwert“ abhängen.

Wir messen die absolute Häufigkeit H . Wenn w der wahre Wert der Einzelerfolgs-Wahrscheinlichkeit ist, ist die absolute Häufigkeit H eine (n, w) -binomialverteilte Zufallsgröße; jede Messung liefert ein anderes H und damit ein anderes Konfidenzintervall.

Mit $H = \text{ZufallszahlBinomialverteilt}[n, w]$ wird H erzeugt. Mit F9 (oder Strg-R in neueren GeoGebra-Versionen) bekommt man eine neue Zufallszahl H .

Nun will man etwa 10-mal H erzeugen und entsprechend 10 Konfidenzintervalle zeigen.

Dazu müssen die Konfidenzintervalle *berechnet* werden.

Die grundlegende Doppelungleichung ist $\mu - k \cdot \sigma \leq H \leq \mu + k \cdot \sigma$. Bei gegebener Sicherheits-Wahrscheinlichkeit δ muss also k zunächst bestimmt werden; dies macht man mit

$$k = \text{InversNormal}(0, 1, (1 + \delta) / 2).$$

Setzt man für μ und σ die entsprechenden Ausdrücke in die Doppelungleichung ein und löst man diese dann nach p auf, bekommt man (ggf. mit CAS-Hilfe)

$$p = \frac{2 \cdot n \cdot H + k^2 \cdot n \pm k \cdot \sqrt{k^2 \cdot n^2 + 4 \cdot n^2 \cdot H - 4 \cdot n \cdot H^2}}{2 \cdot n^2 + 2 \cdot k^2 \cdot n}.$$

Nun zeigt sich, wie gut GeoGebra mit Listen umgehen kann:

Zunächst liefert

$$\text{LH} = \text{Folge}[\text{ZufallszahlBinomialverteilt}[n, w], i, 1, 10]$$

10 binomial verteilte Zufallszahlen. Die Liste der Radikanden ist

$$\text{LRad} = k^2 \cdot n^2 + 4 \cdot n^2 \cdot \text{LH} - 4 \cdot n \cdot \text{LH}^2.$$

Der Nenner ist stets

$$\text{Nenn} = 2 \cdot n^2 + 2 \cdot k^2 \cdot n.$$

Die Liste der „linken p-Werte“ ist

$$\text{Llinks} = (2 \cdot n \cdot \text{LH} + k^2 \cdot n - k \cdot \sqrt{\text{LRad}}) / \text{Nenn},$$

die Liste der „rechten p-Werte“ ist analog

$$\text{Lrechts} = (2 \cdot n \cdot \text{LH} + k^2 \cdot n + k \cdot \sqrt{\text{LRad}}) / \text{Nenn}.$$

Nun braucht man noch die Liste der linken und rechten Endpunkte der Konfidenzintervalle:

$$\text{Plinks} = \text{Folge}[(\text{Llinks}(i), i), i, 1, 10]$$

$$\text{Prechts} = \text{Folge}[(\text{Lrechts}(i), i), i, 1, 10]$$

Schließlich will man die linken und rechten Endpunkte noch mit jeweils einer Strecke verbinden:

$$\text{Folge}[\text{Strecke}[\text{Plinks}(i), \text{Prechts}(i)], i, 1, 10].$$

Stochastische Situationen simulieren, die zu annähernd binomialverteilten Zufallsgrößen führen:

Man nehme eine Urne mit 30 mit „1“ beschrifteten Kugeln und 70 mit „0“ beschrifteten Kugeln. Man ziehe 10-mal *ohne* Zurücklegen. Dann ist die Zufallsgröße „Anzahl der Einsen“ annähernd binomialverteilt. In GeoGebra kann man folgendermaßen vorgehen:

$$L1 = \text{Folge}[\text{Wenn}[k \leq 30, 1, 0], k, 1, 100]$$

liefert nacheinander 30 Einsen und 70 Nullen, die man mit

$$L2 = \text{Mischen}[L1]$$

durcheinander bringt. Davon braucht man die ersten 10 Elemente, also die mit Index 1 bis Index 10:

$$L3 = \text{Teilliste}[L2, 1, 10].$$

Die Anzahl der Einsen wird durch

$$\text{ZähleWenn}[x == 1, L3]$$

geliefert. Nun ist es sinnvoll, die bisherigen Befehle zu einem einzigen zusammenzufassen:

$$\text{ZähleWenn}[x == 1, \text{Teilliste}[\text{Mischen}[\text{Folge}[\text{Wenn}[k \leq 30, 1, 0], k, 1, 100]], 1, 10]],$$

um ihn etwa $N = 200$ -mal anwenden zu können:

$$LL = \text{Folge}[\text{ZähleWenn}[x == 1, \text{Teilliste}[\text{Mischen}[\text{Folge}[\text{Wenn}[k \leq 30, 1, 0], k, 1, 100]], 1, 10]], j, 1, N]$$

Das Säulendiagramm bekommt man mit

$$\text{Balkendiagramm}[LL, 1],$$

wobei das zweite Argument („1“) die Balkenbreite angibt.

Will man dieses Diagramm mit dem der Binomialverteilung vergleichen, muss man

$$\text{Balkendiagramm}[LL, 1, 1/N]$$

eingeben, wobei das letzte Argument („1/N“) eine vertikale Skalierung bewirkt.

Stochastische Situationen simulieren, die zu annähernd normalverteilten Zufallsgrößen führen:

Wie lange muss man beim Würfeln auf die erste „6“ warten? Kurze Wartezeiten sind häufiger als lange Wartezeiten. Wie macht man das in GeoGebra? Mit

$$L = \text{Folge}(\text{Zufallszahl}(1, 6), i, 1, 100)$$

werden 100 Würfelerggebnisse erzeugt (dabei ist 100 nur eine willkürliche große Zahl). Die Wartezeit bis zur ersten „6“ ist einfach der erste Index von L, bei dem man eine „6“ hat:

$$\text{IndexVon}(6, L).$$

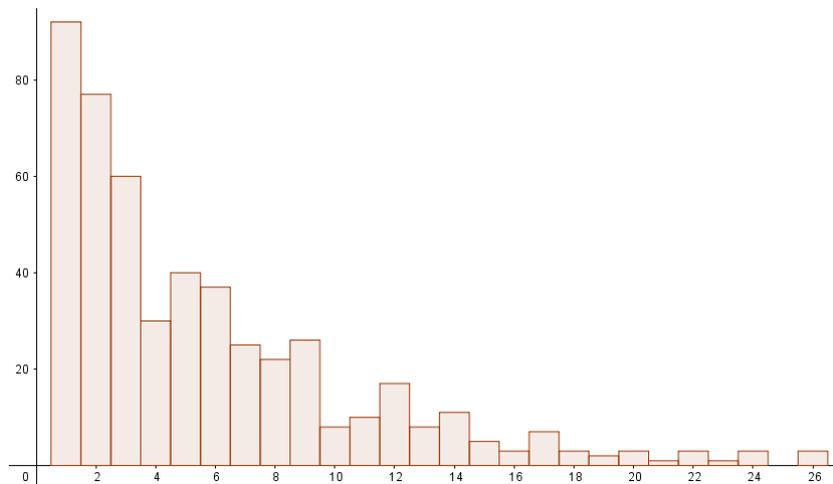
Um $N = 500$ solcher Wartezeiten zu bekommen, verwendet man

$$LL = \text{Folge}(\text{IndexVon}(6, \text{Folge}(\text{Zufallszahl}(1, 6), i, 1, n))), j, 1, N)$$

und bekommt mit

$$\text{Balkendiagramm}(LL, 1)$$

das zugehörige Säulendiagramm:



Wenn nun 10 Leute gleichzeitig würfeln und jeweils die Mittelwerte ihrer Wartezeiten bis zur ersten "6" aufschreiben, so hat das Histogramm der Mittelwerte eine ganz andere Gestalt!

Dazu bilde man

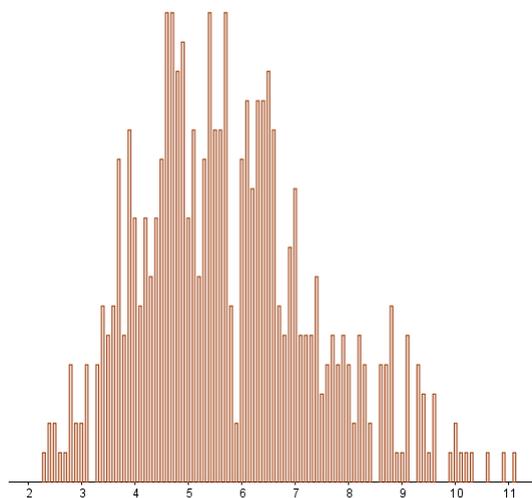
$$A = \text{Folge}(\text{Folge}(\text{IndexVon}(6, \text{Folge}(\text{Zufallszahl}(1, 6), i, 1, n)), j, 1, N), k, 1, 10)$$

und

$$B = (A(1) + A(2) + A(3) + A(4) + A(5) + A(6) + A(7) + A(8) + A(9) + A(10)) / 10.$$

sowie

Balkendiagramm(B, 0.05).



Die arithmetischen Mittel sind etwa normalverteilt. Die Ähnlichkeit zur Normalverteilung steigt, wenn die Anzahl der gleichzeitig Würfelnden zunimmt.

Dies ist sehr bemerkenswert: Die arithmetischen Mittel von (fast) *irgendwie* verteilten Zufallsgrößen (deren Verteilung auch sehr asymmetrisch sein kann wie im Beispiel oben) sind annähernd normalverteilt