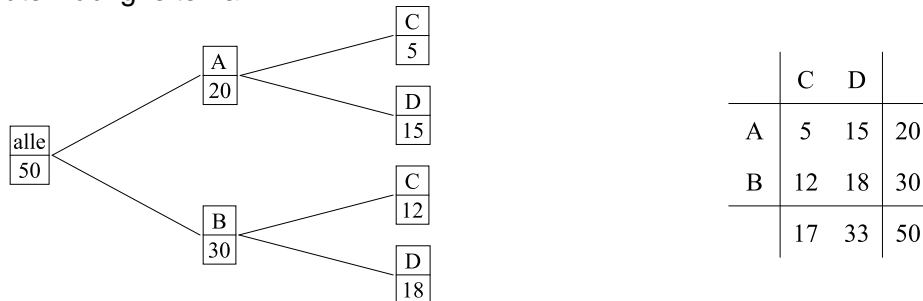


Bedingte Wahrscheinlichkeiten:

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



Interpretiert man die numerischen Einträge als Beschreibungen einer Urne (in der etwa 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 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 $P(C|A) = P(C)$ gilt. Das ist bei dem oben angegebenen Beispiel nicht der Fall.

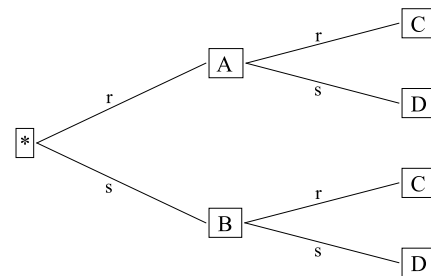
Auch die Ereignisse B und D sind *nicht* stochastisch unabhängig, da $P(D) = \frac{33}{50}$ und

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

**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

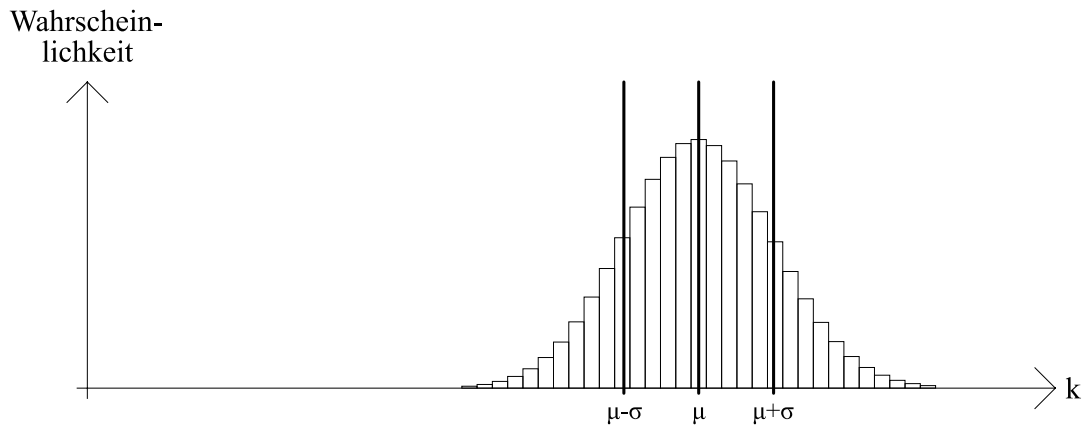
Wahrscheinlichkeit $q = \frac{s}{n}$ eine schwarze Kugel. Die Zufallsvariable A gebe die Anzahl der

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:

Das folgende Säulendiagramm zeigt die Verhältnisse für $n = 100$ und $p = 0,4$; im Unterricht wird man mit Schieberegler n und p variieren. Die Höhen der (gleich breiten) Säulen geben Wahrscheinlichkeiten an.



Wenn die Binomialverteilung einigermaßen symmetrisch ist, so gilt:

Zwischen $\left\{ \begin{array}{l} \mu - \sigma \\ \mu - 2 \cdot \sigma \end{array} \right\}$ und $\left\{ \begin{array}{l} \mu + \sigma \\ \mu + 2 \cdot \sigma \end{array} \right\}$ liegen stets etwa $\left\{ \begin{array}{l} 68\% \\ 95\% \end{array} \right\}$ aller Werte. Dies ist die σ -

Umgebung bzw. die $2 \cdot \sigma$ -Umgebung von μ .

Im Beispiel $n = 100$ und $p = 0,4$ ist $\mu = 40$ und $\sigma = \sqrt{n \cdot p \cdot (1-p)} \approx 4,9$. Die σ -Umgebung ist somit durch $\mu - \sigma \approx 35,1$ und $\mu + \sigma \approx 44,9$ gegeben. Damit treten Werte zwischen 35 und 45 mit mindestens 68 %-iger Wahrscheinlichkeit auf:

$$\binom{100}{35} \cdot 0,4^{35} \cdot 0,6^{65} + \dots + \binom{100}{45} \cdot 0,4^{45} \cdot 0,6^{55} \approx 74\%, \text{ aber}$$

$$\binom{100}{36} \cdot 0,4^{36} \cdot 0,6^{64} + \dots + \binom{100}{44} \cdot 0,4^{44} \cdot 0,6^{56} \approx 64\%.$$

Es ist immer sinnvoll, „nach außen“ zu runden.

Beispiel zur Anwendung: Zur Überprüfung, ob eine Münze fair ist oder gezinkt, hat man sie 100-mal geworfen. In 38 Fällen hat man die nationale Seite und in 62 Fällen die internationale Seite erhalten. Ist dies Ergebnis mit der Annahme „faire Münze“, d.h. mit

$$P(\text{national}) = \frac{1}{2} \text{ verträglich?}$$

Sollte die Münze fair sein, so ist $\mu = 50$ und $\sigma = 5$. In etwa 95 % aller Wurfserien erhält man ein Ergebnis zwischen $\mu - 2 \cdot \sigma = 40$ und $\mu + 2 \cdot \sigma = 60$. Das Ergebnis „38-mal national“ kommt also nur sehr selten vor, so dass man die Münze nicht mehr für fair halten wird. Aber: Auch seltene Ereignisse sind nicht unmöglich, sondern können vorkommen!

Zur Simulation der Binomialverteilung mit GeoGebra:

Eine Münze (mit den Seiten 0/1) wird $n = 100$ -mal geworfen. Jeder Wurf wird mit Zufallszahl[0, 1] simuliert. 100 Würfe bekommt man mit der Liste

$$L = \text{Folge}[\text{Zufallszahl}[0, 1], i, 1, n]$$

(in GeoGebra werden Listen mit dem Folge-Befehl erzeugt).

Die Anzahl der Einsen bekommt man mit

$$\text{Summe}[L]$$

und mit F9 (oder Strg-R in neueren GeoGebra-Versionen) eine neue Wurfserie. Die Befehlsfolge lässt sich durch

$$k = \text{Summe}[\text{Folge}[\text{Zufallszahl}[0, 1], i, 1, n]]$$

zusammenfassen.

Nun will man einen Gesamtüberblick bekommen und $N = 1000$ solcher Wurfserien erzeugen. Das erreicht man mit

$$kk = \text{Folge}[\text{Summe}[\text{Folge}[\text{Zufallszahl}[0, 1], i, 1, n]], j, 1, N].$$

(Zum Ausprobieren wird man mit kleineren Werten für n und N beginnen.)

Nun braucht man das zugehörige Säulendiagramm, das in GeoGebra „Balkendiagramm“ heißt. Hier gibt es zwei Möglichkeiten:

1. Man definiert die x-Werte durch $X = \text{Folge}[i, i, 0, n]$ und die y-Werte durch $Y = \text{Häufigkeit}(X, kk)$ und ruft $a = \text{Balkendiagramm}(X, Y)$ auf – und bekommt als Rückmeldung, dass a undefiniert sei. Das liegt daran, dass Y ein Element weniger hat als X . Man ersetze also Y durch $Z = \text{Anfügen}(Y, 0)$ und rufe $\text{Balkendiagramm}(X, Z)$ auf.
2. Viel einfacher ist es, ohne X und ohne Y und Z einfach

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

aufzurufen. Dabei gibt das zweite Argument („1“) die Balkenbreite an, und das dritte Argument („1/N“) ist ein vertikaler Skalierungsfaktor, den man auch weglassen kann, der aber notwendig ist, wenn man das Balkendiagramm mit der Binomialverteilung vergleichen will.

Die Binomialverteilung als Modellverteilung bekommt man mit

$$\text{Binomial}[n, 0.5].$$

Prognoseintervalle mit GeoGebra:

Der Befehl

`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. Da der Erwartungswert i.a. nicht ganzzahlig ist, ist es sinnvoll, mit $m = \text{round}(\mu)$ zu arbeiten.

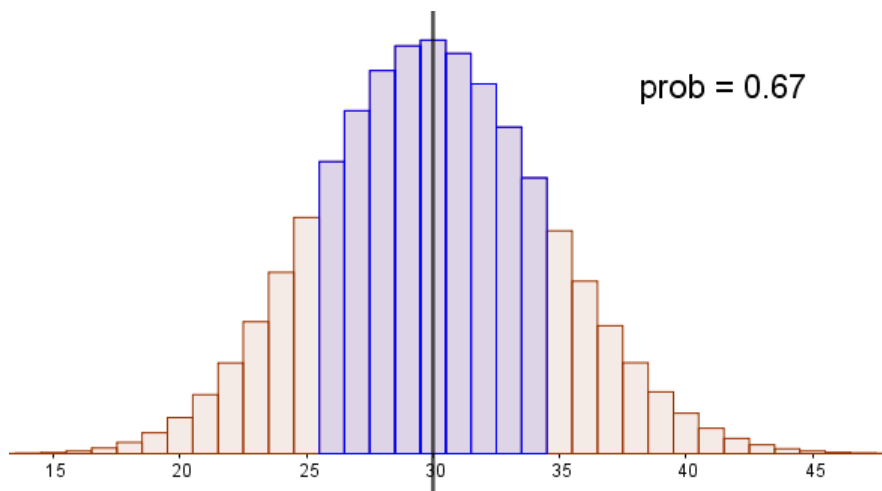
Nun möchte man $\text{prob}(m-g) + \dots + \text{prob}(m+g)$, also sollte g auch über einen Schieberegler veränderbar sein.

Liste der Intervallgrenzen: $L1 = \text{Folge}[k - 0.5, k, m - g, m + g + 1]$

Liste der Intervallhöhen: $L2 = \text{Folge}[\text{Binomial}[n, p, k, \text{false}], k, m - g, m + g]$

(Das letzte Argument („false“) gibt an, dass nicht kumuliert werden soll.)

Mit `Histogramm[L1, L2]` bekommt man in Abhängigkeit von g das Teil-Histogramm.



Die senkrechte Linie bekommt man mit $x = m$.

Nun lässt sich ermitteln, ob ein vermuteter Anteil der Grundgesamtheit bzw. eine vermutete Wahrscheinlichkeit mit einem vorliegenden Messwert verträglich ist: Man entscheide sich für eine Sicherheits-Wahrscheinlichkeit, richte dementsprechend g ein und überprüfe, ob der Messwert im Prognoseintervall liegt.