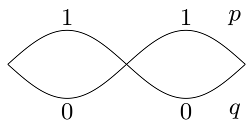
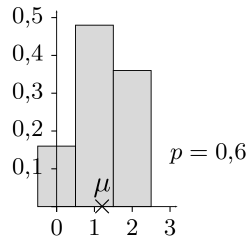
Varianz der Binomialverteilung



k	0	1	2
$P(X=k)$	q^2	$2pq$	p^2
	p_1	p_2	p_3



Wir betrachten eine Bernoulli-Kette der Länge $n = 2$
und berechnen $E(X)$.

$$\begin{aligned} E(X) &= 2pq + 2p^2 \\ &= 2p(q+p) \\ &= 2p \end{aligned}$$

allgemein:

$$\mu = E(X) = np \qquad \text{beachte:} \quad E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n)$$

Der Erwartungswert μ gibt an, wo das „Zentrum" der Verteilung ist, er sagt nichts darüber aus, wie die Verteilung um ihr Zentrum gruppiert ist. Wir benötigen für statistische Anwendungen ein Maß für die Streuung einer Zufallsvariablen um ihren Erwartungswert.

Um eine mittlere Abweichung zu definieren, müssen wir wie beim Erwartungswert die Häufigkeit beachten, mit der die Abweichungen $k_i - \mu$ auftreten. Zudem sind wir nicht am Vorzeichen der Abweichungen interessiert, daher nehmen wir die Quadrate $(k_i - \mu)^2$. Möglich wären auch die Beträge $\mid k_i - \mu \mid$ gewesen, jedoch ist das Rechnen mit ihnen recht unhandlich.

Die Varianz einer Zufallsvariablen X ist die mit den Wahrscheinlichkeiten gewichtete Summe
der Abweichungsquadrate:

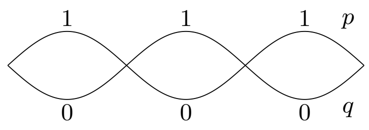
$$V(X) = V(X) = (k_1 - \mu)^2 \cdot p_1 + (k_2 - \mu)^2 \cdot p_2 + \ \ldots\ + (k_n - \mu)^2 \cdot p_n$$

Für die Bernoulli-Kette der Länge $n = 2$ gilt dann:

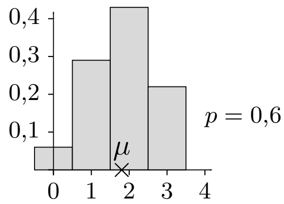
$$\begin{aligned} V(X) &= (0-\mu)^2 \cdot q^2 + (1-\mu)^2 \cdot 2pq + (2-\mu)^2 \cdot p^2 \\ &= \mu^2 \cdot q^2 + (1 - 2\mu + \mu^2) \cdot 2pq + \ldots \\ &= \mu^2 \cdot \underbrace{(q^2 + 2pq + p^2)}_{(p+q)^2 = 1} + \ldots \qquad\qquad \mid\ \mu = 2p \ \text{ einsetzen} \\ &= 4p^2 + 2pq \underbrace{-8p^2q - 8p^3}_{-8p^2(p+q)} + 4p^2 \\ &= 2pq \end{aligned}$$

Es ist zu vermuten, dass für eine Bernoulli-Kette der Länge $n \quad V(X) = npq \quad$ ist
und dass dieses Ergebnis auch dadurch zustande kommt, dass die Varianz für eine Bernoulli-Kette der Länge $n = 1$, nämlich pq, nur mit n multipliziert wird.

Varianz der Binomialverteilung alternativ



k	0	1	2	3
$P(X=k)$	q^3	$3pq^2$	$3p^2q$	p^3
	p_1	p_2	p_3	p_4



Wir betrachten eine Bernoulli-Kette der Länge $n = 3$. Der Erwartungswert μ gibt an, wo das „Zentrum“ der Verteilung ist, er sagt nichts darüber aus, wie die Verteilung um ihr Zentrum gruppiert ist. Wir benötigen für statistische Anwendungen ein Maß für die Streuung einer Zufallsvariablen um ihren Erwartungswert.

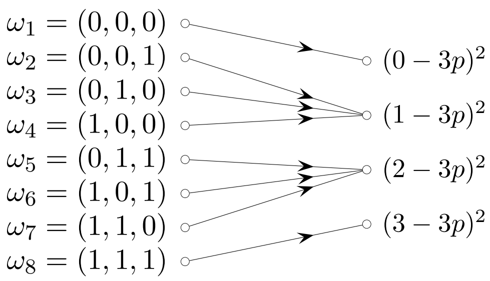
Um eine mittlere Abweichung zu definieren, müssen wir wie beim Erwartungswert die Häufigkeit beachten, mit der die Abweichungen $k_i - \mu$ auftreten. Zudem sind wir nicht am Vorzeichen der Abweichungen interessiert, daher nehmen wir die Quadrate $(k_i - \mu)^2$. Möglich wären auch die Beträge $|\, k_i - \mu \,|$ gewesen, jedoch ist das Rechnen mit ihnen recht unhandlich.

Die Varianz einer Zufallsvariablen X ist die mit den Wahrscheinlichkeiten gewichtete Summe der Abweichungsquadrate:

$$V(X) = (k_1 - \mu)^2 \cdot p_1 + (k_2 - \mu)^2 \cdot p_2 + \ldots + (k_n - \mu)^2 \cdot p_n$$

oder kürzer: $\quad V(X) = E(X-\mu)^2$.

Die Standardabweichung ist dann $\quad \sigma = \sqrt{V(X)}$.



Die Formulierung der Varianz als Erwartungswert der Abweichungsquadrate vereinfacht die Berechnung:

$$\begin{aligned} E(X-\mu)^2 &= E(X^2 - 2X\mu + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - \mu^2 \end{aligned}$$

k	0	1^2	2^2	3^2
$P(X^2=k)$	q^3	$3pq^2$	$3p^2q$	p^3

Mit der Verteilung der Zufallsvariablen X^2 kann die Varianz nun auf diese Weise berechnet werden.

$$\begin{aligned} V(X) &= E(X^2) - \mu^2 \\ &= 3pq^2 + 12p^2q + 9p^3 - (3p)^2 \\ &= 3p\,(\underbrace{q^2 + 4pq}_{q(q+4p) = q(1+3p) = q + 3pq} + 3p^2 - 3p) \\ &= 3p\,(q + \underbrace{3pq + 3p^2 - 3p}_{3p(1-p) + 3p^2 - 3p = 0}) \\ &= 3pq \end{aligned}$$

Es ist zu vermuten, dass für eine Bernoulli-Kette der Länge $n \quad V(X) = npq \quad$ ist
und dass dieses Ergebnis auch dadurch zustande kommt, dass die Varianz für eine Bernoulli-Kette der Länge $n = 1$, nämlich pq, nur mit n multipliziert wird.

Aufg.
In den vorstehenden Umformungen wurde $E(aX) = aE(X)$ und $E(a) = a$ verwendet. Begründe die Berechtigung.

Um die Varianz allgemein für Bernoulli-Ketten der Länge n zu ermitteln,
zerlegen wir wieder die Zufallsvariable X in: $X = X_1 + X_2 + \,...\, + X_n$

oder mit Argumenten: $X(\omega) = X_1(\omega) + X_2(\omega) + \,...\, + X_n(\omega)$,

wobei $X_i(\omega)$ die i-te Stelle von ω angibt.

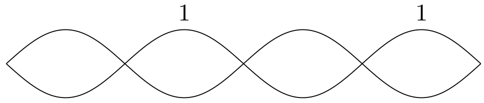
$$\begin{aligned}
V(X) &= E\left((X_1 + X_2 + \,...\, + X_n)^2\right) - (np)^2 \\
&= E\left(X_1^2 + X_2^2 + \,...\, + X_n^2 \;+\; \underbrace{2X_1X_2 + 2X_1X_3 + \,...\, + 2X_{n-1}X_n}_{\binom{n}{2} = \frac{n\,(n-1)}{1 \cdot 2} \quad \textit{Summanden}}\right) - n^2p^2 \\
&= np + n\,(n-1)\,p^2 - n^2p^2 \\
&= np - np^2 \\
&= np\,(1-p) \\
&= npq
\end{aligned}$$

Um die Begründung zu vervollständigen, muss noch
$E(X_i^2) = p$ und $E(X_i \cdot X_j) = p^2$ für $i \neq j$ nachgewiesen werden.

Für die erste Behauptung beachte man, dass die Funktionen X_i^2 und X_i identisch sind,
da X_i und damit X_i^2 nur die Werte 1 und 0 annehmen.

Betrachten wir für die zweite Behauptung die Verteilung von $X_i \cdot X_j$.

$X_i \cdot X_j\ (\omega)$ ist entweder 0 oder 1,
1 genau dann, wenn $X_i(\omega) = 1$ und $X_j(\omega) = 1$ ist.

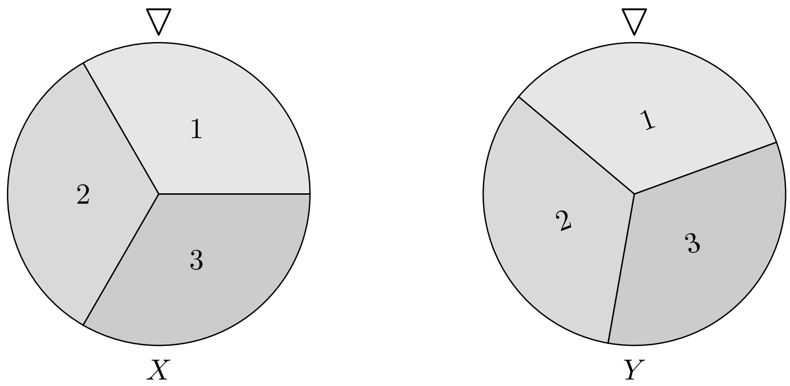


Nun kann $E(X_i \cdot X_j) = p^2$ für $i \neq j$ eingesehen werden.

k	0	1
$P(X_i \cdot X_j = k)$	$1 - p^2$	p^2

Allgemein kann $E(X \cdot Y) = E(X) \cdot E(Y)$ für sogenannte unabhängige Zufallsvariablen (der Wert der einen Zufallsvariablen sagt nichts über den Wert der anderen) gezeigt werden. Für diese Zufallsvariablen gilt dann auch $V(X + Y) = V(X) + V(Y)$, so dass mit den obigen Bezeichnungen gilt:
$V(X) = V(X_1 + X_2 + \,...\, + X_n) = V(X_1) + V(X_2) + \,...\, + V(X_n) = npq\,.$
Von besonderer Bedeutung ist: $V(\frac{1}{n}X) = \frac{1}{n^2}V(X) = \frac{pq}{n}$ (siehe Stichprobenmittel).

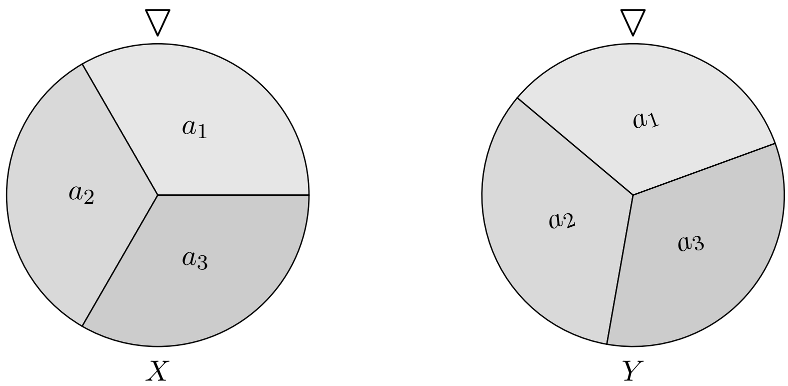
Varianz von $X + Y$



Aufg.
Bestimme $V(X)$ und $V(X+Y)$.

$$\text{Zur Erinnerung} \qquad \begin{aligned} V(X) &= (a_1-\mu)^2 \cdot p_1 + (a_2-\mu)^2 \cdot p_2 + (a_3-\mu)^2 \cdot p_3 \\ &= a_1^2 p_1 + a_2^2 p_2 + a_3^2 p_3 - \mu^2 \quad (= E(X^2) - \mu^2) \end{aligned}$$

Und nun etwas allgemeiner.



$$\begin{aligned} V(X+Y) &= (a_1+b_1)^2 \cdot p_1 p_1 \;+\; (a_1+b_2)^2 \cdot p_1 p_2 \;+\; (a_1+b_3)^2 \cdot p_1 p_3 \;+\; \ldots \;-4\mu^2 \\ &= a_1^2 \cdot p_1(p_1+p_2+p_3) \;+\; a_2^2 \cdot p_2 \;+\; a_3^2 \cdot p_3 \quad + \quad b_1^2 \cdot p_1 \;+\; b_2^2 \cdot p_2 \;+\; b_3^2 \cdot p_3 \\ &\qquad + \underbrace{2a_1 b_1 \cdot p_1 p_1 \;+\; 2a_1 b_2 \cdot p_1 p_2 \;+\; 2a_1 b_3 \cdot p_1 p_3 \;+\; \ldots}_{2(a_1 p_1 + a_2 p_2 + a_3 p_3)\cdot(b_1 p_1 + b_2 p_2 + b_3 p_3)} \;-4\mu^2 \;\; -4\mu^2 \\ &= a_1^2 p_1 \;+\; a_2^2 p_2 \;+\; a_3^2 p_3 \quad + \quad b_1^2 p_1 \;+\; b_2^2 p_2 \;+\; b_3^2 p_3 \quad + \quad 2\mu \cdot \mu \;-\; 4\mu^2 \\ &= a_1^2 p_1 + a_2^2 p_2 + a_3^2 p_3 - \mu^2 \quad + \quad b_1^2 p_1 + b_2^2 p_2 + b_3^2 p_3 - \mu^2 \\ &= V(X) + V(Y) \end{aligned}$$

Zufallsvariable X und Y sind unabhängig, falls für alle Paare $(a_i \mid b_i)$ gilt:
$P(X = a_i, Y = b_i) \;=\; P(X = a_i) \cdot P(Y = b_i)$, dann folgt $E(X \cdot Y) = E(X) \cdot E(Y)$.