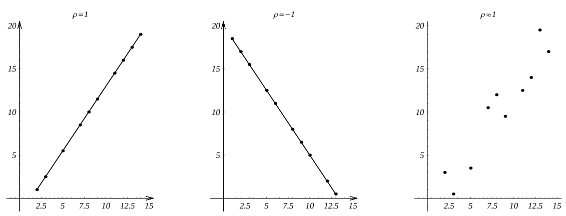
Beispiel F.41

(Illustration von speziellen gemeinsamen Verteilungen und Korrelation)
Die hier diskutierten Beispiele für gemeinsame Verteilungen sind in der folgenden Abbildung graphisch dargestellt. Die Werte der jeweiligen Verteilungen mit positiver Wahrscheinlichkeit sind als Punkte in die x-y-Ebene eingezeichnet, wobei (x, y) Werte der Funktion $X \times Y$ sind. Eine solche Darstellung könnte noch präzisiert werden, indem man zu jedem Punkt die Wahrscheinlichkeit schreibt, was bei einer kleinen Anzahl von Punkten noch übersichtlich wäre. Der Einfachheit halber habe hier jeweils jeder Punkt die gleiche Wahrscheinlichkeit.

1. Sei X eine Zufallsvariable mit Varianz $\sigma_X^2 > 0$ und sei $Y = aX + b$ mit $a \neq 0$. Wir berechnen unter Verwendung der Sätze F.33 und F.36 den Korrelationskoeffizienten von X und Y.

$$\begin{aligned} \operatorname{Var}(Y) &= a^2 \operatorname{Var}(x), \qquad \Rightarrow \qquad \sigma_Y = |a| \cdot \sigma_X, \\ \operatorname{Cov}(X, Y) &= \operatorname{Cov}(X, aX + b) = a \operatorname{Cov}(X, X) = a\sigma_X^2, \\ \rho_{X,Y} &= \frac{a\sigma_X^2}{\sigma_X |a| \sigma_X} = \operatorname{sign}(a). \end{aligned}$$

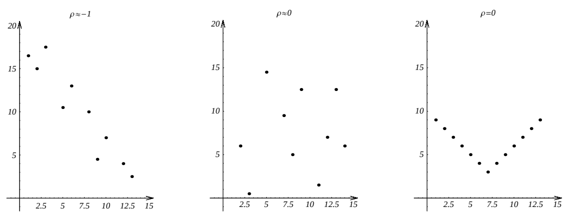
Der Korrelationskoeffizient $\rho_{X,Y}$ ist also 1 oder -1, je nachdem, ob a positiv oder negativ ist. In den Abbildungen (a) und (b) sind Beispiele für solche gemeinsamen Verteilungen von X und Y dargestellt. Die Punkte der gemeinsamen Verteilung liegen auf einer Geraden. Wir bemerken auch, dass im Fall $a = 0$, also $Y = b$, die Zufallsvariable Y deterministisch ist und somit Varianz Null hat. Auch hier liegen die Punkte der gemeinsamen Verteilung von X und Y auf einer Geraden (nicht abgebildet), aber der Korrelationskoeffizient ist im Sinne von Definition F.35 nicht definiert.



(a) Die Punkte liegen auf einer steigenden Geraden

(b) Die Punkte liegen auf einer fallenden Geraden

(c) Die Punkte streuen schwach um eine steigende Gerade



(d) Die Punkte streuen schwach um eine fallende Gerade

(e) Punktwolke ohne zuzuordnender Gerade

(f) Nicht-lineare funktionale Abhängigkeit

Abbildung: Illustration von Korrelationskoeffizienten mit Hilfe von gemeinsamen Verteilungen

2. In den Abbildungen (c) und (d) sind die gemeinsamen Verteilungen von Zufallsvariablen dargestellt, deren Korrelationskoeffizient nahe bei 1 bzw. nahe bei -1 liegt. Die Punkte liegen zwar nicht auf einer Geraden, aber man kann könnte jeder der Verteilungen eine Gerade zuordnen, von der die Punkte „nicht allzu sehr" abweichen. Eine solche Zuordnung geschieht z.B. mit Hilfe von *linearer Regression*.
3. Der in Abbildung (e) dargestellten Verteilung wäre optisch nur schwer eine Gerade zuzuordnen. Der Korrelationskoeffizient in diesem Beispiel liegt nahe bei 0.

4. Wir betrachten nun noch ein sehr spezielles Beispiel. Die gemeinsame Verteilung von X und Y sei

$$P_{X\times Y}(-1,1) = P_{X\times Y}(0,0) = P_{X\times Y}(1,1) = \frac{1}{3}$$

dargestellt. Die Kovarianz von X und Y ist

$$\mathrm{Cov}(X,Y) \quad = \quad \sum_{(x,y)} x\cdot y\cdot P_{X\times Y}(x,y) = \frac{1}{3}\cdot(1\cdot(-1)+0\cdot 0+1\cdot 1) = 0.$$

Dabei haben wir in der ersten Zeile über alle Werte (x,y) mit positiver Wahrscheinlichkeit summiert. Die beiden Zufallsvariablen sind also nicht korreliert. Ihr Korrelationskoeffizient ist gleich 0.
Wir bemerken noch, dass Y *nicht* unabhängig von X ist (s. Definition F.30). Im Gegenteil, es besteht sogar ein funktionaler Zusammenhang zwischen beiden Variablen. Kennt man den Wert von X, so auch den von Y. Dieser Zusammenhang ist aber *nicht* linear (vgl.16).
Analog zu diesem Beispiel sind die Zufallsvariablen, deren gemeinsame Verteilung in Abbildung (f) dargestellt ist, unkorreliert, obwohl ein funktionaler Zusammenhang zwischen ihnen besteht.

Das schwache Gesetz der großen Zahlen

In diesem Abschnitt formulieren wir mit Satz F.43 eine Version des schwachen Gesetzes der großen Zahlen, das insbesondere einen Zusammenhang zwischen dem abstrakt eingeführten Begriff der Wahrscheinlichkeit und relativen Häufigkeiten bei einer Folge aus lauter voneinander unabhängigen Zufallsexperimenten herstellt, die alle den gleichen Erwartungswert haben.

Der folgende Satz liefert uns eine Abschätzung für die Wahrscheinlichkeit der Abweichung einer Zufallsvariablen von ihrem Erwartungswert um mehr als eine vorgegebene Konstante. Diese Abschätzung benutzt nur die Varianz der Zufallsvariablen, ohne irgendwelche weiteren Bedingungen an die Verteilung zu stellen, und ist damit anwendbar sobald man die Varianz kennt. Allerdings ist sie in vielen Fällen auch nur sehr grob oder gar völlig nutzlos, z.B. wenn die rechte Seite in (21) größer gleich 1 ist. Dennoch liefert sie uns einen sehr einfachen Beweis des schwachen Gesetzes der großen Zahlen.

Satz F.42 (Tschebyscheff-Ungleichung)

Sei X eine reelle Zufallsvariable auf (Ω, P). Dann gilt für jedes $\epsilon > 0$:

$$P(|X - E(X)| > \epsilon) \leq \frac{\mathrm{Var}(X)}{\epsilon^2}. \tag{21}$$

Beweis: Sei $Z = X - E(X)$. Wir definieren zu Z^2 eine Minorante, d.h. eine Zufallsvariable Y mit $Y(\omega) \leq (Z(\omega))^2$:

$$Y(\omega) := \begin{cases} 0 & \text{für } |Z(\omega)| < \epsilon, \\ \epsilon^2 & \text{für } |Z(\omega)| \geq \epsilon. \end{cases}$$

Mit Hilfe dieser Minorante können wir den Erwartungswert von Z^2 nach unten abschätzen:

$$\begin{aligned} \mathrm{Var}(X) &= E(Z^2) \geq E(Y) \\ &= \epsilon^2 \cdot P(Y = \epsilon^2) \\ &= \epsilon^2 \cdot P(|X - E(x)| \geq \epsilon). \end{aligned}$$

$\square$

Satz F.43 (Das schwache Gesetz der großen Zahlen)

Seien $X_1, X_2, \ldots$ unabhängige Zufallsvariablen mit den gleichen Erwartungswerten $E(X_1)$ und $\mathrm{Var}(X_i) \leq M$. Dann gilt

$$P\left(\left|\frac{1}{n}(X_1 + \cdots + X_n) - E(X_1)\right| \geq \epsilon\right) \leq \frac{M}{\epsilon^2 n}, \qquad (22)$$

insbesondere

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}(X_1 + \cdots + X_n) - E(X_1)\right| \geq \epsilon\right) = 0.$$

Beweis: Sei $S^{(n)} = \frac{X_1 + \cdots + X_n}{n}$. Dann ist $E(S^{(n)}) = E(X_1)$, und

$$\mathrm{Var}(S^{(n)}) \quad = \quad \frac{1}{n^2}\mathrm{Var}(X_1 + \cdots + X_n) = \frac{1}{n^2} \cdot n \cdot M = \frac{M}{n},$$

wobei wir im vorletzten Schritt die Unabhängigkeit von $(X_i)_i$ verwendet haben. Die Behauptung folgt nun aus der Tschebyscheff-Ungleichung. □

Beispiel F.44 (**n**-maliges Würfeln)

In Beispiel F.34 hatten wir schon den Erwartungswert $E(X_i) = 3.5$ und in Beispiel F.37 die Varianz für die Augenzahl beim einfachen Wurf des Laplace-Würfels berechnet. Wir betrachten nun zum n-fachen Wurf die gemittelte Summe $S^{(n)} = \frac{1}{n}(X_1 + \ldots + X_n)$ der Augenzahlen. Nach dem schwachen Gesetz der großen Zahlen (Satz F.43) ist zu einer vorgegebenen Schranke $\epsilon > 0$ bei häufigem Würfeln die Wahrscheinlichkeit, dass die beobachtete mittlere Augenzahl um mehr als ϵ von ihrem Erwartungswert $E(S^{(n)}) = 3.5$ abweicht klein, vorausgesetzt n ist hinreichend groß. Doch wie oft muss man z.B. würfeln, damit für $\epsilon = 0.1$ die Wahrscheinlichkeit einer Abweichung kleiner ist als 0.01? Hier geben wir mit einer sehr groben Abschätzung zufrieden, die auf der Tschebyscheff-Ungleichung (Satz F.42) beruht, und wollen damit nur (22) an einem Beispiel illustrieren.

Wir erhalten mit $M = \frac{35}{12}$ und $\epsilon = 0.1$:

$$P\left(\left|S^{(n)} - 3.5\right| \geq 0.1\right) \leq \frac{35}{12 \cdot 0.1 \cdot n}. \qquad (23)$$

Die rechte Seite der Abschätzung (23) ist kleiner oder gleich 0.01, falls $n \geq 4200$. D.h. wenn man 4200 mal oder noch häufiger würfelt, dann weicht die mittlere Augenzahl mit einer Wahrscheinlichkeit von höchstens 1% um 0.1 oder mehr vom ihrem Erwartungswert ab.

Bemerkung: Zum schwachen Gesetz der großen Zahlen

Das schwache Gesetz der großen Zahlen sagt, dass in der Situation in Satz F.43 für „große“ n der gemittelte Wert $S^{(n)} = \frac{1}{n}(X_1 + \ldots + X_n)$ mit „großer“ Wahrscheinlichkeit (also einer solchen nahe bei 1) vom Erwartungewert $E(S^{(n)}) = E(X_i)$ „nicht stark“ abweicht. Wenn man den Erwartungswert der Augenzahl bei einem Würfel statistisch durch viele Würfe ermitteln will, führt man aber z.B. *eine* recht lange Versuchsreihe von Würfen durch, die einer Folge $X_1, X_2, \ldots$ entspricht und betrachtet entsprechend die Folge der gemittelten Werte $S^{(1)}, S^{(2)}, \ldots$ Das schwache Gesetz der großen Zahlen sagt, dass für ein vorgegbenes ϵ für hinreichend große n die Wahrscheinlichkeit für eine Abweichung $|S^{(n)} - E(X_1)| > \epsilon$ „klein“ ist, schließt aber nicht aus, das für eine betrachtete Folge von Würfen diese Abweichung „immer wieder mal“ auftritt. Aber das **starke Gesetz der großen Zahlen**, das wir hier nicht als mathematischen Satz formulieren, sagt, dass für *fast alle* Folgen (von Würfen) die Folge der Werte von $S^{(n)}$ tatsächlich gegen $E(X_1)$ konvergiert. Das bedeutet, die Wahrscheinlichkeit für diese Konvergenz ist gleich 1.