

3 Zufallsvariablen und Verteilungen

3.1 Grundlegende Konzepte

Gegeben sei ein Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) . Statt des einzelnen Ergebnisses $\omega \in \Omega$ des hierdurch beschriebenen Zufallsexperimentes ist man häufig nur am Wert $X(\omega) (\in \mathbb{N}, \mathbb{Z}, \mathbb{R}, \dots)$ einer Messgröße X interessiert, wie zum Beispiel Temperatur oder Aktienkurs. Im Zusammenhang mit einer Messgröße X möchte man dann insbesondere den Ereignissen

$$\{X \in B\} = X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\} = X \text{ nimmt einen Wert in } B \text{ an}$$

für $B \in \mathcal{B}(\mathbb{R})$ eine Wahrscheinlichkeit zuordnen können, zB $P(\{X \leq b\})$, $P(\{X \geq a\})$, ... Hierzu ist es notwendig, dass für $B \in \mathcal{B}$ das Ereignis $\{X \in B\}$ in der σ -Algebra \mathcal{A} liegt. Dies motiviert:

Definition 3.1. Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und (E, \mathcal{E}) ein messbarer Raum, (zB $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$). Eine Abbildung $X : \Omega \rightarrow E$ heißt (*E-wertige*) **Zufallsvariable** (kurz: *ZV*), falls

$$\{X \in B\} \in \mathcal{A} \quad \text{für alle } B \in \mathcal{E}.$$

Der Wert $X(\omega)$ heißt **Realisierung** der Zufallsvariablen X (zum Ergebnis ω).

Bemerkung 3.2. (i) Ist Ω eine diskrete Menge (und $\mathcal{A} = \mathcal{P}(\Omega)$), so ist **jede** Abbildung $X : \Omega \rightarrow E$ eine Zufallsvariable.

(ii) Ist $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, so heißt X **reellwertige Zufallsvariable** und im Falle $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ **d-dimensionaler Zufallsvektor**.

(iii) Sind X_1, \dots, X_n reellwertige Zufallsvariablen und ist $f : \mathbb{R}^n \rightarrow \mathbb{R}$ stetig (es reicht: Borel-messbar), so ist auch $f(X_1, \dots, X_n)$ wieder eine reellwertige Zufallsvariable. Insbesondere sind also

$$\alpha X_k \text{ für } \alpha \in \mathbb{R}, X_1 + \dots + X_n, X_k \cdot X_l$$

wieder reellwertige Zufallsvariablen. Einen Beweis findet man in der Literatur zur Maß- und Integrationstheorie (zB [Ba91]).

(iv) Zur Vereinfachung der Notation hat es sich eingebürgert, dass man einfach $P(X \in B)$ anstatt $P(\{X \in B\})$ schreibt, und dann auch $P(X \in A, Y \in B)$ für $P(\{X \in A\} \cap \{Y \in B\})$, usw.

(v) Für jedes Ereignis $A \in \mathcal{A}$ definiert

$$1_A(\omega) = \begin{cases} 1 & \text{für } \omega \in A \\ 0 & \text{sonst,} \end{cases}$$

eine Zufallsvariable (Übungsaufgabe). Sie zeigt durch ihren Wert an, ob das Ergebnis A eingetreten ($X = 1$) oder nicht eingetreten ($X = 0$) ist und heißt daher **Indikatorfunktion** (zum Ereignis A).

Satz 3.3. (und Definition) Es seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, (E, \mathcal{E}) ein messbarer Raum und $X : \Omega \rightarrow E$ eine Zufallsvariable. Dann wird durch

$$P_X(B) := P(X \in B), B \in \mathcal{E}$$

ein Wahrscheinlichkeitsmaß P_X auf (E, \mathcal{E}) definiert (auch mit $P \circ X^{-1}$ bezeichnet). P_X heißt **Verteilung** von X (unter P).

Beweis. Zunächst ist die Mengenfunktion P_X wohldefiniert, da $\{X \in B\} \in \mathcal{A}$ für alle $B \in \mathcal{E}$. Sie ist auch nichtnegativ $P_X(B) \geq 0$ und normiert $P_X(E) = P(X \in E) = P(\Omega) = 1$. Bleibt also wieder lediglich die σ -Additivität zu zeigen. Dazu sei $(B_n)_{n \in \mathbb{N}}$ eine paarweise disjunkte Folge von Ereignissen in \mathcal{E} . Dann ist die Folge $(\{X \in B_n\})_{n \in \mathbb{N}}$ der Urbilder einer Folge paarweise disjunkter Ereignisse in \mathcal{A} und aus der σ -Additivität von P und aus der offensichtlichen Identität $\{X \in \bigcup_{n \geq 1} B_n\} = \bigcup_{n \geq 1} \{X \in B_n\}$ ergibt sich

$$\begin{aligned} P_X\left(\bigcup_{n \geq 1} B_n\right) &= P\left(X \in \bigcup_{n \in \mathbb{N}} B_n\right) = P\left(\bigcup_{n \geq 1} \{X \in B_n\}\right) \\ &= \sum_{n \geq 1} P(X \in B_n) = \sum_{n \geq 1} P_X(B_n). \end{aligned}$$

□

Wir wollen im folgenden zwei Spezialfälle näher ausführen.

3.1.1 Diskrete Zufallsvariablen und Verteilungen

Ist E diskret (und $\mathcal{E} = \mathcal{P}(E)$), so kann die Verteilung P_X vollständig durch die zugehörige Zähldichte

$$p_X(b) = P(X = b) = P(\text{X nimmt den Wert } b \text{ an}), b \in E$$

beschrieben werden. Dies gilt auch für den Fall, in dem nur das Bild $X(\Omega)$ eine diskrete Menge ist.

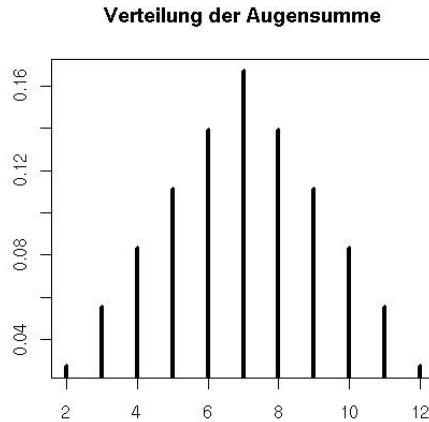
Beispiel 3.4. Beim zweimaligen Würfel eines fairen Würfels sei X die Augensumme. X ist eine Zufallsvariable mit Werten in der Menge $\{2, 3, \dots, 12\}$. Für die zugehörige Verteilung P_X ergibt sich:

$$\begin{aligned} p_X(2) &= P(\{(k, l) \in \Omega : k + l = 2\}) = P(\{(1, 1)\}) = \frac{1}{36} \\ p_X(12) &= P(\{6, 6\}) = \frac{1}{36} \end{aligned}$$

und für die übrigen Werte

$$\begin{aligned} p_X(3) &= p_X(11) = \frac{2}{36}, p_X(4) = p_X(10) = \frac{3}{36} \\ p_X(5) &= p_X(9) = \frac{4}{36}, p_X(6) = p_X(8) = \frac{5}{36} \\ p_X(7) &= \frac{6}{36}. \end{aligned}$$

P_X kann mit Hilfe eines **Stabdiagramms** graphisch veranschaulicht werden.



3.1.2 Reellwertige Zufallsvariablen und Verteilungen

Die Verteilung reellwertiger Zufallsvariablen X kann durch ihre **Verteilungsfunktion**

$$F_X(x) := P(X \leq x) = P_X(]-\infty, x]) , \quad x \in \mathbb{R}$$

beschrieben werden.

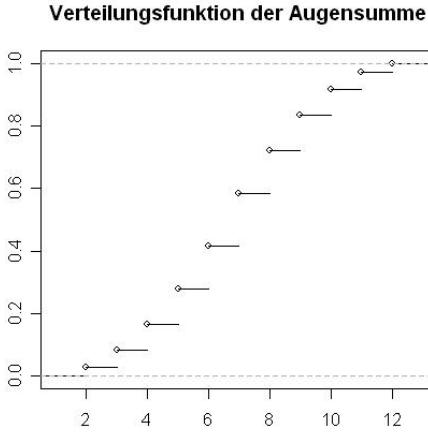
Beispiel 3.5. (i) Eine Zufallsvariable X , die nur einen Wert $c \in \mathbb{R}$ annimmt, nennt man deterministisch (oder entartet), da ihr Wert vom Ausgang des zugrundeliegenden Zufallsexperimentes gar nicht abhängt. Für die zugehörige Verteilungsfunktion gilt

$$F_X(x) = \begin{cases} 0 & \text{für } x < c \\ 1 & \text{für } x \geq c . \end{cases}$$

(ii) Ist X eine diskrete Zufallsvariable, so hat die zugehörige Verteilungsfunktion die Darstellung

$$F_X(x) = \sum_{y \leq x} P_X(\{y\}) = \sum_{y \leq x} P(X = y) .$$

Sie ist also eine stückweise konstante, monoton wachsende Treppenfunktion mit Werten zwischen 0 und 1. Ihre Sprungstellen sind genau diejenigen Werte x die von X mit positiver Wahrscheinlichkeit $P(X = x) > 0$ angenommen werden und die Höhe der Sprungstellen ist gleich $P(X = x)$.



(iii) Ist X gleichverteilt auf $[a, b]$, so gilt für $x \in \mathbb{R}$ aufgrund von (1.16)

$$F_X(x) = P(X \leq x) = P_X(\] - \infty, x]) = \frac{\|] - \infty, x] \cap [a, b] \|}{\|[a, b]\|} = \begin{cases} 0 & \text{für } x < a \\ \frac{x-a}{b-a} & \text{für } x \in [a, b[\\ 1 & \text{für } x \geq b. \end{cases}$$

Satz 3.6. Die Verteilungsfunktion $F = F_X$ einer reellwertigen Zufallsvariablen X besitzt folgende Eigenschaften:

- (a) F ist monoton wachsend,
- (b) F ist rechtsstetig,
- (c) $0 \leq F \leq 1$, $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow \infty} F(x) = 1$.

Beweis. (a) Die Monotonie ist offensichtlich, denn aus $y < x$ folgt $\{X \leq y\} \subseteq \{X \leq x\}$ und damit

$$F_X(y) = P(X \leq y) \leq P(X \leq x) = F_X(x).$$

(b) Wir haben zu zeigen, dass $F(x) = \lim_{y \downarrow x} F(y)$ gilt. Dazu sei (y_n) eine monoton fallende Folge mit $\lim_{n \rightarrow \infty} y_n = x$. Es folgt $\] - \infty, y_n] \downarrow \] - \infty, x]$, d.h. $\] - \infty, y_n] \supseteq \] - \infty, y_{n+1}]$ und $\bigcap_{n=1}^{\infty} \] - \infty, y_n] = \] - \infty, x]$. Aus Satz 1.9 ergibt sich nunmehr dass

$$\lim_{n \rightarrow \infty} F(y_n) = \lim_{n \rightarrow \infty} P_X(\] - \infty, y_n]) = P_X(\] - \infty, x]) = F(x).$$

(c) Wiederum mithilfe von Satz 1.9 ergibt sich wegen

$$\begin{aligned} x_n \downarrow -\infty &\text{ impliziert } \] - \infty, x_n] \downarrow \emptyset \\ \lim_{n \rightarrow \infty} F(x_n) &= \lim_{n \rightarrow \infty} P_X(\] - \infty, x_n]) = P_X(\emptyset) = 0, \end{aligned}$$

und analog wegen

$$\begin{aligned} x_n \uparrow +\infty &\text{ impliziert } \] - \infty, x_n] \uparrow \mathbb{R} \\ \lim_{n \rightarrow \infty} F(x_n) &= \lim_{n \rightarrow \infty} P_X(\] - \infty, x_n]) = P_X(\mathbb{R}) = 1. \end{aligned}$$

□

Definition 3.7. : Eine Funktion $F : \mathbb{R} \rightarrow [0, 1]$ mit (a)- (c) aus Satz 3.6 heißt **Verteilungsfunktion**.

Bemerkung 3.8. (i) In der Maßtheorie zeigt man, dass zu jeder Verteilungsfunktion F genau ein Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ existiert mit $P(]-\infty, x]) = F(x)$. Ebenfalls kann man zu jeder Verteilungsfunktion F einen Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) , und hierauf eine reellwertige Zufallsvariable X konstruieren, deren Verteilungsfunktion F_X mit der vorgegebenen Verteilungsfunktion F übereinstimmt.

(ii) Eine Verteilung ist durch die Angabe ihrer Verteilungsfunktion eindeutig bestimmt: Sind X und Y zwei reellwertige Zufallsvariablen mit $F_X = F_Y$, so ergibt sich hieraus $P_X = P_Y$.

(iii) Ist F_X Verteilungsfunktion von X , so folgt für $x \in \mathbb{R}$

$$P(X = x) = \lim_{n \rightarrow \infty} P\left(\left]x - \frac{1}{n}, x\right]\right) = \lim_{n \rightarrow \infty} F(x) - F\left(x - \frac{1}{n}\right) = F(x) - F(x-).$$

Insbesondere ist F_X stetig in x genau dann wenn $P(X = x) = 0$, d.h. genau dann wenn X den Wert x nur mit Wahrscheinlichkeit 0 annimmt.

(iv) Ist $f : \mathbb{R} \rightarrow [0, \infty[$ eine Dichte, so ist die Funktion

$$F(x) = \int_{-\infty}^x f(y) dy, \quad x \in \mathbb{R} \tag{3.1}$$

eine Verteilungsfunktion. F ist offensichtlich stetig, ja sogar stetig differenzierbar, falls f eine stetige Dichte ist. Jede Verteilungsfunktion F , für die eine Dichte existiert, so dass (3.1) gilt, heißt **absolutstetige** Verteilungsfunktion. Ist X eine Zufallsvariable mit absolutstetiger Verteilungsfunktion F_X und zugehöriger Dichte f_X , so folgt aus (3.1)

$$P(X \in B) = \int_B f_X(y) dy \tag{3.2}$$

für jede Borelmenge $B \in \mathcal{B}(\mathbb{R})$.

Beispiel 3.9. Gleichverteilung auf $[a, b]$ mit Dichte

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } x \in [a, b] \\ 0 & \text{sonst.} \end{cases}$$

3.2 Wichtige diskrete Verteilungen

Bernoulli-Verteilung

Die Verteilung der Indikatorfunktion 1_A des Ereignisses $A \in \mathcal{A}$ nimmt zwei Werte 1 und 0 an. Wir interpretieren das Ereignis $\{1_A = 1\} = A$ als "Erfolg". Dementsprechend bezeichnen wir

$$p := P(X = 1) = P(A)$$

als **Erfolgswahrscheinlichkeit**. Entsprechend gilt für die Wahrscheinlichkeit eines Mißerfolges

$$P(X = 0) = P(A^c) = 1 - P(A) = 1 - p.$$

Definition 3.10. Es sei $p \in [0, 1]$. Das durch die Zähldichte $p : \{0, 1\} \rightarrow [0, 1]$

$$p(1) = p, \text{ und } p(0) = 1 - p$$

definierte Wahrscheinlichkeitsmaß auf $\{0, 1\}$ heißt **Bernoulli-Verteilung zu p** . Zufallsexperimente, die nur zwei mögliche Ausgänge kennen, nennt man entsprechend **Bernoulli-Experimente**.

Beispiele für Bernoulli-Experimente

- Werfen einer fairen Münze: $P(\text{Kopf}) = P(\text{Zahl}) = \frac{1}{2}$
- Geschlecht eines Neugeborenen: $P(\text{weiblich}) = 0.47, P(\text{männlich}) = 0.53$
- Ziehen einer Kugel aus einer Urne mit s schwarzen und w weißen Kugeln:

$$P(\text{gez. Kugel schwarz}) = \frac{s}{s+w}$$

Binomialverteilung

Es seien X_1, \dots, X_n unabhängige Zufallsvariablen, die alle Bernoulli-verteilt sind zu p .

Wir können X_i als Ausgang eines Bernoulli Experiments mit Erfolgswahrscheinlichkeit p interpretieren, wobei die Folge der n Experimente unabhängig ist. Dann zählt die Zufallsvariable

$$S_n := X_1 + \dots + X_n \in \{0, \dots, n\}$$

die **Gesamtanzahl der Erfolge**.

Für die Verteilung P_{S_n} der Summe S_n gilt dann

$$p_{S_n}(k) = P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} =: b(k; n, p)$$

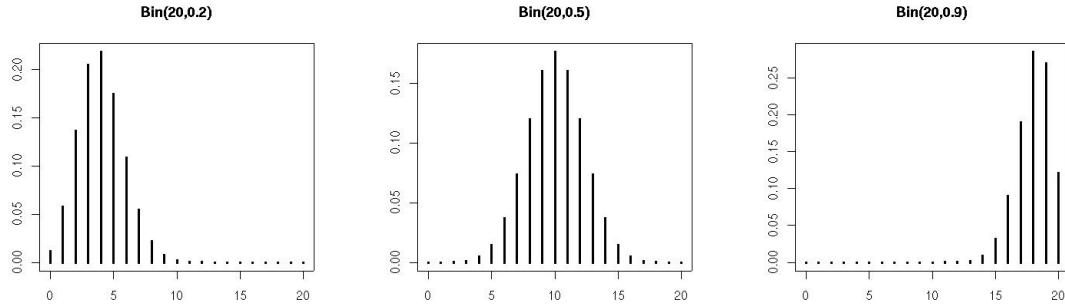
Hierbei ist $\binom{n}{k}$ gerade die Anzahl der n -Tupel mit genau k Einsen (und $n - k$ Nullen), p^k die Wahrscheinlichkeit für k Erfolge und $(1-p)^{n-k}$ die Wahrscheinlichkeit für $n - k$ Mißerfolge.

Definition 3.11. Es sei $n \in \mathbb{N}$ und $p \in [0, 1]$. Das durch die Zähldichte

$$\begin{aligned} b(\cdot; n, p) &: \{0, \dots, n\} \rightarrow [0, 1] \\ k &\mapsto \binom{n}{k} p^k (1-p)^{n-k} \end{aligned}$$

definierte Wahrscheinlichkeitsmaß auf $\{0, \dots, n\}$ heißt **Binomialverteilung zu n und p** und wird mit $\text{Bin}(n, p)$ bezeichnet.

Wir haben insbesondere gesehen: Bei einer Folge von n unabhängigen Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit p ist die Summe der Erfolge binomialverteilt mit Parameter n und p .



Geometrische Verteilung

Wie groß ist die Wahrscheinlichkeit, dass man mit einem fairen Würfel genau k Versuche benötigt, bis zum ersten Mal eine 6 gewürfelt wird?

Für $k = 1$ ist die gesuchte Wahrscheinlichkeit offensichtlich $\frac{1}{6}$, für $k = 2$ ist gleich $\frac{5}{6} \cdot \frac{1}{6}$ und für allgemeine k offensichtlich $(\frac{5}{6})^{k-1} \frac{1}{6}$.

Allgemeiner Gegeben sei eine Folge X_1, X_2, \dots unabhängig Bernoulli-verteilter Zufallsvariablen mit gemeinsamer Erfolgswahrscheinlichkeit $p > 0$. Definiere die **Wartezeit auf den ersten Erfolg**

$$T := \min\{k \geq 1 : X_k = 1\}.$$

Wie im Fall der Wartezeit auf die erste 6 beim Würfeln mit einem fairen Würfel, erhalten wir für die Verteilung von T

$$\begin{aligned} P(T = k) &= P(X_1 = 0, X_2 = 0, \dots, X_{k-1} = 0, X_k = 1) \\ &= P(X_1 = 0) \cdot P(X_2 = 0) \cdot \dots \cdot P(X_{k-1} = 0) \cdot P(X_k = 1) \\ &= (1-p)^{k-1} \cdot p \end{aligned}$$

für $k = 1, 2, 3, \dots$

Definition 3.12. Es sei $p \in]0, 1]$. Das durch die Zähldichte

$$\begin{aligned} g_p : \mathbb{N} &\mapsto [0, 1] \\ k &\mapsto (1-p)^{k-1} p \end{aligned}$$

definierte Wahrscheinlichkeitsmaß auf \mathbb{N} heißt **geometrische Verteilung zu p** und wird mit **Geom (p)** bezeichnet.

Eine wichtige Eigenschaft der geometrischen Verteilung ist ihre Gedächtnislosigkeit im Sinne der Aussage des folgenden Satzes:

$$P(T > k+n \mid T > n) = P(T > k) \quad \forall n \in \mathbb{N}_0$$

Satz 3.13. Es sei T geometrisch verteilt. Dann gilt für $k \geq 2$

$$P(T \geq k-1+n \mid T \geq n) = P(T \geq k) \quad \forall n \in \mathbb{N}. \quad (3.3)$$

Idee: { Umgekehrt: Ist $T : \Omega \rightarrow \mathbb{N}$ eine Zufallsvariable mit der Eigenschaft (3.3), d.h. gedächtnislos, so ist T geometrisch verteilt.

$$a_m := P(T > m), \quad m = 0, 1, 2, \dots$$

$$a_{k+n} = P(T > k+n) = P(T > k+n \mid T > n) \cdot P(T > n) \stackrel{3.3}{=} P(T > k) P(T > n) = a_k \cdot a_n.$$

$$\begin{aligned} \text{Sehe } 1-p = a_1 &\Rightarrow a_{k+n} = a_{k+n-k} \cdot a_k = \dots = a_n^k \quad \forall k, n \in \mathbb{N} \\ &\Rightarrow P(T > k) = a_1^k \\ &\Rightarrow P(T = k) = P(T > k-1) - P(T > k) \\ &= a_1^{k-1} - a_n^k = a_1^{k-1} (1 - a_n^k) = (1-p)^{k-1} \cdot p \end{aligned}$$

↓

noch zu zeigen: $a_1 \leq 1$ und $a_n > 0$: Ang. $a_1 = 0 \Rightarrow a_k = a_1^k = 0 \Rightarrow k = 0 \Rightarrow p = 0$

Beweis. Es sei T geometrisch verteilt. Dann gilt für $m \in \mathbb{N}$

$$\begin{aligned} P(T \geq m) &= \sum_{k=m}^{\infty} P(T = k) = \sum_{k=m}^{\infty} (1-p)^{(k-1)} p \\ &= (1-p)^{m-1} p \sum_{k=0}^{\infty} (1-p)^k = (1-p)^{m-1} p \frac{1}{p} = (1-p)^{m-1}. \end{aligned}$$

für $T \geq k-1+n \mid T \geq n$

also

$$\begin{aligned} P(T \geq k-1+n \mid T \geq n) &= \frac{P(T \geq k-1+n)}{P(T \geq n)} = \frac{(1-p)^{k-1+n-1}}{(1-p)^{n-1}} \\ &= (1-p)^{k-1} = P(T \geq k). \end{aligned}$$

Der Beweis der zweiten Aussage ist eine Übung. □

Als **Verallgemeinerung** der geometrischen Verteilung können wir für jedes $n \geq 1$ die Wartezeit

$$T_n := \min\{k \geq 1 : \sum_{l=1}^k X_l = n\} \quad \in \{n, n+1, \dots\}$$

auf den n -ten Erfolg betrachten. Für die Verteilung der Wartezeit T_n ergibt sich analog zu obigen Überlegungen

$$P(T_n = k) = \binom{k-1}{n-1} (1-p)^{k-n} p^n, \quad k = n, n+1, \dots \quad (3.4)$$

(Übung). Die hierdurch auf $\{n, n+1, \dots\}$ definierte Verteilung heißt **negative Binomialverteilung** zu den Parametern $n \in \mathbb{N}$ und $p \in]0, 1]$.

Poissonverteilung

Für $\lambda > 0$ definiert

$$\pi_\lambda(k) := e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}_0$$

Normierung

eine Zähldichte auf \mathbb{N}_0 , denn aus der Reihenentwicklung der Exponentialfunktion folgt

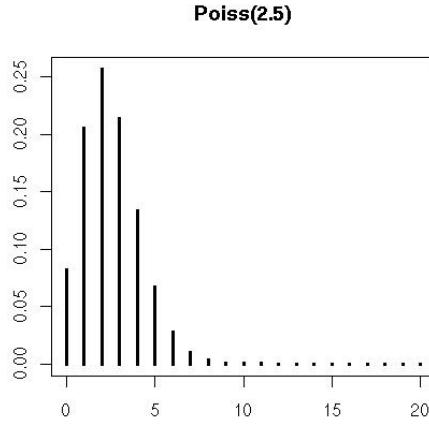
$$\sum_{k=0}^{\infty} \pi_\lambda(k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^\lambda = e^0 = 1.$$

Definition 3.14. Es sei $\lambda > 0$. Das durch die Zähldichte

$$\pi_\lambda : \mathbb{N}_0 \rightarrow [0, 1]$$

$$k \mapsto e^{-\lambda} \frac{\lambda^k}{k!}$$

definierte Wahrscheinlichkeitsmaß auf \mathbb{N}_0 heißt **Poissonverteilung zu λ** und wird mit $\text{Poiss}(\lambda)$ bezeichnet.



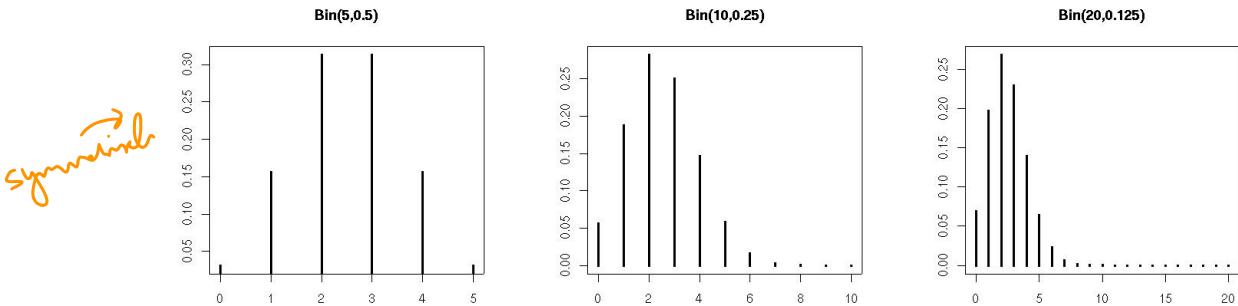
Die Poissonverteilung empfiehlt sich als Näherung der Binomialverteilung $\text{Bin}(n, p)$ für große n und kleine p . Die Approximation ist umso besser, je kleiner der Wert np^2 ist. Diese Näherung wird gerechtfertigt durch die folgende Beobachtung:

Satz 3.15. Poissonscher Grenzwertsatz

Es sei $(p_n) \subset [0, 1]$ eine Folge von Erfolgsparametern mit $\lim_{n \rightarrow \infty} np_n = \lambda > 0$. Dann folgt

$$\lim_{n \rightarrow \infty} b(k; n, p_n) = \pi_\lambda(k) \quad \text{für alle } k \in \mathbb{N}_0.$$

Mit anderen Worten: Die Wahrscheinlichkeitsfunktion der Binomialverteilung $\text{Bin}(n, p_n)$ konvergiert punktweise gegen die Wahrscheinlichkeitsfunktion der Poissonverteilung $\text{Poiss}(\lambda)$. Im folgenden eine Illustration dieser Konvergenz für $\lambda = 2.5$.

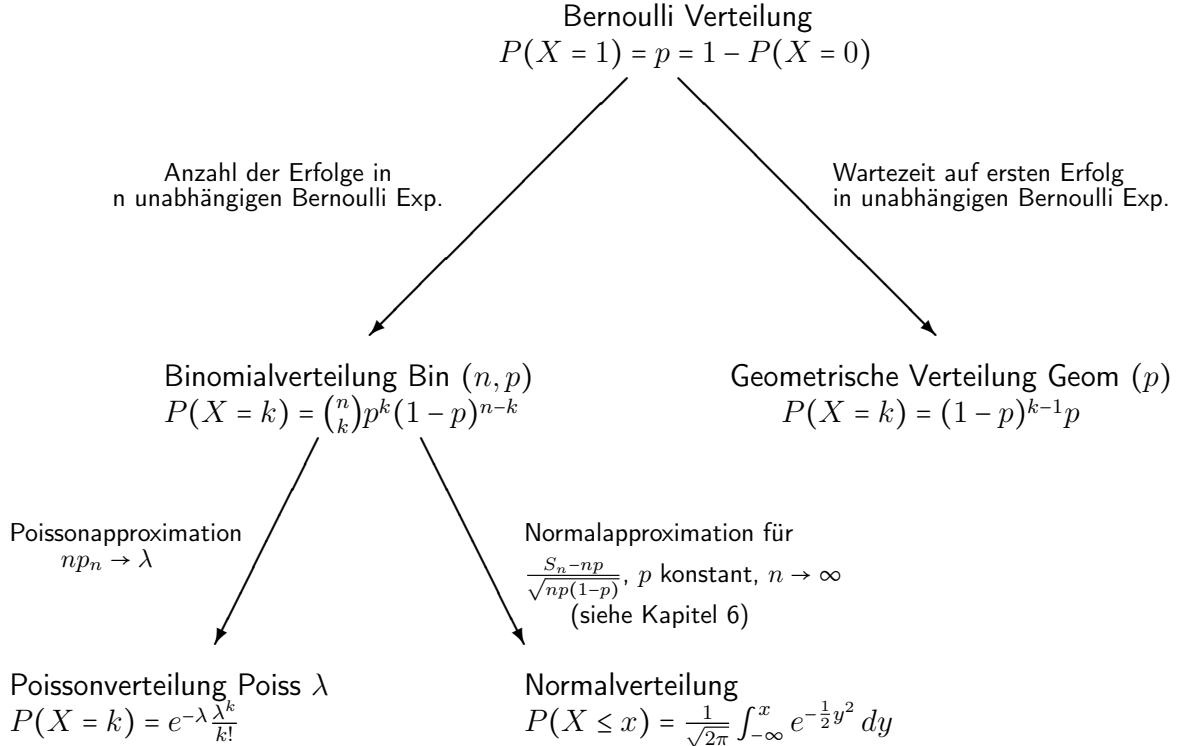


Beweis. Zum Beweis des Poissonschen Grenzwertsatzes beachte man, dass unter der Annahme $\lim_{n \rightarrow \infty} np_n = \lambda$ folgt

$$\begin{aligned}
\lim_{n \rightarrow \infty} b(k; n, p_n) &= \lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\
&= \lim_{n \rightarrow \infty} \frac{1}{k!} \underbrace{\frac{n}{k}}_{\rightarrow 1} \underbrace{\frac{(n-1)}{n}}_{\rightarrow 1} \dots \underbrace{\frac{(n-k+1)}{n}}_{\rightarrow 1} \underbrace{(np_n)^k}_{\rightarrow \lambda^k} \underbrace{\left(1 - \frac{np_n}{n}\right)^{n-k}}_{\sim (1 - \frac{\lambda}{n})^n \rightarrow e^{-\lambda}} \\
&= \frac{1}{k!} \lambda^k e^{-\lambda} = \pi_\lambda(k).
\end{aligned}$$

□

Eine näherungsweise Berechnung von Wahrscheinlichkeiten gewisser Ereignisse mit Hilfe einer Poissonverteilung ist immer dann gerechtfertigt, wenn es sich um seltene Ereignisse handelt.



Hypergeometrische Verteilung

Es sei eine Grundgesamtheit mit N Elementen gegeben, von denen K Elemente die Eigenschaft E besitzen. Aus dieser Grundgesamtheit werde n -mal ohne Zurücklegen gezogen. Wir sind interessiert an der Anzahl k der gezogenen Elemente, die die Eigenschaft E besitzen. Hierzu definieren wir

$$X = \text{Anzahl der gezogenen Elemente mit Eigenschaft } E.$$

Beispiel Hochrechnungen

Ein See enthalte eine (unbekannte) Anzahl N von Fischen. Um N zu schätzen, markiere man zunächst K Fische mit rot. Danach ziehe man n ($n \leq N$) Fische aus dem See. Dann ist X die Anzahl der markierten Fische aus dieser Stichprobe und

$$\hat{N} := \frac{n}{X} K$$

ist eine natürliche Schätzung für die unbekannte Gesamtanzahl N . Zur Begründung beachte man, dass der Anteil $\frac{X}{n}$ an rot markierten Fischen in der Stichprobe dem Anteil $\frac{K}{N}$ aller rot

markierten Fische an der Gesamtpopulation entsprechen sollte, d.h.

$$\frac{X}{n} \sim \frac{K}{N} \quad \text{und damit } N \sim \frac{n}{X} K = \hat{N}.$$

Ist $\frac{n}{N}$ klein, so gibt es keinen großen Unterschied zwischen dem Ziehen ohne Zurücklegen und dem Ziehen mit Zurücklegen. Daher empfiehlt sich in diesem Falle eine Approximation der Verteilung von X durch die Binomialverteilung $\text{Bin}(n, p)$ mit $p = \frac{K}{N}$, also

$$P(X = k) \approx b(k; n, \frac{K}{N}). \quad (3.5)$$

Ist $\frac{n}{N}$ jedoch vergleichsweise groß, so muss die gesuchte Verteilung exakt berechnet werden:

$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}, \quad k = 0, \dots, n. \quad (3.6)$$

Zur Herleitung der Formel (3.6) für die gesuchte Wahrscheinlichkeit beachte man, dass $\binom{K}{k}$ (bzw. $\binom{N-K}{n-k}$) gerade die Anzahl der k (bzw. $n - k$)-elementigen Teilmengen einer K (bzw. $N - K$)-elementigen Grundmenge ist, während $\binom{N}{n}$ die Anzahl aller n -elementigen Teilmengen der Grundgesamtheit aus N Elementen ist.

Definition 3.16. Es sei $K \leq N$, $n \leq N$. Das durch die Zähldichte

$$H(\cdot; n, N, K) : \{0, \dots, n\} \rightarrow [0, 1]$$

$$k \mapsto \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

definierte Wahrscheinlichkeitsmaß auf $\{0, \dots, n\}$ heißt **Hypergeometrische Verteilung zu n, N und K** und wird mit $\text{Hyp}(n, N, K)$ bezeichnet.

Begründung von (3.5) Für $N, K_N \rightarrow \infty$ mit $p_N := \frac{K_N}{N} \rightarrow p$, $N \rightarrow \infty$, gilt (mit $K = K_N$)

$$\begin{aligned} P(X = k) &= \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} = \frac{\cancel{K!}}{\cancel{k!(K-k)!}} \cdot \frac{\cancel{(N-k)!}}{\cancel{(n-k)!(N-K-(n-k))!}} \cdot \frac{\cancel{(n)(N-n)!}}{N!} \\ &= \binom{n}{k} \frac{K!}{(K-k)!} \frac{(N-K)!}{((N-K)-(n-k))!} \frac{(N-n)!}{N!} \\ &= \binom{n}{k} \frac{K}{N} \frac{K-1}{N} \cdots \frac{K-k+1}{N} \frac{N-K}{N} \frac{N-K-1}{N} \cdots \frac{N-K-n+k+1}{N} \\ &\quad \cdot \frac{N}{N} \frac{N}{N-1} \cdots \frac{N}{N-n+1} \rightarrow \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned}$$

Hypergeometrische Verteilung Hyp (n, N, K)

$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

Binomialapproximation
 $N, K \rightarrow \infty, \frac{K}{N} \rightarrow p$

Binomialverteilung Bin (n, p)

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

10.5

3.3 Einige wichtige stetige Verteilungen

Gleichverteilung

Für $a < b$ heißt eine Zufallsvariable mit Dichte

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } x \in [a, b] \\ 0 & \text{sonst} \end{cases}$$

gleichverteilt auf $[a, b]$. Die zugehörige Verteilungsfunktion ist

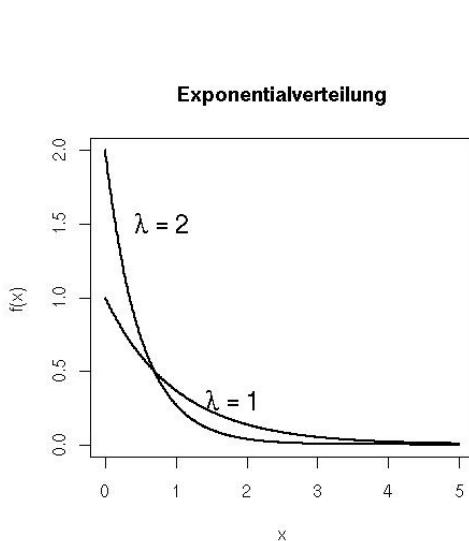
$$F(x) = \begin{cases} 0 & \text{für } x < a \\ \frac{x-a}{b-a} & \text{für } x \in [a, b] \\ 1 & \text{für } x > b. \end{cases}$$

Exponentialverteilung

Für $\lambda > 0$ ist

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

eine Dichte. Die zugehörige Verteilung heißt **Exponentialverteilung** zum Parameter λ . Sie wird mit $\text{Exp}(\lambda)$ bezeichnet.



$$\begin{aligned}
 p_h &= \int_{h-1}^h \lambda e^{-\lambda x} dx \\
 &= e^{-\lambda(h-1)} \int_0^1 \lambda e^{-\lambda x} dx \\
 &= (e^{-\lambda h})^{h-1} (1 - e^{-\lambda}) \\
 &= (1 - e^{-\lambda})^{h-1} \lambda \\
 \text{für } p &= 1 - e^{-\lambda} \\
 \Downarrow & \\
 x &= \log(1-p) \\
 \text{stetig analog umgekehrt für } p &= 1 - e^{-\lambda}
 \end{aligned}$$

Die zugehörige Verteilungsfunktion ist

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ 1 - e^{-\lambda x} & \text{für } x \geq 0. \end{cases}$$

Die Exponentialverteilung ist das stetige Analogon der geometrischen Verteilung, die ja die Verteilung von Wartezeiten auf den ersten Erfolg in einer Folge von unabhängigen Bernoulli Experimenten beschreibt. Dementsprechend verwendet man die Exponentialverteilung zur Modellierung von stetig verteilten Wartezeiten. Analog zum Fall geometrisch verteilter Zufallsvariablen (siehe Satz 3.13) ist auch eine exponentialverteilte Zufallsvariable gedächtnislos.

Satz 3.17. Es sei T exponentialverteilt. Dann gilt für $s > 0$

$$P(T > s + t \mid T > t) = P(T > s) \quad \forall t > 0. \quad (3.7)$$

Umgekehrt: Ist T eine nichtnegative Zufallsvariable mit der Eigenschaft (3.7), d.h. gedächtnislos, und gilt $P(T > 0) > 0$, so ist T exponentialverteilt.

Beweis. Es sei T exponentialverteilt mit Parameter $\lambda > 0$ und F die zugehörige Verteilungsfunktion. Dann gilt für $t > 0$

$$\underbrace{P(T \leq t)}_{P(T > t)} = 1 - F(t) = e^{-\lambda t}$$

also

$$P(T > s + t \mid T > t) = \frac{P(T > s + t)}{P(T > t)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} = P(T > s).$$

Sei umgekehrt T eine nichtnegative Zufallsvariable mit (3.7), F ihre Verteilungsfunktion und $\bar{F}(t) := 1 - F(t)$. Dann folgt für $t, s > 0$ die Funktionalgleichung

$$\begin{aligned}
 \bar{F}(t+s) &= P(T > s + t) = P(T > s + t \mid T > t)P(T > t) \\
 \stackrel{\text{bzw.}}{=} P(T > s)P(T > t) &= \bar{F}(s)\bar{F}(t).
 \end{aligned} \quad (3.8)$$

In der Analysis zeigt man, dass jede strikt positive rechtsstetige Funktion, die dieser Funktionalgleichung genügt von der Form $\bar{F}(t) = e^{-\lambda t}$ für ein $\lambda \in \mathbb{R}$ sein muss. Da außerdem $\lim_{t \rightarrow \infty} \bar{F}(t) = 0$ gilt, muss $\lambda > 0$ sein und damit $P(T \leq t) = 1 - \bar{F}(t) = 1 - e^{-\lambda t}$, d.h. T exponentialverteilt. Um einzusehen, dass \bar{F} strikt positiv ist, beachte man, dass aus $P(T > 0) > 0$ folgt, dass $\bar{F}(t) > 0$ zunächst für kleine t gelten muss und dann durch wiederholte Anwendung von (3.8) $\bar{F}(t) = \bar{F}(\frac{t}{n})^n > 0$ für alle $t > 0$. \square

Analog zum diskreten Fall können wir auch die Familie der Exponentialverteilungen einbetten in eine Familie allgemeinerer Wartezeitenverteilungen, den Gamma-Verteilungen. Eine Zufallsvariable T heißt Gamma-verteilt mit Parameter $q, \lambda > 0$, falls ihre Dichte die Form

$$f(x) = \begin{cases} \frac{\lambda^q}{\Gamma(q)} x^{q-1} e^{-\lambda x} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases} \quad (3.9)$$

hat. Sie wird mit $\Gamma_{q,\lambda}$ bezeichnet. Hierbei ist

$$\Gamma(q) = \int_0^\infty x^{q-1} e^{-x} dx, q > 0$$

die Gammafunktion. Man beachte, dass $\text{Exp}(\lambda) = \Gamma_{1,\lambda}$. Nun gilt, dass die Summe unabhängig $\text{Exp}(\lambda)$ -verteilter Wartezeiten T_1, \dots, T_n Gamma-verteilt ist mit Parameter n und λ (siehe Beispiel 3.30).

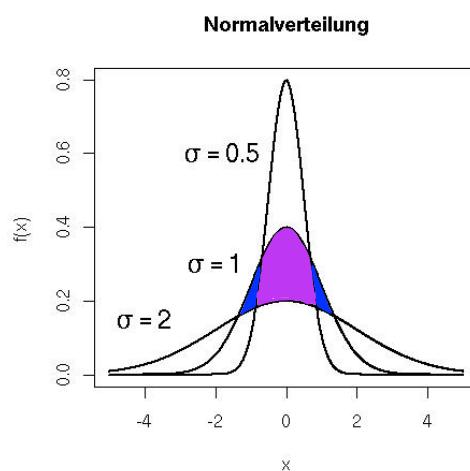
Normalverteilung

Für $m \in \mathbb{R}$ und $\sigma > 0$ ist

$$f_{m,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}}$$

Normierung
max Dichtewert
 m

eine Dichte. Die zugehörige Verteilung heißt **Normalverteilung** mit Mittel m und Varianz σ^2 . Sie wird mit $N(m, \sigma^2)$ bezeichnet. Im Falle $m = 0$ und $\sigma^2 = 1$ spricht man von der **Standardnormalverteilung**.



f_{m,σ^2} besitzt ein absolutes Maximum in $x = m$ und Wendepunkte in $m \pm \sigma$. Wegen ihrer Form wird f auch als **Gaußsche Glockenkurve** bezeichnet. σ bestimmt **Breite** und **Höhe** der Glockenkurve.

Bemerkung Ist X normalverteilt mit Mittel $m \in \mathbb{R}$ und Varianz $\sigma^2 > 0$, so ist $Y := \frac{X-m}{\sigma}$ standardnormalverteilt, denn:

$$\begin{aligned} P\left(\frac{X-m}{\sigma} \leq t\right) &= P(X \leq m + t\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{m+t\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{y^2}{2}} dy, \end{aligned}$$

wobei in der zweiten Zeile mit $y = \frac{x-m}{\sigma}$ substituiert wurde.

Allgemein: $X \sim N(m, \sigma^2)$
 \Downarrow
 $Y = aX + b \sim N(a+b, a^2\sigma^2)$
 $\Updownarrow a \neq 0, b \in \mathbb{R}$

3.4 Gemeinsame Verteilung und Unabhängigkeit von Zufallsvariablen

Definition 3.18. Es sei (Ω, \mathcal{A}, P) ein beliebiger W.raum und X_1, X_2, \dots, X_n Zufallsvariablen mit Werten in den messbaren Räumen (E_i, \mathcal{E}_i) , $i = 1, \dots, n$. Dann heißen X_1, \dots, X_n (stochastisch) unabhängig, falls für beliebige Teilmengen $B_i \in \mathcal{E}_i$ die Ereignisse

$$\{X_1 \in B_1\}, \dots, \{X_n \in B_n\}$$

(stochastisch) unabhängig sind.

Bemerkung 3.19. (i)

$$\begin{aligned} X_1, \dots, X_n \text{ unabhängig} &\iff P(X_1 \in B_1, \dots, X_n \in B_n) \\ &= P(X_1 \in B_1) \cdot \dots \cdot P(X_n \in B_n) \\ &\text{für beliebige } B_1 \in \mathcal{E}_1, \dots, B_n \in \mathcal{E}_n \end{aligned}$$

(ii) Eine beliebige Familie $(X_i)_{i \in I}$ von Zufallsvariablen heißt (stochastisch) unabhängig, falls jede endliche Teilfamilie $(X_j)_{j \in J}$, $J \subseteq I$, $|J| < \infty$, (stochastisch) unabhängig ist.

(iii) Die Unabhängigkeit von Zufallsvariablen bleibt unter Transformationen erhalten, d.h., sind f_1, \dots, f_n messbare Abbildungen zwischen messbaren Räumen (E_i, \mathcal{E}_i) und (D_i, \mathcal{D}_i) , $i = 1, \dots, n$, so sind auch die transformierten Zufallsvariablen

$$f_1(X_1), \dots, f_n(X_n)$$

unabhängig. Um dies einzusehen beachte man, dass für beliebige Ereignisse $B_i \in \mathcal{D}_i$ zunächst aufgrund der Messbarkeit der Transformationen die Urbilder $f_i^{-1}(B_i)$ in \mathcal{E}_i liegen, dies ist gerade die Definition einer messbaren Abbildung $f_i : E_i \rightarrow D_i$, und damit aufgrund der Unabhängigkeit der Zufallsvariablen X_1, \dots, X_n die Ereignisse

$$\{X_1 \in f_1^{-1}(B_1)\}, \dots, \{X_n \in f_n^{-1}(B_n)\}$$

stochastisch unabhängig. Wegen $\{X_i \in f_i^{-1}(B_i)\} = \{f_i(X_i) \in B_i\}$ folgt hieraus gerade die Unabhängigkeit der Ereignisse

$$\{f_1(X_1) \in B_1\}, \dots, \{f_n(X_n) \in B_n\}$$

und damit die Unabhängigkeit der transformierten Zufallsvariablen $f_1(X_1), \dots, f_n(X_n)$.

Wir wollen im folgenden die Unabhängigkeit von Zufallsvariablen mithilfe ihrer Verteilungen charakterisieren. Dazu benötigen wir:

Definition 3.20. Es sei (Ω, \mathcal{A}, P) ein beliebiger W.raum und X_1, X_2, \dots, X_n Zufallsvariablen mit Werten in den messbaren Räumen (E_i, \mathcal{E}_i) , $i = 1, \dots, n$. Weiter sei $E := E_1 \times \dots \times E_n$ das kartesische Produkt, $\mathcal{E} = \mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$ die Produkt σ -Algebra sowie

$$\mathbb{X} := (X_1, \dots, X_n) : \Omega \longrightarrow E, \omega \mapsto (X_1(\omega), \dots, X_n(\omega)).$$

Dann heißt die Verteilung

$$P_{\mathbb{X}}(B) = P(\mathbb{X} \in B) = P((X_1, \dots, X_n) \in B), B \in \bigotimes_{i=1}^n \mathcal{E}_i \quad (3.10)$$

auf dem Produktraum E die gemeinsame Verteilung von X_1, \dots, X_n .

Bemerkung 3.21. Für Zylindermengen $B = B_1 \times \dots \times B_n$ mit $B_i \in \mathcal{E}_i$, $i = 1, \dots, n$ gilt insbesondere

$$P_{\mathbb{X}}(B_1 \times \dots \times B_n) = P(X_1 \in B_1, \dots, X_n \in B_n).$$

Im folgenden werden wir wieder in den beiden Spezialfällen diskreter und reellwertiger Zufallsvariablen alternative Charakterisierungen der gemeinsamen Verteilung herleiten, mit denen dann ihre Unabhängigkeit einfach überprüft werden kann.

3.4.1 Gemeinsame Verteilung und Unabhängigkeit diskreter Zufallsvariablen

Es seien X_1, \dots, X_n diskrete Zufallsvariablen mit Werten in höchstens abzählbaren Mengen E_i , $i = 1, \dots, n$. Dann ist auch $\mathbb{X} := (X_1, \dots, X_n)$ mit Werten im (abzählbaren) kartesischen Produkt $E = E_1 \times \dots \times E_n$ eine diskrete Zufallsvariable und daher ihre Verteilung, also die gemeinsame Verteilung der X_1, \dots, X_n , eindeutig durch die Zähldichte

$$p_{\mathbb{X}}(e_1, \dots, e_n) = P(X_1 = e_1, \dots, X_n = e_n), (e_1, \dots, e_n) \in E$$

bestimmt.

Satz 3.22. Unter den obigen Annahmen sind die folgenden Aussagen äquivalent:

- (i) X_1, \dots, X_n sind unabhängig.
- (ii) $P(X_1 = e_1, \dots, X_n = e_n) = P(X_1 = e_1) \cdot \dots \cdot P(X_n = e_n)$ für alle $(e_1, \dots, e_n) \in E_1 \times \dots \times E_n$.
- (iii) $p_{\mathbb{X}}(e_1, \dots, e_n) = p_{X_1}(e_1) \cdot \dots \cdot p_{X_n}(e_n)$ für alle $(e_1, \dots, e_n) \in E_1 \times \dots \times E_n$.

Proof. Die Äquivalenz von (ii) und (iii) ist offensichtlich, ebenso die Implikation (i) \Rightarrow (ii) nach Definition der Unabhängigkeit. Bleibt also nur noch (ii) \Rightarrow (i) zu zeigen. Dazu seien $B_i \subseteq E_i$ beliebige Teilmengen, $i = 1, \dots, n$. Dann folgt:

$$\begin{aligned} P(X_1 \in B_1, \dots, X_n \in B_n) &= \sum_{(e_1, \dots, e_n) \in B_1 \times \dots \times B_n} P(X_1 = e_1, \dots, X_n = e_n) \\ &= \sum_{(e_1, \dots, e_n) \in B_1 \times \dots \times B_n} P(X_1 = e_1) \cdot \dots \cdot P(X_n = e_n) \\ &= \sum_{e_1 \in B_1} P(X_1 = e_1) \cdot \dots \cdot \sum_{e_n \in B_n} P(X_n = e_n) \\ &= P(X_1 \in B_1) \cdot \dots \cdot P(X_n \in B_n). \end{aligned}$$

□

Beispiel 3.23. Bernoulli Experimente

Es sei $(A_n)_n$ eine Folge unabhängiger Ereignisse mit $P(A_n) = p$ und $X_n = 1_{A_n}$ der Ausgang des n -ten Versuchs. Dann ist (X_n) eine Folge **unabhängiger** $\{0, 1\}$ -wertiger Zufallsvariablen, denn für beliebige $i_1, \dots, i_n \in \{0, 1\}$ gilt

$$\begin{aligned} P(X_1 = i_1, \dots, X_n = i_n) &= P(A_1^{(i_1)} \cap \dots \cap A_n^{(i_n)}) \\ &= P(A_1^{(i_1)}) \cdot \dots \cdot P(A_n^{(i_n)}) \\ &= P(X_1 = i_1) \cdot \dots \cdot P(X_n = i_n). \end{aligned}$$

wobei

$$A_j^{(i_j)} = \begin{cases} A_j & \text{für } i_j = 1 \\ A_j^c & \text{für } i_j = 0. \end{cases}$$

Hierbei haben wir benutzt, dass nach Satz 2.11 mit A_1, \dots, A_n auch die Ereignisse $A_1^{(i_1)}, \dots, A_n^{(i_n)}$ (stochastisch) unabhängig sind. Für die Zähldichte der gemeinsamen Verteilung der X_1, \dots, X_n gilt

$$p_{(X_1, \dots, X_n)}(i_1, \dots, i_n) = p^{\sum_{j=1}^n i_j} (1-p)^{\sum_{j=1}^n (1-i_j)}.$$

Sie stimmt also mit dem Produkt der Wahrscheinlichkeitsmaße $P_j(1) = p = 1 - P_j(\{0\})$, $j = 1, \dots, n$ zu einem einzelnen Bernoulli-Experiment überein (siehe Beispiel 2.14).

3.4.2 Gemeinsame Verteilung und Unabhängigkeit reellwertiger Zufallsvariablen

Es seien X_1, \dots, X_n reellwertige Zufallsvariablen auf einem beliebigen Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) . Dann ist der n -dimensionale Vektor $\mathbb{X} = (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$ messbar bezüglich der Borelschen σ -Algebra $\mathcal{B}(\mathbb{R}^n)$ auf \mathbb{R}^n , also eine vektorwertige Zufallsvariable. Analog zum Fall $n = 1$ können wir die Verteilung einer vektorwertigen Zufallsvariablen durch die Verteilungsfunktion

$$F_{\mathbb{X}}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n), (x_1, \dots, x_n) \in \mathbb{R}^n$$

(eindeutig!) beschreiben. Die Funktion $F_{\mathbb{X}} : \mathbb{R}^n \rightarrow \mathbb{R}$ heißt **gemeinsame Verteilungsfunktion** der Zufallsvariablen X_1, \dots, X_n . Sind X_1, \dots, X_n unabhängig, so sind für alle x_1, \dots, x_n die Ereignisse

$$\{X_1 \leq x_1\}, \dots, \{X_n \leq x_n\}$$

unabhängig und damit ergibt sich für die gemeinsame Verteilungsfunktion

$$\begin{aligned} F_{\mathbb{X}}(x_1, \dots, x_n) &= P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \cdot \dots \cdot P(X_n \leq x_n) \\ &= F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n). \end{aligned}$$

M.a.W., in diesem Falle ist die gemeinsame Verteilungsfunktion das Produkt der Verteilungsfunktionen F_{X_i} . Dies beweist die Implikation (i) \Rightarrow (ii) im folgenden Satz. Der Beweis der umgekehrten Implikation (ii) \Rightarrow (i) benötigt maßtheoretische Hilfsmittel.

Satz 3.24. Es seien X_1, \dots, X_n reellwertige Zufallsvariablen auf einem beliebigen Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) . Dann sind äquivalent:

- (i) X_1, \dots, X_n sind unabhängig.
- (ii) $F_{\mathbb{X}}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n)$ für alle $(x_1, \dots, x_n) \in \mathbb{R}^n$.

Analog zum Fall $n = 1$ heißt $F_{\mathbb{X}}$ **absolutstetig**, falls eine Dichte $f_{\mathbb{X}} : \mathbb{R}^n \rightarrow [0, \infty[$ existiert mit

$$F_{\mathbb{X}}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{\mathbb{X}}(y_1, \dots, y_n) dy_n \dots dy_1$$

für alle $(x_1, \dots, x_n) \in \mathbb{R}^n$ und für jede Borelmenge $B \in \mathcal{B}(\mathbb{R}^n)$ gilt in diesem Fall analog zu (3.2)

$$P(X \in B) = \int \dots \int_B f_{\mathbb{X}}(y_1, \dots, y_n) dy_n \dots dy_1.$$

Die Dichte $f_{\mathbb{X}}$ bezeichnet man als die **gemeinsame Dichte** der Zufallsvariablen X_1, \dots, X_n . Besitzen die Zufallsvariablen X_1, \dots, X_n eine gemeinsame Dichte, so besitzen auch alle X_i eine Dichte, die man aus der gemeinsamen Dichte auch explizit errechnen kann. Dies ist der Gegenstand des folgenden Satzes.

Satz 3.25. Es seien X_1, \dots, X_n reellwertige Zufallsvariablen mit gemeinsamer Dichte $f_{\mathbb{X}}$. Dann besitzen alle Zufallsvariablen X_i , $i = 1, \dots, n$, ebenfalls eine Dichte f_{X_i} und es gilt

$$f_{X_i}(x) = \underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty}}_{n-1\text{-mal}} f_{\mathbb{X}}(y_1, \dots, y_{i-1}, x, y_{i+1}, \dots, y_n) dy_n \dots \widehat{dy_i} \dots dy_1 \quad (3.11)$$

für $i = 1, \dots, n$. Hierbei bedeutet $\widehat{dy_i}$, dass dieser Term in der Aufzählung weggelassen ist.

Beweisskizze: Eine formale Rechnung zeigt dass

$$\begin{aligned} P(X_i \leq x) &= \lim_{x_1, \dots, \hat{x}_i, \dots, x_n \uparrow \infty} P(X_1 \leq x_1, \dots, X_i \leq x, \dots, X_n \leq x_n) \\ &= \lim_{x_1, \dots, \hat{x}_i, \dots, x_n \uparrow \infty} F_{\mathbb{X}}(x_1, \dots, x, \dots, x_n) \\ &= \lim_{x_1, \dots, \hat{x}_i, \dots, x_n \uparrow \infty} \int_{-\infty}^{x_1} \dots \int_{-\infty}^x \dots \int_{-\infty}^{x_n} f_{\mathbb{X}}(y_1, \dots, y_n) dy_n \dots dy_1 \\ &= \int_{-\infty}^x \left(\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{\mathbb{X}}(y_1, \dots, y_{i-1}, y_i, y_{i+1}, \dots, y_n) dy_n \dots \widehat{dy_i} \dots dy_1 \right) dy_i. \end{aligned}$$

Eine Rechtfertigung der obigen Rechnung findet man in der Literatur zur Maß- und Integrationstheorie. Für uns von Belang wird in dieser Vorlesung zunächst nur die Formel (3.11) für die Berechnung der Dichte f_{X_i} aus der gemeinsamen Dichte $f_{\mathbb{X}}$ sein.

Analog zu Satz 3.22 haben wir nun folgende Charakterisierung der Unabhängigkeit von Zufallsvariablen in Größen ihrer gemeinsamen Verteilung:

Satz 3.26. Es seien X_1, \dots, X_n reellwertige Zufallsvariablen.

(i) Besitzen X_1, \dots, X_n eine gemeinsame Dichte $f_{\mathbb{X}}$, so sind äquivalent:

(a) X_1, \dots, X_n unabhängig.

(b) $f_X(x_1, \dots, x_n) = f_{X_1}(x_1) \cdot \dots \cdot f_{X_n}(x_n)$ für alle $(x_1, \dots, x_n) \in \mathbb{R}^n$.

(ii) Sind X_1, \dots, X_n unabhängig und besitzen sie Dichten f_{X_1}, \dots, f_{X_n} , so besitzen sie auch eine gemeinsame Dichte und es gilt $f_{\mathbb{X}}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdot \dots \cdot f_{X_n}(x_n)$.

Beweis. (i) Analog zum Beweis von Satz 3.22 gilt: Sind X_1, \dots, X_n unabhängig, so folgt

$$\begin{aligned} F_{\mathbb{X}}(x_1, \dots, x_n) &= P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \cdot \dots \cdot P(X_n \leq x_n) \\ &= F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n) \end{aligned}$$

für alle $(x_1, \dots, x_n) \in \mathbb{R}^n$, und damit

$$\begin{aligned} \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{\mathbb{X}}(y_1, \dots, y_n) dy_n \dots dy_1 &= \int_{-\infty}^{x_1} f_{X_1}(y_1) dy_1 \dots \int_{-\infty}^{x_n} f_{X_n}(y_n) dy_n \\ &= \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{X_1}(y_1) \cdot \dots \cdot f_{X_n}(y_n) dy_n \dots dy_1 \end{aligned}$$

also $f_X(x_1, \dots, x_n) = f_{X_1}(x_1) \cdot \dots \cdot f_{X_n}(x_n)$ für alle $(x_1, \dots, x_n) \in \mathbb{R}^n$. Umgekehrt folgt hieraus mithilfe derselben Rechnung, dass

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \cdot \dots \cdot P(X_n \leq x_n)$$

für alle $(x_1, \dots, x_n) \in \mathbb{R}^n$. Hieraus folgt aber für alle Borelmengen $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$, dass

$$P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \cdot \dots \cdot P(X_n \in B_n)$$

und damit die Unabhängigkeit der Zufallsvariablen X_1, \dots, X_n . □

3.4.3 Die Verteilung der Summe unabhängiger Zufallsvariablen

Es seien zunächst X und Y unabhängige ganzzahlige Zufallsvariablen mit zugehörigen Zähldichten p_X und p_Y . Dann gilt für die Zähldichte p_Z der Summe $Z = X + Y$:

$$\begin{aligned} p_Z(n) &= P(Z = n) = P(X + Y = n) = \sum_{k \in \mathbb{Z}} P(X = k, Y = n - k) \\ &= \sum_{k \in \mathbb{Z}} P(X = k) P(Y = n - k) \\ &= \sum_{k \in \mathbb{Z}} p_X(k) p_Y(n - k) =: p_X * p_Y(n), n \in \mathbb{Z}. \end{aligned}$$

Dies motiviert folgende

Definition 3.27. Es seien p_1 und p_2 Zähldichten auf \mathbb{Z} . Dann ist auch

$$p_1 * p_2(n) := \sum_{k \in \mathbb{Z}} p_1(k) p_2(n - k), n \in \mathbb{Z}$$

eine Zähldichte auf \mathbb{Z} . $p_1 * p_2$ heißt **Faltung der Zähldichten** p_1 und p_2 .

Wir haben also insbesondere gezeigt: die Zähldichte p_Z der Summe unabhängig verteilter \mathbb{Z} -wertiger Zufallsvariablen X und Y ist gegeben durch die Faltung $p_X * p_Y$ der Zähldichten p_X und p_Y ihrer Summanden.

Beispiel 3.28. (i) Sind X und Y unabhängig Poisson-verteilt mit Parametern λ bzw. μ , so ist deren Summe $Z = X + Y$ wieder Poisson-verteilt mit Parameter $\lambda + \mu$. In der Tat gilt für die Faltung

$$\begin{aligned} p_{X+Y}(k) &= \sum_{l=0}^k \pi_\lambda(l) \pi_\mu(k-l) = \sum_{l=0}^k e^{-\lambda} \frac{\lambda^l}{l!} e^{-\mu} \frac{\mu^{k-l}}{(k-l)!} \\ &= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{l=0}^k \binom{k}{l} \lambda^l \mu^{k-l} = e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^k}{k!} \end{aligned}$$

für $k = 0, 1, 2, \dots$

(ii) Sind X und Y unabhängig binomialverteilt mit Parameter (n, p) bzw. (m, p) , so ist deren Summe $Z = X + Y$ wieder binomialverteilt mit Parameter $(n+m, p)$. Dies ergibt sich allein schon aus der Tatsache, dass wir die Verteilung von X (bzw. von Y) als Verteilung der Summe von unabhängig Bernoulli-verteilten Zufallsvariablen X_1, \dots, X_n (bzw. Y_1, \dots, Y_m) mit Erfolgsparameter p auf einem gemeinsamen Wahrscheinlichkeitsraum realisieren können. Dann aber ist die Verteilung der Summe $X + Y$ gerade die Verteilung der Summe

$$X_1 + \dots + X_n + Y_1 + \dots + Y_m$$

von $n+m$ unabhängig Bernoulli-verteilten Zufallsvariablen mit Erfolgsparameter p , und damit binomialverteilt mit Parameter $(n+m, p)$.

Man kann dies aber auch alternativ über die Faltung nachrechnen:

$$\begin{aligned} p_{X+Y}(k) &= \sum_{l=0}^k b(l; n, p) b(k-l; m, p) = \sum_{l=0}^k \binom{n}{l} p^l (1-p)^{n-l} \binom{m}{k-l} p^{k-l} (1-p)^{m-(k-l)} \\ &= \left(\sum_{l=0}^k \binom{n}{l} \binom{m}{k-l} \right) p^k (1-p)^{n+m-k} = \binom{n+m}{k} p^k (1-p)^{n+m-k} \end{aligned}$$

für $k = 0, \dots, n+m$.

Im folgenden seien X und Y nun unabhängige reellwertige Zufallsvariablen mit Dichten f_X bzw. f_Y . Es folgt aus Satz 3.26, dass $f_{(X,Y)} = f_X \otimes f_Y$ die gemeinsame Dichte ist und damit ergibt sich für die Verteilungsfunktion F_Z der Summe $Z = X + Y$

$$\begin{aligned} F_Z(z) &= P(X + Y \leq z) = \int \int_{\{(x,y) | x+y \leq z\}} f_X(x)f_Y(y) dy dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_X(x)f_Y(y) dy dx = \int_{-\infty}^{\infty} \int_{-\infty}^z f_X(x)f_Y(y-x) dy dx \\ &= \int_{-\infty}^z \int_{-\infty}^{\infty} f_X(x)f_Y(y-x) dx dy =: \int_{-\infty}^z f_X * f_Y(y) dy, z \in \mathbb{R} \end{aligned}$$

wobei

$$f_X * f_Y(y) = \int_{-\infty}^{\infty} f_X(x)f_Y(y-x) dx.$$

Dies motiviert folgende

Definition 3.29. Es seien f_1 und f_2 zwei Dichten auf \mathbb{R} . Dann ist auch

$$f_1 * f_2(x) := \int_{-\infty}^{+\infty} f_1(y)f_2(x-y) dy, x \in \mathbb{R}$$

eine Dichte auf \mathbb{R} . $f_1 * f_2$ heißt **Faltung der Dichten** f_1 und f_2 .

Wir haben also insbesondere gezeigt: die Summe $Z = X + Y$ unabhängig verteilter \mathbb{R} -wertiger Zufallsvariablen X und Y mit Dichten f_X und f_Y , besitzt wieder eine Dichte f_Z und diese ist gegeben durch die Faltung $f_X * f_Y$ der Dichten f_X und f_Y ihrer Summanden.

Beispiel 3.30. (i) Sind X und Y unabhängig exponentialverteilt mit Parametern λ bzw. μ , so besitzt deren Summe $Z = X + Y$ die Dichte

$$\begin{aligned} f_Z(z) &= f_X * f_Y(z) = \int_{-\infty}^z f_X(y)f_Y(z-y) dy = \lambda\mu e^{-\mu z} \int_0^z e^{-(\lambda-\mu)y} dy \\ &= \begin{cases} \lambda\mu \frac{e^{-\lambda z} - e^{-\mu z}}{\mu - \lambda} & \text{für } \lambda \neq \mu \\ \lambda^2 z e^{-\lambda z} & \text{für } \lambda = \mu. \end{cases} \end{aligned}$$

für $z \geq 0$ und $f_Z(z) = 0$ sonst.

(ii) Sind X und Y unabhängig Gamma-verteil mit Parametern (p, λ) (bzw. (q, λ)), so ist die Summe $Z = X + Y$ wieder Gamma-verteil mit Parametern $(p+q, \lambda)$. Um dies einzusehen beachte, dass aus

$$f_X(x) = \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x} 1_{\{x>0\}} \quad (\text{bzw. } f_Y(x) = \frac{\lambda^q}{\Gamma(q)} x^{q-1} e^{-\lambda x} 1_{\{x>0\}})$$

folgt

$$\begin{aligned} f_Z(x) &= f_X * f_Y(z) = \frac{\lambda^{p+q}}{\Gamma(p)\Gamma(q)} \int_0^z x^{p-1} e^{-\lambda x} (z-x)^{q-1} e^{-\lambda(z-x)} dx \\ &= \frac{\lambda^{p+q}}{\Gamma(p)\Gamma(q)} \int_0^z x^{p-1} (z-x)^{q-1} dx e^{-\lambda z} \\ &= \frac{\lambda^{p+q}}{\Gamma(p+q)} z^{p+q-1} e^{-\lambda z} \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^1 u^{p-1} (1-u)^{q-1} du \\ &= \frac{\lambda^{p+q}}{\Gamma(p+q)} z^{p+q-1} e^{-\lambda z}. \end{aligned}$$

In der dritten Zeile wurde dabei die Substitution $u = \frac{x}{z}$ benutzt und in der letzten Zeile die Tatsache, dass

$$\int_0^1 u^{p-1} (1-u)^{q-1} du = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$

3.5 Erwartungswert und Varianz reellwertiger Zufallsvariablen

Erwartungswert und Varianz sind die beiden wichtigsten Kennzahlen einer Zufallsvariablen. Im ganzen Abschnitt sei (Ω, \mathcal{A}, P) ein beliebiger Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine reellwertige Zufallsvariable.

3.5.1 Erwartungswert diskreter Zufallsvariablen

Definition 3.31. (Erwartungswert für diskrete Zufallsvariablen) Es sei X diskret mit reellen Werten $x_n, n \in \mathbb{N}$.

(i) Ist $X \geq 0$, d.h. also $x_n \geq 0$ für alle n , so ist der **Erwartungswert** $E(X)$ von X definiert als der Mittelwert

$$E(X) := \sum_n x_n P(X = x_n) \in [0, \infty]. \quad (3.12)$$

Insbesondere ist der Wert $E(X) = +\infty$ zugelassen.

(ii) Im allgemeinen Fall sagen wir, dass der **Erwartungswert von X existiert**, falls $E(|X|) < \infty$. In diesem Falle ist die (möglicherweise) unendliche Reihe (3.12) absolut konvergent und somit insbesondere der Erwartungswert $E(X)$ von X durch (3.12) wohldefiniert.

Der Erwartungswert $E(X)$ ist also nichts weiter als das mit den Wahrscheinlichkeiten $P(X = x_n)$ gewichtete Mittel der Funktionswerte x_n . Er lässt sich daher als Schwerpunkt der Verteilung von X interpretieren. Man beachte, dass der Erwartungswert $E(X)$ nur von der **Verteilung** der Zufallsvariablen X abhängt und nicht von der konkreten Funktionsvorschrift $\omega \mapsto X(\omega)$.

Beispiel 3.32. (i) Münzwurf: $P(X = 0) = P(X = 1) = \frac{1}{2}$. Dann folgt

$$E(X) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = \frac{1}{2}.$$

(ii) Würfeln: $P(X = i) = \frac{1}{6}$, $i = 1, \dots, 6$. Dann folgt

$$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

Der Erwartungswert stimmt also in diesem Falle mit dem arithmetischen Mittel der Funktionswerte überein.

(iii) Indikatorfunktion $X = 1_A$ für $A \in \mathcal{A}$: $P(1_A = 1) = P(A)$ und $P(1_A = 0) = P(A^c) = 1 - P(A)$. Damit gilt

$$E(1_A) = 0 \cdot P(1_A = 0) + 1 \cdot P(1_A = 1) = 0 \cdot P(A^c) + 1 \cdot P(A) = P(A).$$

Bemerkung 3.33. Rechenregeln für Erwartungswerte

Es seien X, Y beliebige (diskrete) Zufallsvariablen, deren Erwartungswerte existieren. Dann gilt:

(i) **Linearität** $E(aX + bY) = aE(X) + bE(Y)$ für alle $a, b \in \mathbb{R}$.

(ii) **Nichtnegativität** $X \geq 0$ (d.h. $X(\omega) \geq 0$ für alle $\omega \in \Omega$)

$$\implies E(X) \geq 0.$$

(iii) **Monotonie** $X \leq Y$ (d.h. $Y - X \geq 0$)

$$\implies E(X) \leq E(Y).$$

(iv) **Dreiecksungleichung**

$$|E(X)| \leq E(|X|).$$

(v) **Transformationssatz** Es sei $h : \mathbb{R} \rightarrow \mathbb{R}$ eine beliebige Funktion. Sind x_n die Werte von X und y_k die Werte von $h(X)$, so gilt:

(1) Ist $h \geq 0$, also $y_k \geq 0$, so ist

$$\begin{aligned} E(h(X)) &= \sum_k y_k P(h(X) = y_k) = \sum_k y_k \sum_{n:h(x_n)=y_k} P(X = x_n) \\ &= \sum_n h(x_n) P(X = x_n). \end{aligned} \tag{3.13}$$

(2) Ist $E(|h(X)|) < \infty$, so konvergiert die Reihe $\sum_n h(x_n)P(X = x_n)$ absolut und für den Erwartungswert $E(h(X))$ gilt ebenfalls (3.13).

Satz 3.34. Es seien X_1, \dots, X_n unabhängige (diskrete) Zufallsvariablen, deren Erwartungswerte existieren. Dann gilt: der Erwartungswert von $X_1 \cdot \dots \cdot X_n$ existiert und

$$E(X_1 \cdot \dots \cdot X_n) = E(X_1) \cdot E(X_2) \cdot \dots \cdot E(X_n).$$

Beweis. Für $k = 1, \dots, n$ seien x_l^k , $l = 1, 2, \dots$, die Werte der Zufallsvariablen X_k . Dann gilt

$$\begin{aligned} &\sum_{l_1} \dots \sum_{l_n} |x_{l_1}^1 \cdot \dots \cdot x_{l_n}^n| P(X_1 = x_{l_1}^1, \dots, X_n = x_{l_n}^n) \\ &= \sum_{l_1} \dots \sum_{l_n} |x_{l_1}^1 \cdot \dots \cdot x_{l_n}^n| \prod_{k=1}^n P(X_k = x_{l_k}^k) = \sum_{l_1} \dots \sum_{l_n} \prod_{k=1}^n |x_{l_k}^k| P(X_k = x_{l_k}^k) \\ &= \sum_{l_1} |x_{l_1}^1| P(X_1 = x_{l_1}^1) \dots \sum_{l_n} |x_{l_n}^n| P(X_n = x_{l_n}^n) = E(|X_1|) \cdot \dots \cdot E(|X_n|) < \infty, \end{aligned}$$

also existiert der Erwartungswert $E(X_1 \cdot \dots \cdot X_n)$ und Wiederholung derselben Rechnung ohne $|\cdot|$ liefert

$$E(X_1 \cdot \dots \cdot X_n) = \prod_{k=1}^n E(X_k).$$

□

Beispiel 3.35. (i) Sind X_1, \dots, X_n unabhängig Bernoulli-verteilt mit Erfolgswahrscheinlichkeit p , so folgt

$$E(X_k) = 0 \cdot P(X_k = 0) + 1 \cdot P(X_k = 1) = p.$$

Insbesondere gilt für den Erwartungswert der Summe

$$S_n = X_1 + \dots + X_n$$

$$E(S_n) = E(X_1) + \dots + E(X_n) = p + \dots + p = np.$$

Da S_n binomialverteilt ist mit Parameter n und p , folgt insbesondere: Für den Erwartungswert einer binomialverteilten Zufallsvariablen S_n mit Parametern n und p gilt:

$$E(S_n) = np.$$

Die Anwendung des Transformationssatzes ergibt weiterhin, dass für $\alpha \in \mathbb{R}$

$$\begin{aligned} E(e^{\alpha X_1}) &= e^{\alpha 0} P(X_1 = 0) + e^{\alpha 1} P(X_1 = 1) \\ &= e^{\alpha 0}(1 - p) + e^{\alpha 1} p = (1 - p) + pe^\alpha, \end{aligned}$$

also

$$E(e^{\alpha X_i}) = (1 - p) + pe^\alpha \quad \text{für } i = 1, \dots, n,$$

und damit folgt

$$\begin{aligned} E(e^{\alpha S_n}) &= E(e^{\alpha \sum_{i=1}^n X_i}) = E(e^{\alpha X_1} e^{\alpha X_2} \dots e^{\alpha X_n}) \\ &= E(e^{\alpha X_1}) E(e^{\alpha X_2}) \dots E(e^{\alpha X_n}) = (1 - p + pe^\alpha)^n. \end{aligned}$$

(ii) Ist X Poiss(λ)-verteilt, so folgt

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k P(X = k) = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} \\ &= \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda \lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda. \end{aligned}$$

Weiterhin folgt mit dem Transformationssatz

$$\begin{aligned} E(e^{\alpha X}) &= \sum_{k=0}^{\infty} e^{\alpha k} P(X = k) = \sum_{k=0}^{\infty} e^{\alpha k} e^{-\lambda} \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^\alpha \lambda)^k}{k!} = e^{-\lambda} e^{e^\alpha \lambda} = e^{-\lambda(1-e^\alpha)}. \end{aligned}$$

3.5.2 Erwartungswert allgemeiner reellwertiger Zufallsvariablen

Für eine beliebige reellwertige Zufallsvariable X definieren wir den Erwartungswert durch

$$E(X) := \int_{\Omega} X dP \quad \text{falls} \quad \int_{\Omega} |X| dP < \infty.$$

Das Integral wird dabei in der Maß- und Integrationstheorie wie folgt über drei Schritte definiert, die man auch als **maßtheoretische Induktion** bezeichnet:

(i) Ist $X \geq 0$ eine diskrete Zufallsvariable, so lässt sich X darstellen in der Form

$$X = \sum_n x_n 1_{A_n}, A_n := \{X = x_n\}$$

Dann definiert man (vgl. mit Definition 3.31)

$$\int X dP := \sum_n x_n P(A_n) = \sum_n x_n P(X = x_n).$$

(ii) Ist $X \geq 0$ beliebig, so kann man X durch eine monoton wachsende Folge von nichtnegativen diskreten Zufallsvariablen X_n approximieren, die punktweise gegen X konvergiert, d.h., $\lim_{n \rightarrow \infty} X_n(\omega) \uparrow X(\omega)$ für alle $\omega \in \Omega$. Zum Beispiel

$$X_n = \sum_{k=0}^{n2^n-1} \frac{k}{2^n} 1_{\{\frac{k}{2^n} \leq X < \frac{k+1}{2^n}\}} + n 1_{\{n \leq X\}} \uparrow X.$$

Dann ist $\int X_n dP$ aufgrund der Monotonie des Integrals (Erwartungswertes) eine monoton wachsende Folge und damit (uneigentlich) konvergent. Nun definiert man

$$\int X dP := \lim_{n \rightarrow \infty} \int X_n dP.$$

Die Wohldefiniertheit von $\int X dP$ zeigt man in der Maß- und Integrationstheorie dadurch, dass der soeben definierte Grenzwert **unabhängig** von der Wahl der approximierenden Folge (X_n) ist! Wir können also den Erwartungswert einer beliebigen nichtnegativen Zufallsvariablen über den Grenzwert der Erwartungswerte approximierender Folgen diskreter Zufallsvariablen ausrechnen.

(iii) Ist X eine beliebige Zufallsvariable, so sind Positivteil $X^+ := \max\{X, 0\}$ und Negativteil $X^- := \max\{-X, 0\}$ wieder nichtnegative Zufallsvariablen (insbesondere messbar nach Bemerkung 3.2). Beachte dass $X = X^+ - X^-$. Ist nun $\int |X| dP < \infty$, so ist auch $\int X^+ dP < \infty$ und $\int X^- dP < \infty$, und daher

$$\int X dP := \int X^+ dP - \int X^- dP$$

wohldefiniert.

Die Eigenschaften des Erwartungswertes diskreter Zufallsvariablen (siehe Bemerkung 3.33), mit Ausnahme des Transformationssatzes, übertragen sich Wort für Wort auf den allgemeinen Fall. Ebenso gilt die Aussage des Satzes 3.34 zum Erwartungswert des Produktes unabhängiger Zufallsvariablen auch im allgemeinen Fall. Einen Beweis findet man in der Literatur zur Maß- und Integrationstheorie.

Für eine absolutstetig verteilte Zufallsvariable X mit (Riemann- oder Lebesgue-integrierbarer) Dichte f gilt

$$E(|X|) = \int_{-\infty}^{\infty} |x| f(x) dx.$$

Ist daher $E(|X|) < \infty$, so existiert der Erwartungswert von X und es gilt

$$E(X) := \int_{-\infty}^{\infty} x f(x) dx.$$

In diesem Falle können wir folgende Version des Transformationssatzes angeben: Ist $h : \mathbb{R} \rightarrow \mathbb{R}$ Borel-messbar (z.B. stückweise stetig), so gilt

$$E(h(X)) = \int_{-\infty}^{\infty} h(x) f(x) dx \quad \text{falls} \quad \int_{-\infty}^{\infty} |h(x)| f(x) dx < \infty.$$

Beispiel 3.36. (i) X gleichverteilt auf $[a, b]$. Dann existiert der Erwartungswert von X und es gilt

$$E(X) = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \frac{1}{2} x^2 \Big|_a^b = \frac{1}{2}(a+b).$$

Der Erwartungswert einer auf $[a, b]$ gleichverteilten ZV ist somit gleich dem Mittelwert der Randpunkte a und b .

(ii) X exponentialverteilt mit Parameter $\lambda > 0$. Dann existiert der Erwartungswert von X und es gilt

$$E(X) = \int_0^\infty x \lambda e^{-\lambda x} dx = -x e^{-\lambda x} \Big|_0^\infty + \int_0^\infty e^{-\lambda x} dx = -\frac{1}{\lambda} e^{-\lambda x} \Big|_0^\infty = \frac{1}{\lambda}.$$

(iii) X standardnormalverteilt. Dann existiert der Erwartungswert und es gilt

$$E(X) = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0.$$

Ist X normalverteilt mit Mittel $m \in \mathbb{R}$ und Varianz $\sigma^2 > 0$, so ist $Y := \frac{X-m}{\sigma}$ standardnormalverteilt und daher

$$E(X) = \sigma E\left(\frac{X-m}{\sigma}\right) + E(m) = \sigma \cdot 0 + m = m.$$

3.5.3 Die Varianz einer reellwertigen Zufallsvariablen

Für eine reellwertige Zufallsvariable X und $p = 1, 2, 3, \dots$, heißt $E(X^p)$ das p -te **Moment** von X . X heißt **quadratisch integrierbar**, falls X endliches zweites Moment besitzt, d.h. also $E(X^2) < \infty$.

Satz 3.37. Cauchy-Schwarzsche Ungleichung Es seien X und Y reellwertige Zufallsvariablen mit $E(X^2) < \infty$ und $E(Y^2) < \infty$. Dann ist auch $E(|XY|) < \infty$ und es gilt

$$|E(XY)| \leq E(X^2)^{\frac{1}{2}} E(Y^2)^{\frac{1}{2}}.$$

Für $Y \equiv 1$ ergibt sich hieraus insbesondere $|E(X)| \leq E(X^2)^{\frac{1}{2}}$.

Beweis. Aus der elementaren Ungleichung $|XY| \leq \frac{1}{2}X^2 + \frac{1}{2}Y^2$ folgt zunächst $E(|XY|) < \infty$ und damit existiert der Erwartungswert von XY . Sind $E(X^2) = E(Y^2) = 0$ so folgt aus der elementaren Ungleichung ebenfalls $E(|XY|) = 0$ und mithilfe der Dreiecksungleichung somit auch $|E(XY)| \leq E(|XY|) = 0$. Dies beweist die Ungleichung in diesem Spezialfall. Ist nun $E(X^2) > 0$ oder $E(Y^2) > 0$, so können wir oBdA annehmen, dass $E(Y^2) > 0$. Mit $\alpha := \frac{E(XY)}{E(Y^2)}$ ergibt sich dann

$$0 \leq E((X - \alpha Y)^2) = E(X^2) - 2\alpha E(XY) + \alpha^2 E(Y^2) = E(X^2) - \frac{E(XY)^2}{E(Y^2)},$$

äquivalent hierzu

$$E(XY)^2 \leq E(X^2)E(Y^2)$$

woraus sich dann die behauptete Ungleichung auch in diesem Fall ergibt. \square

Definition 3.38. (i) Es sei X eine reellwertige Zufallsvariable mit endlichem Erwartungswert $|E(X)| < \infty$. Dann heißt die mittlere quadratische Abweichung von X um ihren Erwartungswert $E(X)$

$$V(X) := E((X - E(X))^2) \in [0, \infty]$$

die **Varianz** der Zufallsvariablen X . Aus Satz 3.37 folgt insbesondere, dass $V(X) < \infty \iff E(X^2) < \infty$.

Die Wurzel $S(X) := \sqrt{V(X)}$ aus der Varianz heißt **Standardabweichung** von X .

(ii) Es seien X und Y Zufallsvariablen mit $E(X^2) < \infty$ und $E(Y^2) < \infty$. Dann heißt der Erwartungswert

$$\text{Cov}(X, Y) := E((X - E(X))(Y - E(Y)))$$

die **Kovarianz** von X und Y . Gilt $\text{Cov}(X, Y) = 0$, so heißen X und Y **unkorreliert**.

Satz 3.39. Rechenregeln für Varianzen

Es seien X, Y, X_1, \dots, X_n quadratisch integrierbare Zufallsvariablen. Dann gilt:

(i) $V(aX + b) = a^2 V(X)$ für alle $a, b \in \mathbb{R}$ und $\text{Cov}((aX + b), (cY + d)) = ac \text{Cov}(X, Y)$ für alle $a, b, c, d \in \mathbb{R}$.

(ii) Die Kovarianz ist eine positiv semidefinite symmetrische Bilinearform, d.h.,

$$\text{Cov}(X, X) = V(X) \geq 0, \quad \text{Cov}(X, Y) = \text{Cov}(Y, X),$$

$$\text{Cov}(X, aX_1 + bX_2) = a \text{Cov}(X, X_1) + b \text{Cov}(X, X_2)$$

für alle $a, b \in \mathbb{R}$.

(iii) **Verschiebungssatz** $V(X) = E(X^2) - (E(X))^2$ und $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$.

(iv) Sind X, Y unkorreliert, so folgt $V(X + Y) = V(X) + V(Y)$.

Allgemeiner gilt die **Identität von Bienaym **

Sind X_1, \dots, X_n paarweise unkorreliert, so folgt

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n).$$

(v) X und Y unabhängig $\Rightarrow X, Y$ unkorreliert, d.h. $\text{cov}(X, Y) = 0$. Die Umkehrung gilt i.A. nicht.

Beweis. (i) Aus $E(aX + b) = aE(X) + b$ folgt

$$V(aX + b) = E((aX + b - E(aX + b))^2) = E((aX - aE(X))^2) = a^2 V(X).$$

Analog zeigt man

$$\begin{aligned} \text{Cov}(aX + b, cY + d) &= E((aX + b - E(aX + b))(cY + d - E(cY + d))) \\ &= E(a(X - E(X))c(Y - E(Y))) = ac \text{Cov}(X, Y). \end{aligned}$$

(ii) Positiv Semidefinitheit und Symmetrie sind offensichtlich. Die Bilinearität folgt dann aus

$$\begin{aligned}\text{Cov}(X, aX_1 + bX_2) &= E((X - E(X))(aX_1 + bX_2 - E(aX_1 + bX_2))) \\ &= E((X - E(X))(a(X_1 - E(X_1)) + b(X_2 - E(X_2)))) \\ &= aE((X - E(X))(X_1 - E(X_1))) + bE((X - E(X))(X_2 - E(X_2))) \\ &= a \text{Cov}(X, X_1) + b \text{Cov}(X, X_2).\end{aligned}$$

(iii) Es gilt

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

und analog

$$\begin{aligned}\text{Cov}(X, Y) &= E((X - E(X))(Y - E(Y))) \\ &= E(XY - XE(Y) - E(X)Y + E(X)E(Y)) \\ &= E(XY) - 2E(X)E(Y) + E(X)E(Y) = E(XY) - E(X)E(Y).\end{aligned}$$

(iv) Sind X und Y unkorreliert, so folgt

$$\begin{aligned}\text{V}(X + Y) &= \text{Cov}(X + Y, X + Y) = \text{Cov}(X, X) + 2 \text{Cov}(X, Y) + \text{Cov}(Y, Y) \\ &= \text{V}(X) + \text{V}(Y).\end{aligned}$$

Der Beweis der Identität von Bienaym  schlie lich ist eine leichte Verallgemeinerung.

(v) Sind X und Y unabh ngig, so folgt $E(XY) = E(X)E(Y)$ (Satz 3.34) und damit aufgrund des Verschiebungssatzes

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0.$$

Um einzusehen, dass die Umkehrung i.A. falsch ist betrachte man folgendes **Beispiel**

Es seien $A_1, A_2 \subseteq A_3$ Ereignisse mit $A_1 \cap A_2 = \emptyset$ und $0 < P(A_1) = P(A_2) < P(A_3) < 1$. Dann sind die Zufallsvariablen $X := 1_{A_1} - 1_{A_2}$ und $Y := 1_{A_3}$ unkorreliert, denn

$$XY = 1_{A_1 \cap A_3} - 1_{A_2 \cap A_3} = 1_{A_1} - 1_{A_2} = X$$

sowie $E(X) = P(A_1) - P(A_2) = 0$ und damit

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(X) - E(X)E(Y) = 0,$$

aber nicht unabh ngig, denn

$$P(X = 1, Y = 1) = P(A_1 \cap A_3) = P(A_1) \neq P(A_1)P(A_3) = P(X = 1)P(Y = 1).$$

□

Beispiele

- (i) Sind X_1, \dots, X_n unabhängig Bernoulli-verteilt mit Erfolgswahrscheinlichkeit p , so folgt für die Varianz der Summe $S_n = X_1 + \dots + X_n$

$$\text{V}(S_n) = \text{V}(X_1 + \dots + X_n) = \text{V}(X_1) + \dots + \text{V}(X_n).$$

Für die Varianz der Bernoulli-verteilten Zufallsvariablen X_k errechnet man sofort

$$\text{V}(X_k) = E(X_k^2) - (E(X_k))^2 = p - p^2 = p(1-p),$$

so dass

$$\text{V}(S_n) = np(1-p).$$

Da S_n binomialverteilt ist mit Parameter n und p , folgt insbesondere: Für die Varianz einer binomialverteilten Zufallsvariablen S_n mit Parameter n und p gilt

$$\text{V}(S_n) = np(1-p).$$

- (ii) Ist X Poiss(λ)-verteilt, so folgt

$$\begin{aligned} E(X^2) &= \sum_{k=0}^{\infty} k^2 P(X=k) = \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} \\ &= \sum_{k=0}^{\infty} (k-1+1) e^{-\lambda} \frac{\lambda \cdot \lambda^{k-1}}{(k-1)!} \\ &= \lambda \sum_{k=1}^{\infty} (k-1) e^{-\lambda} \frac{\lambda^{k-1}}{(k-1)!} + \lambda \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} + \lambda = \lambda^2 + \lambda, \end{aligned}$$

also

$$\text{V}(X) = E(X^2) - (E(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Varianz und Erwartungswert einer Poisson-verteilten Zufallsvariablen sind also gleich!

- (iii) X exponentialverteilt mit Parameter $\lambda > 0$. Dann gilt $E(X) = \frac{1}{\lambda}$ und wegen

$$\begin{aligned} E(X^2) &= \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx = -x^2 e^{-\lambda x} \Big|_0^{\infty} + 2 \int_0^{\infty} x e^{-\lambda x} dx \\ &= \frac{2}{\lambda} E(X) = \frac{2}{\lambda^2}, \end{aligned}$$

ergibt sich

$$\text{V}(X) = E(X^2) - (E(X))^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

- (iv) X standardnormalverteilt. Dann ist $E(X) = 0$ und

$$\begin{aligned} E(X^2) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \left(-\frac{d}{dx} e^{-\frac{x^2}{2}} \right) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = 1. \end{aligned}$$

Somit folgt

$$\text{V}(X) = E(X^2) - (E(X))^2 = 1.$$

Ist X normalverteilt mit Erwartungswert $m \in \mathbb{R}$ und Varianz $\sigma^2 > 0$ und damit $Y = \frac{X-m}{\sigma}$ standardnormalverteilt, so folgt

$$\text{V}(X) = \text{V}(\sigma Y + m) = \sigma^2 \text{V}(Y) = \sigma^2.$$

Insbesondere erklären sich hiermit die Bezeichnungen Mittel und Varianz für die beiden Parameter m und σ^2 einer normalverteilten ZV.