Wahrscheinlichkeitstheorie und Statistik

4. Woche

Dávid Tóth (BME SZIT)

23., 26. September 2024

Gemeinsame Verteilung diskreter Zufallsvariablen

- Es ist ein Vorteil des Begriffs der Zufallsvariablen, dass verschiedene zufällige Quantitäten zusammen (in einem Wahrscheinlichkeitsraum) behandelt werden können.
- Wir werden das gemiensame Verhalten verschiedene Zufallsvariablen beschreiben.
- Um die verschiedene Variablen zusammen behandeln zu können, schreiben wir ihre Werte in einer Sequenz, also bilden wir einen Vektor aus ihnen.

Gemeinsame Verteilung diskreter Zufallsvariablen

Definition. Sei Ω eine abzählbare (d.h.: endliche oder abzählbar unendliche) Ergebnismenge, dann ist *ein diskreter Zufallsvektor* eine Funktion $\underline{X} : \Omega \to \mathbb{R}^n$.

Bemerkungen.

- In diesem Kurs benutzen wir fast immer Zeilenvektoren.
- Die Koordinaten eines Zufallsvektors $\underline{X} = (X_1, \ldots, X_n)$ sind $\Omega \to \mathbb{R}$ Funktionen, also Zufallsvariablen.
- Umgekehrt, falls n Zufallsvariablen mit gleicher Definitionsmenge in einen Vektor geordnet werden, so erhalten wir einen Zufallsvektor $\Omega \to \mathbb{R}^n$.

Gemeinsame Verteilung diskreter Zufallsvariablen

Wie bei den Zufallsvariablen, das Verhlaten eines Zufallsvektors kann anhand der Wahrscheinlichkeiten beschrieben werden, mit den der Vektor die verischiedene Werte (=Vektoren in $\mathbb{R}^n$) annimmt.

Definition. Sei $\underline{X} = (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ ein diskreter Zufallsvektor, und für einen Vektor $\underline{t} = (t_1, \ldots, t_n) \in \operatorname{ran} \underline{X}$ sei

$$f_{\underline{X}}(t) := \mathbb{P}(\underline{X} = \underline{t}) = \mathbb{P}(X_1 = t_1, \ldots, X_n = t_n).$$

Die so erhaltene Funktion $f_X : \operatorname{ran} \underline{X} \to [0; 1]$ wird der *Wahrscheinlichkeitsfunktion des Zufallsvektors* $\underline{X}$ genannt.

Gemeinsame Verteilung diskreter Zufallsvariablen

- Wie früher, die Wahrscheinlichkeiten solcher Ereignisse, die anhand der Werten des Zufallsvektors beschrieben werden können, sind schon durch die Wahrscheinlichkeitsfunktion bestimmt.
- Wir sagen, dass die Wahrscheinlichkeitsfunktion die *gemeinsame Verteilung* der Koordinaten $X_1, \dots, X_n$ bestimmt.

Gemeinsame Verteilung diskreter Zufallsvariablen

- Die gemeinsame Verteilung von *zwei einfachen* Variablen X und Y kann in einer Tabelle angegeben werden, die die Werte

$$f_{X,Y}(s,t) = \mathbb{P}(X = s, Y = t)$$

der gemeinsamen Wahrscheinlichkeitsfunktion enthält, wo $s \in \operatorname{ran} X$ und $t \in \operatorname{ran} Y$.

- Die Spalten gehören zu einer der Variablen, die Zeilen gehören zur anderen.
- Im Schnitt einer Spalte und eriner Zeile steht der entsprechende Wert der Wahrscheinlichkeitsfunktion.

Gemeinsame Verteilung diskreter Zufallsvariablen

Beispiel. Seien $\operatorname{ran} X = \{2, 3, 5\}$, $\operatorname{ran} Y = \{0, 1, 2\}$. Die Wahrscheinlichkeiten $\mathbb{P}(X = s,\ Y = t)$ sind in der folgenden Tabelle zusammengefasst (z.B.: $\mathbb{P}(X = 3, Y = 0) = 0{,}15$):

$Y \backslash X$	2	3	5
0	0,05	0,15	0,1
1	0,1	0,2	0,1
2	0,05	0,2	0,05

Gemeinsame Verteilung diskreter Zufallsvariablen

$Y \backslash X$	2	3	5
0	0,05	0,15	0,1
1	0,1	0,2	0,1
2	0,05	0,2	0,05

Berechnen wir die Wahrscheinlichkeit $\mathbb{P}(X \leq 4, Y > 0)$. Diese Bedingung wird (zwischen den oberen Paaren) gerade durch $(2,1)$, $(2,2)$, $(3,1)$, $(3,2)$ erfüllt. Also gilt

$$\begin{aligned} \mathbb{P}(X \leq 4, Y > 0) &= f_{X,Y}(2,1) + f_{X,Y}(2,2) + f_{X,Y}(3,1) + f_{X,Y}(3,2) \\ &= 0{,}1 + 0{,}05 + 0{,}2 + 0{,}2 = 0{,}55. \end{aligned}$$

Gemeinsame Verteilung diskreter Zufallsvariablen

- Eine solche Tabelle von nicht negativen Zahlen beschreibt die gemeinsame Verteilung von zwei Zufallsvariablen dann und nur dann, falls die Summe der Zahlen in der Tabelle gleich 1 ist.
- Diese Eigenschaft Charakterisiert die Wahrscheinlichkeitsfunktion.
- Eine entsprechende Behauptung gilt auch für mehr Variablen.

Gemeinsame Verteilung diskreter Zufallsvariablen

Definition. Falls die gemeinsame Verteilung eines Zufallsvektors $\underline{X} = (X_1, \ldots, X_n)$ angegeben ist, dann heißen die Verteilungen der einzelne Variablen X_i *Randverteilungen* (peremeloszlás, marginális eloszlás).

Gemeinsame Verteilung diskreter Zufallsvariablen

Zuerst bestimmen wir die Randverteilungen im Fall $n = 2$: wir betrachten den Zufallsvektor (X, Y).

Um die Ranverteilungen anzugeben, sollen wir die Funktionen f_X und f_Y bestimmen.

Die Ereignisse $\{Y = t\}$, wo t die Menge $\operatorname{ran} Y$ durchläuft, bilden ein vollständiges Ereignissystem, also zefällt das Ereignis $\{X = s\}$ in disjunkte Ereignisse folgenderweise:

$$\{X = s\} = \bigcup_{t \in \operatorname{ran} Y} \{X = s, Y = t\}.$$

Folglich gilt

$$f_X(s) = \mathbb{P}(X = s) = \sum_{t \in \operatorname{ran} Y} \mathbb{P}(X = s, Y = t) = \sum_{t \in \operatorname{ran} Y} f_{X,Y}(s, t).$$

für alle $s \in \operatorname{ran} X$.

Gemeinsame Verteilung diskreter Zufallsvariablen

Ähnlich ergibt sich die Verteilung von Y:

$$f_Y(t) = \mathbb{P}(Y = t) = \sum_{s \in \operatorname{ran} X} \mathbb{P}(X = s,\ Y = t) = \sum_{s \in \operatorname{ran} X} f_{X,Y}(s,t).$$

D.h.: Falls die Wertemenge der Vektor (X, Y) endlich ist und die zugehörige gemeinsame Wahrscheinlichkeitsfunktion in einer Tabelle angegeben werden kann, dann ergibt sich die Wahrscheinlichkeitsfunktion von X (bzw. Y) an der Stelle s (bzw. t), wenn die Werte von $f_{X,Y}$ in der zu s gehörigen Spalte (bzw. in der zu t gehörigen Zeile) aufsummiert werden.

Gemeinsame Verteilung diskreter Zufallsvariablen

Beispiel.

$Y \backslash X$	2	3	5
0	0,05	0,15	0,1
1	0,1	0,2	0,1
2	0,05	0,2	0,05

$$\begin{aligned} f_X(2) &= \mathbb{P}(X = 2) \\ &= \mathbb{P}(X = 2, Y = 0) + \mathbb{P}(X = 2, Y = 1) + \mathbb{P}(X = 2, Y = 2) \\ &= 0{,}05 + 0{,}1 + 0{,}05 = 0{,}2, \end{aligned}$$

$$f_X(3) = \mathbb{P}(X = 3) = 0{,}15 + 0{,}2 + 0{,}2 = 0{,}55,$$

$$f_X(5) = \mathbb{P}(X = 5) = 0{,}1 + 0{,}1 + 0{,}05 = 0{,}25,$$

Gemeinsame Verteilung diskreter Zufallsvariablen

Beispiel.

$Y \backslash X$	2	3	5
0	0,05	0,15	0,1
1	0,1	0,2	0,1
2	0,05	0,2	0,05

$$f_Y(0) = \mathbb{P}(Y = 0) = 0{,}05 + 0{,}15 + 0{,}1 = 0{,}3,$$

$$f_Y(1) = \mathbb{P}(Y = 1) = 0{,}1 + 0{,}2 + 0{,}1 = 0{,}4,$$

$$f_Y(2) = \mathbb{P}(Y = 2) = 0{,}05 + 0{,}2 + 0{,}05 = 0{,}3.$$

Gemeinsame Verteilung diskreter Zufallsvariablen

Für n Variablen $X_1, \ldots, X_n$ ergeben sich die Ranverteilungen ähnlich:

Das Ereignis $\{X_i = t_i\}$ zerfällt nach den Werten der anderen $n-1$ Variablen:

$$f_{X_i}(t_i) = \mathbb{P}(X_i = t_i) = \sum_{\substack{s_j \in \operatorname{ran} X_j \\ j \neq i}} f_{\underline{X}}(s_1, \ldots, s_{i-1}, t_i, s_{i+1}, \ldots, s_n).$$

Gemeinsame Verteilung diskreter Zufallsvariablen

Später werden wir den Erwartungswert des Produkts von 2 Varibلen benötigen. Um ihn zu bestimmen, benutzen wir die Definition des Erwartungswertes:

$$
\begin{aligned}
\mathbb{E}(XY) &= \sum_{m \in \operatorname{ran} XY} m \cdot \mathbb{P}(XY = m) \\
&= \sum_{m \in \operatorname{ran} XY} m \cdot \mathbb{P}\left(\bigcup_{st=m} \{X = s, Y = t\} \right) \\
&= \sum_{m \in \operatorname{ran} XY} m \cdot \sum_{st=m} \mathbb{P}(X = s, Y = t) \\
&= \sum_{s \in \operatorname{ran} X} \sum_{t \in \operatorname{ran} Y} st \cdot f_{X,Y}(s, t).
\end{aligned}
$$

Gemeinsame Verteilung diskreter Zufallsvariablen

Beispiel.

$Y \backslash X$	2	3	5
0	0,05	0,15	0,1
1	0,1	0,2	0,1
2	0,05	0,2	0,05

$$\begin{aligned}
\mathbb{E}(XY) &= \sum_{s\in\{2,3,5\}} \sum_{t=0}^{2} st \cdot \mathbb{P}(X = s, Y = t) \\
&= 2 \cdot 0 \cdot 0{,}05 + 3 \cdot 0 \cdot 0{,}15 + 5 \cdot 0 \cdot 0{,}1 \\
&\quad + 2 \cdot 1 \cdot 0{,}1 + 3 \cdot 1 \cdot 0{,}2 + 5 \cdot 1 \cdot 0{,}1 \\
&\quad + 2 \cdot 2 \cdot 0{,}05 + 3 \cdot 2 \cdot 0{,}2 + 5 \cdot 2 \cdot 0{,}05 \\
&= 0{,}2 + 0{,}6 + 0{,}5 + 0{,}2 + 1{,}2 + 0{,}5 = 3{,}2.
\end{aligned}$$

Unabhängigkeit diskreter Zufallsvariablen

- Die Unabhängigkeit von *Ereignissen* bedeutet, dass die Eintritte einiger Ereignisse die Wahrscheinlichkeiten der anderen Ereignisse nicht beeinflussen.
- Die analoge Eigenschaft kann auch für *Zufallsvariablen* formuliert werden: die Unabhängigkeit von Variablen bedeutet, dass die Werte einiger Variablen das Verhalten der anderen Variablen nicht beeinflussen.
- Beispiel: zwei Würfel werden geworfen, die Augenzahlen sind unabhängig voneinander.

Unabhängigkeit diskreter Zufallsvariablen

- Zur Unabhängigkeit von zwei Variablen ist es nicht genügend, die Unabhängigkeit von zwei zugehörige Ereignisse zu zeigen.
- Z.B: zwei Würzel werden geworfen, seien X die Summe und Y das Produkt der Augenzahlen. Dann sind die Ereignisse $\{X \text{ ist gerade}\}$ und $\{Y \leq 4\}$ unabhängig.
- Aber die zwei Quantitäten sind offenbar *nicht* unabhängig. Z.B.: falls das Produkt ungerade ist, dann muss die Summe gerade sein.
- Mehere Variablen sind nur dann unabhängig, falls *alle Ereignisse*, die anhand der Variablen ausgedrückt werden kann, unabhängig sind.
- Dazu ist es aber genügend, weniger anzunehmen:

Unabhängigkeit diskreter Zufallsvariablen

Definition. Die diskreten Zufallsvariablen $X, Y : \Omega \to \mathbb{R}$ sind *unabhängig*, falls die Ereignisse $\{X = s\}$ und $\{Y = t\}$ für alle $s \in \operatorname{ran} X$ und $t \in \operatorname{ran} Y$ unabhängig sind, d.h. falls die Gleichung

$$f_{X,Y}(s,t) = \mathbb{P}(X = s, Y = t) = \mathbb{P}(X = s)\mathbb{P}(Y = t) = f_X(s)f_Y(t)$$

für alle $s \in \operatorname{ran} X$ und $t \in \operatorname{ran} Y$ gilt.

Bemerkungen.

- Die obige Gleichungen können in der Gleichung $f_{X,Y} = f_X \cdot f_Y$ zwischen *Funktionen* zusammengefasst werden.
- Aus den obigen Gleichungen folgt, dass alle Paare von Ereignissen, die anhand X bzw. Y ausgedrückt werden können, unabhängig sind.

Unabhängigkeit diskreter Zufallsvariablen

Beispiel. Ein Würfel wird geworfen, sei X der Rest bei der Division der Augenzahl durch 2, und sei Y der Rest bei der Division durch 3.

Dann gelten $\operatorname{ran} X = \{0, 1\}$ und $\operatorname{ran} Y = \{0, 1, 2\}$, und beide Variablen sind gleichverteilt, also nimmt X (bzw. Y) jeden Wert mit Wahrscheinlichkeit $1/2$ (bzw. $1/3$) an.

Es ist leicht zu sehen, dass die Reste modulo 2 und 3 eine Zahl zwischen 1 und 6 eindeutig bestimmen. Also gilt

$$
\begin{aligned}
f_{X,Y}(s,t) &= \mathbb{P}(X = s, Y = t) \\
&= \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} = \mathbb{P}(X = s)\mathbb{P}(Y = t) = f_X(s) f_Y(t)
\end{aligned}
$$

für alle $s \in \operatorname{ran} X$ und $t \in \operatorname{ran} Y$, d.h.: X und Y sind unabhängig.

Unabhängigkeit diskreter Zufallsvariablen

- Es ist oft leicht zu zeigen, dass zwei Variablen X und Y *nicht* unabhängig sind, denn in diesem Fall ist es genug, zwei Werte $s \in \operatorname{ran} X$ und $t \in \operatorname{ran} Y$ zu finden, für die die Gleichung $\mathbb{P}(X = s, Y = t) = \mathbb{P}(X = s)\mathbb{P}(Y = t)$ *nicht* gilt.
- Um die Unabhängigkeit zu zeigen, muss man alle Gleichungen der obigen Form prüfen.
- Falls die Variablen einfach sind, dann kann ihre gemeinsame Verteilung in einer Tabelle angegeben werden, die diese Analyse einfacher macht.

Unabhängigkeit diskreter Zufallsvariablen

Beispiel. Seien X und Y Zufallsvariablen mit $\operatorname{ran} X = \{0, 1, 2\}$ und $\operatorname{ran} Y = \{0, 1\}$. Nehmen wir an, dass ihre gemeinsame Verteilung durch die folgende Tabelle angegeben ist:

$Y \backslash X$	0	1	2
0	$1/12$	$1/8$	$1/8$
1	$1/6$	$1/4$	$1/4$

Wir bestimmen zunächst die Randverteilungen (wir summen die Werte in den Spalten und Zeilen auf):

$Y \backslash X$	0	1	2	
0	$1/12$	$1/8$	$1/8$	$1/3 = f_Y(0)$
1	$1/6$	$1/4$	$1/4$	$2/3 = f_Y(1)$
	$1/4 = f_X(0)$	$3/8 = f_X(1)$	$3/8 = f_X(2)$	

Unabhängigkeit diskreter Zufallsvariablen

$Y \backslash X$	0	1	2	
0	$1/12$	$1/8$	$1/8$	$1/3 = f_Y(0)$
1	$1/6$	$1/4$	$1/4$	$2/3 = f_Y(1)$
	$1/4 = f_X(0)$	$3/8 = f_X(1)$	$3/8 = f_X(2)$	

Bei dieser Form der Tabelle ist es leicht zu sehen, ob das Produkt der zu den Spalten und Zeilen gehörigen Werte von f_X und f_Y die im Schnitt stehende Wahrscheinlichkeit geben.

In diesem Fall gelten alle nötigen Gleichungen, also sind X und Y unabhängig.

Unabhängigkeit diskreter Zufallsvariablen

Behauptung. Seien X und Y unabhängige diskrete Zufallsvariablen mit endlichen Erwartungswerten $\mathbb{E}(X)$ und $\mathbb{E}(Y)$. Dann ist auch $\mathbb{E}(XY)$ endlich und $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ gilt.

Beweis. Die Variablen X und Y sind unabhängig, also gilt $f_{X,Y} = f_X \cdot f_Y$. Weiter,

$$\mathbb{E}(XY) = \sum_{s \in \operatorname{ran} X} \sum_{t \in \operatorname{ran} Y} st \cdot f_{X,Y}(s,t) = \sum_{s \in \operatorname{ran} X} \sum_{t \in \operatorname{ran} Y} st \cdot f_X(s) f_Y(t)$$

$$= \left(\sum_{s \in \operatorname{ran} X} s \cdot f_X(s) \right) \left(\sum_{t \in \operatorname{ran} Y} t \cdot f_Y(t) \right) = \mathbb{E}(X)\mathbb{E}(Y).$$

Bemerkung. Die Endlichkeit der Werte $\mathbb{E}(X)$ und $\mathbb{E}(Y)$ garantiert die Richtigkeit der obigen Rechnung.

Unabhängigkeit diskreter Zufallsvariablen

Unabhängigkeit von mehreren Variablen:

Definition. Die diskreten Zufallsvariablen $X_1, \dots, X_n : \Omega \to \mathbb{R}$ sind *paarweise* bzw. *gemeinsam unabhängig*, falls die Ereignisse $\{X_1 = t_1\}, \dots, \{X_n = t_n\}$ für alle $t_1 \in \operatorname{ran} X_1, \dots, t_n \in \operatorname{ran} X_n$ paarweise bzw. gemeinsam unabhängig sind.

Beispiel. Seien $A_1, \dots, A_n \subset \Omega$ gemeinsam unabhänige Ereignisse. Dann sind die zugehörigen Indikatorvariablen $\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n}$ gemeinsam unabhängig.

Unabhängigkeit diskreter Zufallsvariablen

Beweis. Die Wertemenge der Indikatorvariablen ist die Menge $\{0, 1\}$, also müssen wir zeigen, dass die Ereignisse

$$\{\mathbb{1}_{A_1} = t_1\}, \ldots, \{\mathbb{1}_{A_n} = t_n\}$$

gemeinsam unabhänig sind, wo $t_i = 0$ oder 1 für alle $1 \leq i \leq n$.

Für eine Zahl $t = 0$ oder 1 und eine Menge A sei

$$A^t = \begin{cases} A, & \text{falls } t = 1, \\ \overline{A}, & \text{falls } t = 0, \end{cases}$$

dann gilt $\{\mathbb{1}_{A_i} = t_i\} = A_i^{t_i}$.

Unabhängigkeit diskreter Zufallsvariablen

Beweis. Die Ereignisse $A_1, \dots, A_n$ sind gemeinsam unabhängig, folglich sind auch $A_1^{t_1}, \dots, A_n^{t_n}$ gemeinsam unabhängig.

Also gilt

$$\mathbb{P}\left(\bigcap_{i \in I} \{\mathbb{1}_{A_i} = t_i\}\right) = \mathbb{P}\left(\bigcap_{i \in I} A_i^{t_i}\right) = \prod_{i \in I} \mathbb{P}(A_i^{t_i}) = \prod_{i \in I} \mathbb{P}(\mathbb{1}_{A_i} = t_i),$$

für alle Indexmengen $\emptyset \neq I \subset \{1, \dots, n\}$.

Transformierte Zufallsvariablen

Sei $X : \Omega \to \mathbb{R}$ eine diskrete Zufallsvariable.

- Wir werden oft solche Variablen betrachten, die Funktionen von X sind.
- Sei $g : \mathbb{R} \to \mathbb{R}$ eine Funktion und betrachten wir die transformierte Variable $g(X)$.
- Oft interessiert man sich nicht für die Verteilung von $g(X)$, sondern für verscheidene Parameter der Verteilung, z.B. für den Erwartungswert $\mathbb{E}(g(X))$.
- Unser Hauptbeispiel ist die Funktion $g(t) = t^2$, also die transformierte Variable X^2.

Definition. Seien X eine diskrete Zufallsvariable und k eine positive ganze Zahl, dann heißt der Erwartungswert $\mathbb{E}(X^k)$ (falls er definiert ist) der *k-te Moment* von X (k-adik momentum).

Transformierte Zufallsvariablen

Um den Erwartungswert $\mathbb{E}(g(X))$ zu berechnen, muss man die Verteilung von $g(X)$ *nicht* bestimmen, er kann auch anhand der Verteilung von X berechnet werden:

Satz. (Transformationsformel) Sei X eine diskrete Zufallsvariable und sei $g : \mathbb{R} \to \mathbb{R}$ eine Funktion. Dann ist auch $g(X)$ eine Zufallsvariable, und falls $\mathbb{E}(g(X))$ endlich ist, dann gilt

$$\mathbb{E}(g(X)) = \sum_{k \in \operatorname{ran} X} g(k) f_X(k) = \sum_{k \in \operatorname{ran} X} g(k) \mathbb{P}(X = k),$$

wo f_X die Wahrscheinlichkeitsfunktion von X ist. Im Spezialfall $g(t) = t^2$ gilt

$$\mathbb{E}(X^2) = \sum_{k \in \operatorname{ran} X} k^2 f_X(k) = \sum_{k \in \operatorname{ran} X} k^2 \mathbb{P}(X = k),$$

falls $\mathbb{E}(X^2)$ endlich ist.

Transformierte Zufallsvariablen

Beweis. Nach der Definition des Erwartungswertes gilt

$$\mathbb{E}(g(X)) = \sum_{s \in \operatorname{ran} g(X)} s \cdot \mathbb{P}(g(X) = s) = \sum_{s \in \operatorname{ran} g(X)} s \cdot \mathbb{P}(X \in g^{-1}(s)),$$

wo $g^{-1}(s)$ das Urbild der Menge $\{s\}$ undter der Funktion g, d.h.

$$g^{-1}(s) = \{t \in \mathbb{R} : \; g(t) = s\}.$$

Für eine fixe $s \in \operatorname{ran} g(X)$ gilt

$$\begin{aligned} s \cdot \mathbb{P}(X \in g^{-1}(s)) &= \sum_{t \in \operatorname{ran} X \cap g^{-1}(s)} g(t) \cdot \mathbb{P}(X = t) \\ &= \sum_{t \in \operatorname{ran} X \cap g^{-1}(s)} g(t) \cdot f_X(t), \end{aligned}$$

Transformierte Zufallsvariablen

Beweis. Dann erhalten wir

$$\mathbb{E}(g(X)) = \sum_{s \in \operatorname{ran} g(X)} \; \sum_{t \in \operatorname{ran} X \cap g^{-1}(s)} g(t) \cdot f_X(t).$$

Hier listen wir jedes Element von $\operatorname{ran} X$ gerade einmal, also folgt

$$\mathbb{E}(g(X)) = \sum_{k \in \operatorname{ran} X} g(k) f_X(k).$$

Bemerkung. Falls die Wertemenge von X abzählbar unendlich ist, dann folgt die Richtigkeit der Operationen aus der absoluten Konvergenz der (möglicherweise unendlichen) Summe

$$\mathbb{E}(g(X)) = \sum_{s \in \operatorname{ran} g(X)} s \cdot \mathbb{P}(g(X) = s).$$

Transformierte Zufallsvariablen

Beispiel. Sei $X \sim \mathrm{Poi}(\lambda)$ eine Poisson-verteilte Variable für eine $\lambda > 0$. Dann gilt

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

Transformierte Zufallsvariablen

Beispiel. Sei $X \sim \mathrm{Geo}(p)$ geometrisch verteilt mit dem Parameter $p \in (0;1]$. Wegen der Transformationsformel ergibt sich

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{k=1}^{\infty} k^2 \cdot \mathbb{P}(X=k) = \sum_{k=1}^{\infty} k^2 (1-p)^{k-1} p \\
&= 1^2 \cdot p + \sum_{k=2}^{\infty} k^2 (1-p)^{k-1} p \\
&= p + \sum_{m=1}^{\infty} (m+1)^2 (1-p)^m p \\
&= p + (1-p) \cdot \sum_{m=1}^{\infty} (m^2 + 2m + 1)(1-p)^{m-1} p.
\end{aligned}$$

Transformierte Zufallsvariablen

Die letzte Summe können wir als drei Summe schreiben:

$$\sum_{m=1}^{\infty}(m^2+2m+1)(1-p)^{m-1}p =$$

$$= \sum_{m=1}^{\infty} m^2(1-p)^{m-1}p + 2\sum_{m=1}^{\infty} m(1-p)^{m-1}p + \sum_{m=1}^{\infty}(1-p)^{m-1}p$$

$$= \mathbb{E}(X^2) + 2\mathbb{E}(X) + 1 = \mathbb{E}(X^2) + \frac{2}{p} + 1.$$

Also gilt

$$\begin{aligned}\mathbb{E}(X^2) &= p + (1-p)\mathbb{E}(X^2) + \frac{2(1-p)}{p} + (1-p) \\ &= \frac{2-p}{p} + (1-p)\mathbb{E}(X^2).\end{aligned}$$

Transformierte Zufallsvariablen

Nach Umordnung ergibt sich

$$p\mathbb{E}(X^2) = \frac{2-p}{p}, \qquad \mathbb{E}(X^2) = \frac{2-p}{p^2}.$$

Varianz

Bemerkung.

- Der Erwartungswert gibt den durchschnittlichen Wert einer Zufallsvariablen, aber er sagt nichts darüber, wie weit die Werte der Variablen vom Erwartungswert liegen.
- Zum Beispiel: sei X_n eine Zufalsvariable, die die Werte $\pm n$ mit Wahrscheinlichkeit $\frac{1}{2}$ annimmt, wo $n \in \mathbb{N}^+$ eine positive ganze Zahl ist.
- Dann gilt $\mathbb{E}(X_n) = 0$ für jede n, aber die Werte dieser Variablen liegen für eine große n beliebig weit von 0.

Varianz

Definition. Sei X eine diskrete Zufallsvariable. Falls der Erwartungswert

$$\mathbb{E}((X - \mathbb{E}(X))^2)$$

endlich ist, dann heißt er die *Varianz* (szórásnégyzet, variancia) der Zufallsvariablen X. Sie ist mit $\mathbb{D}^2(X)$ bezeichnet.

Die positive Quadratwurzel der Varianz heißt die *Standardabweichung* (szórás). Sie ist mit $\mathbb{D}(X)$ bezeichnet.

Bemerkung. Die Varianz ist definiert als die mittlere quadratische Abweichung einer Zufallsvariablen von ihrem Erwartungswert.

Varianz

Bemerkungen.

- Die Varianz einer diskreten Zufallsvariablen ist nicht immer definiert, weil $\mathbb{E}(X)$ nicht notwendig definiert oder endlich ist.
- Wenn $\mathbb{E}(X)$ endlich ist, dann ist $(X - \mathbb{E}(X))^2$ eine nichtnegative Zufallsvariable. Deshalb ist $\mathbb{E}((X - \mathbb{E}(X))^2)$ eine nichtnegative Zahl oder Unendlich.
- Im ersten Fall werden wir sagen, dass $\mathbb{D}^2(X)$ (oder $\mathbb{D}(X)$) endlich ist, und das wird in den folgenden Sätzen immer angenommen.

Varianz

Behauptung. Sei X eine diskrete Zufallsvariable, für die $\mathbb{D}^2(X)$ endlich ist. Dann sind auch $\mathbb{D}^2(X+c)$ und $\mathbb{D}^2(cX)$ für jede $c \in \mathbb{R}$ endlich, und

$$\mathbb{D}(X+c) = \mathbb{D}(X) \qquad (\mathbb{D}^2(X+c) = \mathbb{D}^2(X))$$

und

$$\mathbb{D}(cX) = |c|\,\mathbb{D}(X) \qquad (\mathbb{D}^2(cX) = c^2\mathbb{D}^2(X))$$

gelten.

Varianz

Beweis.

$$
\begin{aligned}
\mathbb{D}^2(X) &= \mathbb{E}([X - \mathbb{E}(X)]^2) \\
&= \mathbb{E}([X + c - \mathbb{E}(X) - c]^2) \\
&= \mathbb{E}([X + c - (\mathbb{E}(X) + \mathbb{E}(c))]^2) \\
&= \mathbb{E}([X + c - \mathbb{E}(X + c)]^2) =: \mathbb{D}^2(X + c)
\end{aligned}
$$

Varianz

Beweis.

$$\begin{aligned}
c^2\mathbb{D}^2(X) &= c^2\mathbb{E}([X - \mathbb{E}(X)]^2) \\
&= \mathbb{E}(c^2[X - \mathbb{E}(X)]^2) \\
&= \mathbb{E}([cX - c\mathbb{E}(X)]^2) \\
&= \mathbb{E}([cX - \mathbb{E}(cX)]^2) =: \mathbb{D}^2(cX)
\end{aligned}$$

Varianz

Behauptung. Sei X eine diskrete Zufallsvariable. Dann gilt

$$\mathbb{D}^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2,$$

d.h. wenn der Ausdruck auf einer Seite der Gleichung definiert und endlich ist, dann ist der Ausdruck auf der anderen Seite der Gleichung definiert und endlich, und sie sind gleich.

Beweis.

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

Folgerung. Sei X eine diskrete Zufallsvariable. Dann ist $\mathbb{D}^2(X)$ dann und nur dann endlich, falls $\mathbb{E}(X^2)$ und $\mathbb{E}(X)$ endlich sind.

Varianz

Beispiel. Eine Münze wird geworfen, sei der Wert von X 1, falls das Ergebnis eine Zahl ist, und 2, falls das Ergebnis ein Kopf ist.

- Wegen der Definition des Erwartungswertes haben wir

$$\mathbb{E}(X) = 1 \cdot \mathbb{P}(X = 1) + 2 \cdot \mathbb{P}(X = 2) = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = \frac{3}{2}.$$

- Wegen der Transformationsformel gilt

$$\mathbb{E}(X^2) = 1^2 \cdot \mathbb{P}(X = 1) + 2^2 \cdot \mathbb{P}(X = 2) = 1 \cdot \frac{1}{2} + 4 \cdot \frac{1}{2} = \frac{5}{2}.$$

- Also folgt

$$\mathbb{D}^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{5}{2} - \frac{9}{4} = \frac{1}{4}.$$

Varianz

Beispiel. Eine Würfel wird geworfen, sei X die Augenzahl.

- Früher bestimmten wir die Werte $\mathbb{E}(X) = \dfrac{7}{2}$ und $\mathbb{E}(X^2) = \dfrac{91}{6}$.
- Also folgt

$$\mathbb{D}^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}.$$

Varianz

Beispiel. Sei A ein Ereignis mit $\mathbb{P}(A) = p$, und sei $\mathbb{1}_A$ die Indikatorvariable von A.

- Wir sahen schon, dass $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A) = p$ gilt.
- Weiter gilt auch $\mathbb{1}_A^2 = \mathbb{1}_A$, also $\mathbb{E}(\mathbb{1}_A^2) = \mathbb{E}(\mathbb{1}_A) = p$.
- Deshalb ist die Varianz von $\mathbb{1}_A$

$$\mathbb{D}^2(\mathbb{1}_A) = \mathbb{E}(\mathbb{1}_A^2) - \mathbb{E}(\mathbb{1}_A)^2 = p - p^2 = p(1 - p).$$

Varianz

Beispiel. Sei $X \sim \mathrm{Poi}(\lambda)$ Poisson-verteilt mit dem Parameter $\lambda > 0$.

- Wir sahen früher, dass $\mathbb{E}(X) = \lambda$ und $\mathbb{E}(X^2) = \lambda^2 + \lambda$ gelten.
- Dann erhalten wir

$$\mathbb{D}^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Varianz

Beispiel. Sei $X \sim \mathrm{Geo}(p)$ geometrisch verteilt mit Parameter p.

- Wir sahen früher, dass $\mathbb{E}(X) = \dfrac{1}{p}$ und $\mathbb{E}(X^2) = \dfrac{2-p}{p^2}$ gelten.
- Also folgt

$$\mathbb{D}^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

Unabhängigkeit und Varianz

Satz. Seien X und Y *unabhängige* diskrete Zufallsvariablen, für die $\mathbb{D}^2(X)$ und $\mathbb{D}^2(Y)$ endlich sind. Dann ist $\mathbb{D}^2(X+Y)$ auch endlich, und es gilt

$$\mathbb{D}^2(X+Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y).$$

Beweis. Wir zeigten früher, dass

$$\mathbb{D}^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2, \qquad \mathbb{D}^2(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2$$

und auch $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ gelten, also folgt

$$\begin{aligned}\mathbb{D}^2(X) + \mathbb{D}^2(Y) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 + \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 \\ &= \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) - \\ &\qquad - (\mathbb{E}(X)^2 + 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(Y)^2)\end{aligned}$$

Unabhängigkeit und Varianz

Beweis.

$$\begin{aligned}
\mathbb{D}^2(X) + \mathbb{D}^2(Y) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 + \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 \\
&= \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) - \\
&\qquad - (\mathbb{E}(X)^2 + 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(Y)^2) \\
&= \mathbb{E}(X^2 + 2XY + Y^2) - (\mathbb{E}(X) + \mathbb{E}(Y))^2 \\
&= \mathbb{E}((X+Y)^2) - \mathbb{E}(X+Y)^2 \\
&= \mathbb{D}^2(X+Y).
\end{aligned}$$

Unabhängigkeit und Varianz

Ähnlich ergibt sich das folgende

Korollar. Seien $X_1, \ldots, X_n$ paarweise unabhängige diskrete Zufallsvariablen, für die $\mathbb{D}^2(X_1), \ldots, \mathbb{D}^2(X_n)$ endlich sind. Dann ist auch $\mathbb{D}^2(X_1 + \ldots + X_n)$ endlich, und

$$\mathbb{D}^2(X_1 + \ldots + X_n) = \mathbb{D}^2(X_1) + \ldots + \mathbb{D}^2(X_n).$$

Unabhängigkeit und Varianz

Wir berechnen die Varianz einer binomialverteilten Zufallsvariablen:

- Sei $X \sim \mathrm{Bin}(n; p)$.
- Betrachten wir gemeinsam unabhängige Ereignisse $A_1, \ldots, A_n \subset \Omega$ mit $\mathbb{P}(A_i) = p$.
- Dann gilt $\mathbb{1}_{A_1} + \cdots + \mathbb{1}_{A_n} \sim \mathrm{Bin}(n; p)$,
- und die Variablen $\mathbb{1}_{A_1}, \ldots, \mathbb{1}_{A_n}$ sind unabhängig.
- Der letzte Satz gibt

$$\begin{aligned} \mathbb{D}^2(X) &= \mathbb{D}^2(\mathbb{1}_{A_1} + \cdots + \mathbb{1}_{A_n}) \\ &= \mathbb{D}^2(\mathbb{1}_{A_1}) + \cdots + \mathbb{D}^2(\mathbb{1}_{A_n}) = np(1-p). \end{aligned}$$

Geometrische Wahrscheinlichkeitsräume

- In vielen Situationen ist es zweckmäßig, alle Elemente eines Intervalls als möglicher Wert einer zufälligen Quantität zu betrachten.
- Z.B.: Höhe/Lebensalter einer zufällig gewählten Person, die Wartezeit bis zu einem zufälligen Ereignis, usw.
- Wenn $a < b$, dann ist die Mächtigkeit des Intervalls $[a; b]$ nicht abzählbar, also ist die bisher aufgebaute Theorie nicht genügend, diese Probleme zu behandeln.

Geometrische Wahrscheinlichkeitsräume

Zuerst betrachten wir solche Situationen, wo die Wahl uniform zufällig ist:

1. Wenn eine reelle Zahl in einem Intervall $[a; b]$ gewählt wird, dann wollen wir erreichen, dass die Zahl in der ersten Hälfte oder im ersten Drittel des Intervalls mit Wahrscheinlichkeit $1/2$ bzw. $1/3$ fällt.

Im Allgemeinen: die Wahrscheinlichkeit dafür, dass die Zahl in einem Teilintervall $I \subset [a; b]$ fällt, sollte $\ell(I)/(b-a)$ sein, wo $\ell(I)$ die Länge des Intervalls I ist.

Geometrische Wahrscheinlichkeitsräume

Zuerst betrachten wir solche Situationen, wo die Wahl uniform zufällig ist:

2. Die Wahl von zwei unabhängig und uniform zufällig gewählten reellen Zahlen in den Intervallen $[a; b]$ bzw. $[c; d]$ entspricht der Wahl eines Punktes im Rechteck $R = [a; b] \times [c; d]$. (Das wird später präzis formuliert und in einem Beispiel bewiesen.)

Die Wahrscheinlichkeit, dass beide Zahl in der ersten Hälfte des entsprechenden Intervalls fallen, sollte $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ sein.

Im Allgemeinen: die Wahrscheinlichkeit dafür, dass der gewählte Punkt in einer Teilmenge $M \subset R$ fällt, sollte $A(M)/A(R)$ sein, wo A der Flächeninhalt bezeichnet.

Geometrische Wahrscheinlichkeitsräume

Geometrische Wahrscheinlichkeitsräume im Allgemeinen:

- Ergebnismenge:
  eine Teilmenge der Linie/der Ebene/des Raumes endlicher und positiver Länge/Flächeninhalts/Volumens
- Ereignisse: die Teilmenge der Ergebnismenge, für die die Länge/der Flächeninhalt/das Volumen definiert ist
- $B \subset \Omega \Rightarrow \mathbb{P}(B) = \dfrac{\ell(B)}{\ell(\Omega)}$ bzw. $\dfrac{A(B)}{A(\Omega)}$ bzw. $\dfrac{V(B)}{V(\Omega)}$

Bemerkung.

Die Länge/der Flächeninhalt/das Volumen eines Punktes ist 0, also gilt $\mathbb{P}(\{x\}) = 0$ für jeden Punkt x in einem geometrischen Wahrscheinlichkeitsraum. (Das passiert nie in einer abzählbaren Ergebnismenge.)

Ähnlich ist die Wahrscheinlichkeit jeder eindimensionalen Kurve der Ebene/des Raumes und jeder zweidimensionalen Fläche des Raumes Null.

Geometrische Wahrscheinlichkeitsräume

- Die Länge/der Flächeninhalt/das Volumen kann nicht für alle Teilmengen der Linie/der Ebene/des Raums definiert werden!
- Die Gesamtheit der Mengen, für die dies möglich ist, hat eine bestimmte Struktur, sie bilden eine sogenannte σ-Algebra.
- Die Konstruktion der obigen Begriffe führt dann zur Definition des Lebesgue-Maßes in $\mathbb{R}^n$.
- Diese Überlegungen können in viel größerer Allgemeinheit durchgeführt werden, das zur allgemeinen Maßtheorie und zu einem unfassenden Aufbau der Wahrscheinlichkeitstheorie fürht.
- Die allgemeinen Grundbegriffe werden nächst auch hier definiert.

Wahrscheinlichkeitsräume

Definition. Sei Ω eine Menge. Dann ist $\mathcal{F} \subset \mathcal{P}(\Omega)$ eine *σ-Algebra*, falls

- $\Omega \in \mathcal{F}$,
- $A \in \mathcal{F} \Rightarrow \overline{A} \in \mathcal{F}$
  ($\mathcal{F}$ ist komplementstabil),
- $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$
  ($\mathcal{F}$ ist sigma-vereinigungsstabil)

gelten.

Beispiele. Die Mengen $\mathcal{F}_1 = \{\emptyset, \Omega\}$ und $\mathcal{F}_2 = \mathcal{P}(\Omega)$ sind immer σ-Algebren.

Zufallsvariablen

Bemerkung.

- In unseren Wahrscheinlichkeitsmodellen werden die Ereignisse immer eine σ-Algebra bilden.
- $\mathcal{F}_1$ enthält zu wenig Teilmengen, sie ist unnützlich.
- Zwar bei einer abzählbaren Ergebnismenge nahmen wir an, dass $\mathcal{F} = \mathcal{F}_2 = \mathcal{P}(\Omega)$ gilt, aber es ist auch dort nicht notwendig, und im Allgemeinen ist es einfach nicht möglich, denn $\mathcal{F}_2$ enthält oft zu viel Teilmengen und dann kann kein Wahrscheinlichkeitsmaß mit Definitionsmenge $\mathcal{F}_2$ angegeben werden.
- Es ist im Allgemeinen nicht so einfach, eine gute σ-Algebra und dazu ein Maß zu definieren/beschreiben.

Wahrscheinlichkeitsräume

Definition. Seien Ω eine Menge, $\mathcal{F} \subset \mathcal{P}(\Omega)$ eine σ-Algebra und $\mathbb{P} : \mathcal{F} \to [0; 1]$ ein Wahrscheinlichkeitsmaß (d.h. $\mathbb{P}(\Omega) = 1$ und $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ gelten für beliebige paarweise dijunkte Ereignisse $A_1, A_2, \ldots \subset \mathcal{F}$).
Dann ist der Tripel $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum.

Die Menge Ω heißt die Ergebnismenge, und die Elementen von $\mathcal{F}$ werden Ereignisse genannt.

Behauptung. Die oben beschriebenen geometrischen Räume sind Wahrscheinlichkeitsräume im Sinne der obigen Definition. Die σ-Algebra $\mathcal{F}$ besteht in diesen Fällen aus den Mengen, für die die Länge/der Flächeninhalt/das Volumen definiert ist. Diese sind die sogenannten Lebesgue-messbaren Mengen.

Wahrscheinlichkeitsräume

Behauptung. Sei $\mathcal{F} \subset \mathcal{P}(\Omega)$ eine σ-Algebra. Dann gelten

- $\emptyset \in \mathcal{F}$
- $A, B \in \mathcal{F} \Rightarrow A \cup B, A \cap B, A \setminus B \in \mathcal{F}$
  ($\mathcal{F}$ ist vereinigungsstabil, schnittstabil und differenzmengenstabil)
- $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$
  ($\mathcal{F}$ ist σ-schnittstabil)

Beweis: Übung.

Bemerkung. Die Definition der σ-Algebra und die obige Behauptung geben, dass die Ergebnisse der früher definierten Verknüpfungen von Ereignissen immer Ereignisse sind.

Wahrscheinlichkeitsräume

Die im diskreten Fall angegebenen Eingeschaften des Wahrscheinlichkeitsmaßes und ihre Beweise bleiben auch in einem allgemeinen Wharscheinlichkeitsraum gültig:

Behauptung. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeistraum. Dann gelten

- $\mathbb{P}(\emptyset) = 0$,
- $A, B \in \mathcal{F}$, $A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$,
- $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$ für alle $A \in \mathcal{F}$,
- $A, B \in \mathcal{F}$, $B \subset A \Rightarrow \mathbb{P}(B) \leq \mathbb{P}(A)$,
- die Siebformel und die Boolsche Ungleichung gelten.

Wahrscheinlichkeitsräume

Behauptung. (Stetigkeit von Wahrscheinlichkiestmaßen) Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeistraum, und seien $A_1, A_2, \ldots \in \mathcal{F}$ Ereignisse.

1. $\mathbb{P}$ ist *stetig von unten*: falls
$$A_1 \subset A_2 \subset \ldots, \qquad \text{und} \qquad A = \bigcup_{n=1}^{\infty} A_n,$$
dann gilt $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.
2. $\mathbb{P}$ ist *stetig von oben*: falls
$$A_1 \supset A_2 \supset \ldots, \qquad \text{und} \qquad A = \bigcap_{n=1}^{\infty} A_n,$$
dann gilt $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

Wahrscheinlichkeitsräume

Beweis. 1. Sei $A_0 = \emptyset$, und definieren wir die Ereignisse B_k durch

$$B_k = A_k \setminus A_{k-1} \quad (k \geq 1).$$

Dann sind $B_1, B_2, \dots$ paarweise disjunkte Ereignisse: falls $i > j \geq 1$, dann

$$B_j \subset A_j \subset A_{i-1}, \quad A_{i-1} \cap B_i = \emptyset$$

also $B_i \cap B_j = \emptyset$. Weiter gilt

$$\bigcup_{k=1}^{\infty} B_k = \bigcup_{k=1}^{\infty} A_k = A,$$

weil es für jedes Element $\omega \in A$ einen kleinsten Index $k \geq 1$ gibt, für den $\omega \in A_k$ und $\omega \notin A_{k-1}$, also auch $\omega \in B_k$ gelten.

Wahrscheinlichkeitsräume

Beweis. 1. Wegen der σ-Additivität von $\mathbb{P}$ gilt dann

$$\mathbb{P}(A) = \sum_{k=1}^{\infty} \mathbb{P}(B_k) = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{P}(B_k).$$

Durch vollständige Induktion ergibt sich

$$\sum_{k=1}^{n} \mathbb{P}(B_k) = \mathbb{P}(A_n).$$

Für $n = 1$ gilt nämlich $\mathbb{P}(B_1) = \mathbb{P}(A_1)$, und falls die Behauptung für $n-1$ richtig ist, dann folgt

$$\sum_{k=1}^{n} \mathbb{P}(B_k) = \mathbb{P}(B_n) + \sum_{k=1}^{n-1} \mathbb{P}(B_k) = \mathbb{P}(A_n \setminus A_{n-1}) + \mathbb{P}(A_{n-1}) = \mathbb{P}(A_n).$$

Beweis. 2. Definieren wir die Ereignisse C_k durch $C_k = \Omega \setminus A_k$ $(k \geq 1)$, dann gilt

$$C_1 \subset C_2 \subset \dots, \qquad \bigcup_{k=1}^{\infty} C_k = \Omega \setminus \left(\bigcap_{k=1}^{\infty} A_k \right)$$

Wegen 1. ergibt sich

$$\begin{aligned} 1 - \mathbb{P}\left(\cap_{k=1}^{\infty} A_k\right) &= \mathbb{P}\left(\cup_{k=1}^{\infty} C_k\right) = \lim_{n\to\infty} \mathbb{P}(C_n) \\ &= \lim_{n\to\infty} (1 - \mathbb{P}(A_n)) = 1 - \lim_{n\to\infty} \mathbb{P}(A_n), \end{aligned}$$

und die Behauptung folgt nach Umordnung.

Zufallsvariablen

Beispiel 1.

- Eine Zahl wird im Intervall $[0; 1]$ zufällig gewählt.
- Sei $\Omega = [0; 1]$ und sei Y der Wert der gewählten Zahl.
- Also ist $Y : \Omega \to \mathbb{R}$, $Y(\omega) = \omega$ eine Funktion mit einer überabzählbaren Defintions- und Bildmenge.
- Wir definierten die Zufallsvariablen in diesen Fällen noch nicht.

Zufallsvariablen

Beispiel 2.

- Sei K der Einheitskreis in der Ebene (mit Mittelpunkt O, wo O der Ursprung ist).
- Wir wählen einen Punkt zufällig in K. Also betrachten wir K als Ergebnismenge.
- Sei Z der Abstand zwischen O und dem gewählten Punkt. Dann ist $Z : K \to \mathbb{R}$, $Z(P) = |OP|$ eine Funktion (auch mit einer überabzählbaren Defintions- und Bildmenge).
- Falls wir $P = (x, y)$ schreiben, so gilt

$$Z(P) = Z((x, y)) = \sqrt{x^2 + y^2}.$$

Zufallsvariablen

Auch in diesen Fällen können wir über die Ereginisse $\{Y = t\}$ oder $\{Z = t\}$ sprechen, weil diese Menge Ereignisse von geometrischen Wahrscheinlichkeitsräumen sind:

$$\{Y = t\} = \{t\}, \ \{Z = t\} = \partial K_t \text{ für jede Zahl } t \in [0;1],$$

wo ∂K_t der Rand des Kreises K_t mit Mittelpunkt O und Radius t ist.

Die Länge bzw. der Flächeninhalt dieser Menge sind Null, also gilt

$$\mathbb{P}(Y = t) = \mathbb{P}(Z = t) = 0$$

für jede $t \in \mathbb{R}$.

Zufallsvariablen

D.h.: die Wahrscheinlichkeiten $\mathbb{P}(Y = t)$ bzw, $\mathbb{P}(Z = t)$ können nicht eine Verteilung beschreiben (sie geben zu wenig Information).

Stattdessen werden wir die Wahrscheinlichkeiten $\mathbb{P}(Y \leq t)$ und $\mathbb{P}(Z \leq t)$ $(t \in \mathbb{R})$ benutzen:

$$\{Y \leq t\} = \begin{cases} \emptyset, & \text{falls } t < 0 \\ [0; t], & \text{falls } 0 \leq t < 1 \\ [0; 1], & \text{falls } t \geq 1, \end{cases}$$

und dann

$$\mathbb{P}(Y \leq t) = \begin{cases} 0, & \text{falls } t < 0, \\ \dfrac{\ell([0;t])}{\ell([0;1])} = t, & \text{falls } 0 \leq t < 1, \\ 1, & \text{falls } t \geq 1. \end{cases}$$

Zufallsvariablen

D.h.: die Wahrscheinlichkeiten $\mathbb{P}(Y = t)$ bzw, $\mathbb{P}(Z = t)$ können nicht eine Verteilung beschreiben (sie geben zu wenig Information).

Stattdessen werden wir die Wahrscheinlichkeiten $\mathbb{P}(Y \leq t)$ und $\mathbb{P}(Z \leq t)$ $(t \in \mathbb{R})$ benutzen:

$$\{Z \leq t\} = \begin{cases} \emptyset, & \text{falls } t < 0 \\ K_t, & \text{falls } 0 \leq t < 1 \\ K_1, & \text{falls } t \geq 1, \end{cases}$$

also

$$\mathbb{P}(Z \leq t) = \begin{cases} 0, & \text{falls } t < 0, \\ \dfrac{A(K_t)}{A(K_1)} = \dfrac{t^2\pi}{\pi} = t^2, & \text{falls } 0 \leq t < 1, \\ 1, & \text{falls } t \geq 1. \end{cases}$$

Zufallsvariablen

In den obigen Fällen sind die Menge $\{Y \leq t\}$ und $\{Z \leq t\}$ auch Ereignisse in geometrischen Wahrscheinlichkeitsräumen.

Wie wird es im Allgemeinen garantiert, dass die Mengen dieser Typen Ereignisse sind?

Antwort: wir verlangen, diese Mengen Ereignisse zu sein!

Auch die Mengen $\{X < t\}$, $\{X \in [a; b]\}$, usw. sollten Ereignisse sein, so können die Wahrscheinlichkeiten $\mathbb{P}(X < t)$, $\mathbb{P}(X \leq t)$, $\mathbb{P}(a \leq X \leq b)$ definiert werden.

Zufallsvariablen

Definition. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine Funktion $X : \Omega \to \mathbb{R}$ ist eine *Zufallsvariable*, falls $\{X \leq t\} \in \mathcal{F}$ für jede $t \in \mathbb{R}$ gilt (d.h. falls die Menge $\{X \leq t\} \subset \Omega$ ein Ereignis ist).

Die Variable X heißt *diskret*, falls ihre Bildmenge abzählbar (endlich oder abzählbar undendlich) ist.

Die Variable X heißt *einfach*, falls ihre Bildmenge endlich ist.

Achtung: in der obigen Definition nehmen wir nichts über die Mächtigkeit von Ω an, und zwar wir die diskrete Variablen früher mit abzählbaren Definitionsmengen betracheten, aber das war nur ein Spezialfall der letzten völlig allgemeinen Definition.

Zufallsvariablen

Behauptung. Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Dann sind auch die folgende Teilmenge von Ω für jede $s, t \in \mathbb{R}$ Ereignisse:

$$\{X < t\}, \{X > t\}, \{X \geq t\}, \{X = t\}, \{s < X < t\}, \dots$$

Zum Beweis kann man die Eigenschaften der σ-Agebra $\mathcal{F}$ benutzen. Z.B.

$$\{X > t\} = \overline{\{X \leq t\}} \in \mathcal{F}.$$

Zufallsvariablen

Behauptung. Wenn $X, Y : \Omega \to \mathbb{R}$ Zufallsvariablen sind, dann sind auch

- $X \pm Y : \Omega \to \mathbb{R}$, $(X \pm Y)(\omega) := X(\omega) \pm Y(\omega)$ $(\forall \omega \in \Omega)$
- $X \cdot Y : \Omega \to \mathbb{R}$, $(X \cdot Y)(\omega) := X(\omega) \cdot Y(\omega)$ $(\forall \omega \in \Omega)$

Zufallsvariablen. Wenn zusätzlich $Y \neq 0$ gilt, dann ist auch

- $\dfrac{X}{Y} : \Omega \to \mathbb{R}$, $\dfrac{X}{Y}(\omega) := \dfrac{X(\omega)}{Y(\omega)}$ $(\forall \omega \in \Omega)$

eine Zufallsvariable.

Zufallsvariablen

Zurück zu unseren Beispielen Y und Z.

Noch einmal:

$$\{Y \leq t\} = \begin{cases} \emptyset, & \text{falls } t < 0 \\ [0; t], & \text{falls } 0 \leq t < 1 \\ [0; 1], & \text{falls } t \geq 1, \end{cases}$$

und

$$\{Z \leq t\} = \begin{cases} \emptyset, & \text{falls } t < 0 \\ K_t, & \text{falls } 0 \leq t < 1 \\ K_1, & \text{falls } t \geq 1, \end{cases}$$

Diese Mengen sind Ereignisse, also sind Y und Z Zufallsvariablen (im Sinne der obigen Definition).

Verteilungsfunktion

Wie kann die Verteilung einer Zufallsvarible im Allgemeinen definiert/beschrieben werden?

Erinnerung: $\mathbb{P}(Y = t) = \mathbb{P}(Z = t) = 0$ für jede $t \in \mathbb{R}$, also sagen diese Wahrscheinlichkeiten über die Verteilung von Y und Z nichts.

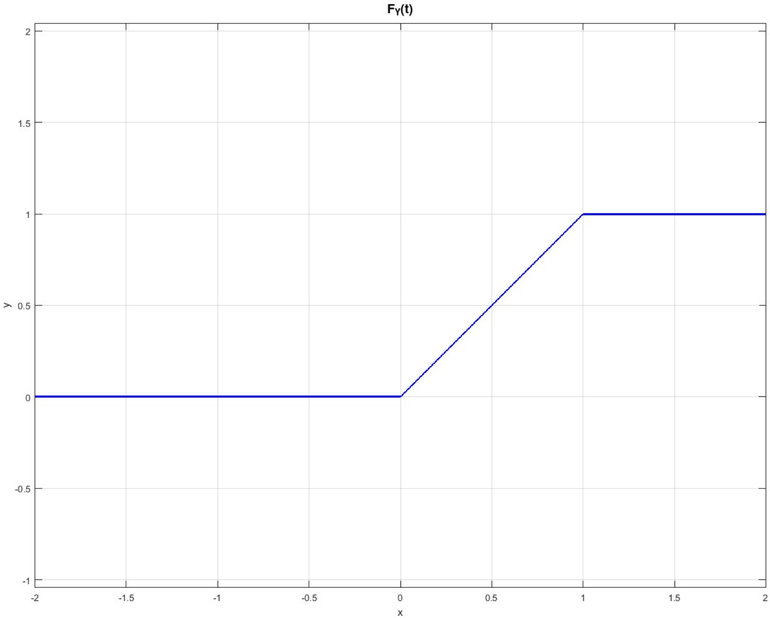
Stattdessen kann man die Wahrscheinlichkeiten $\mathbb{P}(Y \leq t)$, $\mathbb{P}(Z \leq t)$ $(t \in \mathbb{R})$ benutzen. Diese sind Funktionen von t:

$$\mathbb{P}(Y \leq t) = \begin{cases} 0, & \text{falls } t < 0, \\ t, & \text{falls } 0 \leq t < 1, \\ 1, & \text{falls } t \geq 1. \end{cases}$$

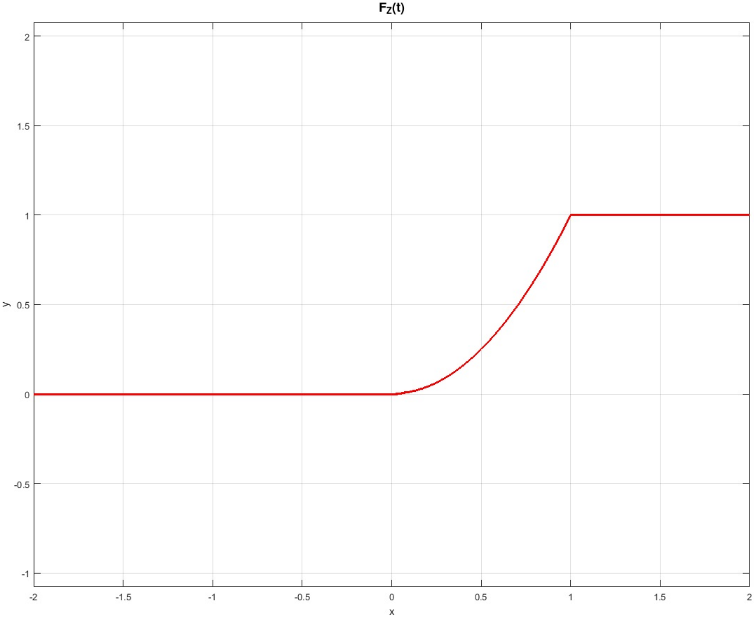
und

$$\mathbb{P}(Z \leq t) = \begin{cases} 0, & \text{falls } t < 0, \\ t^2, & \text{falls } 0 \leq t < 1, \\ 1, & \text{falls } t \geq 1. \end{cases}$$

Verteilungsfunktion



Verteilungsfunktion



Verteilungsfunktion

Definition. Sei X eine Zufallsvariable. Die Funktion

$$F_X : \mathbb{R} \to [0;1]$$
$$F_X(t) = \mathbb{P}(X \leq t) \qquad (t \in \mathbb{R})$$

heißt die *Verteilungsfunktion* (eloszlásfüggvény) von X.

Bemerkung. Die Verteilungsfunktion gibt/beschreibt die Verteilung einer Zufallsvariablen, d.h.: sie bestimmt die Wahrscheinlichkeit $\mathbb{P}(X \in A)$ für jede "schöne" Teilmenge $A \subset \mathbb{R}$.

Verteilungsfunktion

Behauptung. Wenn F_X die Verteilungsfunktion der Zufallsvariablen X ist, dann gilt

$$F_X(t) - F_X(s) = \mathbb{P}(s < X \leq t)$$

für beliebige reelle Zahlen $s < t$.

Beweis.

$$\{X \leq t\} = \{X \leq s\} \cup \{s < X \leq t\}, \quad \{X \leq s\} \cap \{s < X \leq t\} = \emptyset,$$

also

$$\mathbb{P}(X \leq t) = \mathbb{P}(X \leq s) + \mathbb{P}(s < X \leq t),$$

d.h.

$$\mathbb{P}(s < X \leq t) = \mathbb{P}(X \leq t) - \mathbb{P}(X \leq s) = F_X(t) - F_X(s).$$

Verteilungsfunktion

Wie erkennt man, dass eine Funktion ist auch die Verteilungsfunktion einer Zufallsvariablen ist?

Behauptung. Eine Funktion $F : \mathbb{R} \to [0; 1]$ ist genau dann die Verteilungsfunktion einer Zufallsvariablen, falls F die folgenden Eigenschaften erfüllt:

1. F ist (nicht unbedingt streng) monoton steigend,
2. F ist rechtseitig stetig,
3. $\lim_{t\to-\infty} F(t) = 0$, $\lim_{t\to\infty} F(t) = 1$.

Z.B.: wenn F_X die Verteilungsfunktion der Zufallsvariablen X ist, und $s < t$ gilt, dann gild auch $\{X \leq s\} \subset \{X \leq t\}$, also

$$F_X(s) = \mathbb{P}(X \leq s) \leq \mathbb{P}(X \leq t) = F_X(t).$$

Verteilungsfunktion

Beispiel. Ein Würfel wird geworfen, sei X das Ergebnis. Dann gilt $\operatorname{ran} X = \{1,2,3,4,5,6\}$, also ist X diskret.
Aber $F_X(t)$ ist definiert für jede reelle Zahl $t \in \mathbb{R}$, zum Beispiel

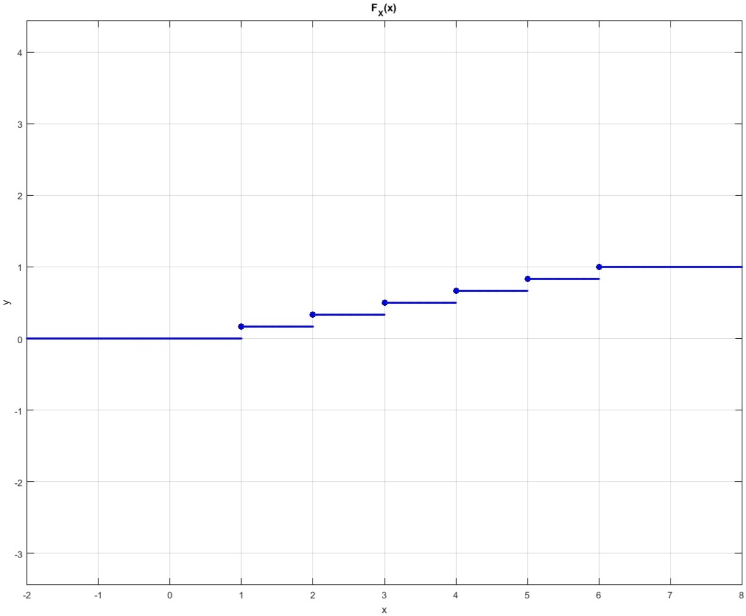
$$F_X(2{,}3) = \mathbb{P}(X \leq 2{,}3) = \mathbb{P}(X = 1 \text{ oder } X = 2) = \frac{2}{6}.$$

Für eine beliebige Zahl t:

$$F_X(t) = \begin{cases} 0, & \text{falls } t < 1, \\ \dfrac{\lfloor t \rfloor}{6}, & \text{falls } 1 \leq t < 6, \\ 1, & \text{falls } t \geq 6. \end{cases}$$

Bemerkung. Diese Funktion ist nicht stetig. Im Allgemeinen: die Verteilungsfunktion einer diskreten Variablen ist nie stetig.

Verteilungsfunktion



Verteilungsfunktion

Im diskreten Fall benutzten wir die Wahrscheinlichkeiten $\mathbb{P}(X = t)$, um die Verteilungen zu beschreiben. Wie erhält man diese aus der Verteilungsfunktion?

Behauptung. Sei X eine Zufallsvariable mit Verteilungsfunktion F_X. Dann gilt

$$\mathbb{P}(X = t) = \mathbb{P}(X \leq t) - \mathbb{P}(X < t) = F_X(t) - \lim_{s \to t-0} F_X(s).$$

Beweis. Sei $s_1, s_2, \ldots \in \mathbb{R}$ eine streng monoton steigende Folge mit $\lim_{n\to\infty} s_n = t$. Dann gilt

$$\{X \leq s_1\} \subset \{X \leq s_2\} \subset \ldots, \quad \bigcup_{k=1}^{\infty} \{X \leq s_k\} = \{X < t\}$$

und die Behauptung folgt aus der (unteren) Stetigkeit von $\mathbb{P}$.