Die Zufallsvariable und ihre Verteilung

Die Zufallsvariable

In der Wahrscheinlichkeitstheorie bzw. Statistik betrachtet man **Zufallsvariablen**.

> Eine Zufallsvariable ist eine Funktion, die Ergebnissen eines Zufallsexperimentes reelle Zahlen zuordnet.

Wenn das Zufallsexperiment ein Intelligenztest ist, so wird einer Person z.B. der Intelligenzquotient zugeordnet. Der zugeordnete Wert wird auch **Ausprägung** der Zufallsvariable genannt.

Man unterscheidet zwischen folgenden Arten von Zufallsvariablen nach ihren Ausprägungen:

- **Quantitative Variablen**, wenn die Ausprägungen Zahlenwerte sind (z.B. die Anzahl der Punkte in einem Intelligenztest). Weiters unterscheidet man:
  - *stetige Variable*: wenn die Anzahl der Ausprägungen unendlich ist (z.B. das Gewicht einer Person, zwischen zwei Merkmalsausrägungen können beliebig viele andere liegen)
  - *diskrete Variable*: wenn die Anzahl der Ausprägungen endlich ist (z.B. die Anzahl der Zeilen einer Matrix, hier liegen zwischen zwei Merkmalsausprägungen nur endlich viele andere)
- **Qualitative Variablen**, deren Ausprägung ein kategorischer Wert ist. Man unterscheidet
  - *dichotome Variablen:* zwei Kategorien (z.B. Ja/Nein)
  - *polynome Variablen:* mehr als zwei Kategorien (z.B. Afrikaner, Asiate, Europäer usw.)

Variablen werden mit lateinischen Großbuchstaben bezeichnet (z.B. X), deren Realisierungen mit lateinischen Kleinbuchstaben (x_i ist die i-te Realisierung der Zufallsvariable X).

Die Wahrscheinlichkeitsverteilung

Die **Wahrscheinlichkeitsverteilung** einer Zufallsvariable X ist definiert als

$$F_X(t) = P(X \leq t).$$

D.h. die Wahrscheinlichkeit, daß die Zufallsvariable X kleiner als der fixe Wert t ist, wird durch die Funktion $F_X(t)$ beschrieben (X ist der Index von F, weil dadurch hingewiesen wird, daß F die Verteilungsfunktion der Zufallsvariable X ist, und nicht einer anderen Zufallsvariable).

Die Wahrscheinlichkeitsverteilung ist sozusagen die theroretische Verteilung eines Ereignisses. Wenn man etwa das Zufallsexperiment Würfelwurf betrachtet, so bestimmt die Wahrscheinlichkeitsverteilung, mit welchen Wahrscheinlichkeiten die einzelnen Ausprägungen auftreten.

Häufig ist die theroretische Wahrscheinlichkeitsverteilung einer Zufallsvariable nicht bekannt. Man führt eine Anzahl von Experimenten durch, um diese Verteilung (näherungsweise) zu bestimmen.

Die Verteilungsfunktion hat folgende Eigenschaften:

1. F_X ist monoton nicht fallend, d.h. wenn $t_1 < t_2$, dann ist $F(t_1) \leq F(t_2)$
2. F_x ist rechtsseitig stetig, d.h. der rechtsseitige Grenzwert $\lim_{h \to 0} F_X(t+h) = F_X(t)$
3. $\lim_{x \to \infty} F_X(t) = 1$ und $\lim_{x \to 0} F_X(t) = 0$

Im folgenden unterscheiden wir zwischen der Verteilung einer **stetigen** und einer **diskreten** Zufallsvariable.

- **X diskret verteilt**

  Die Verteilungsfunktion ist

  $$F_X(t) = P(X \leq t) = \underbrace{\sum_{i:x_1 \leq t} P(X = x_i)}_{\text{Summation über alle } x_i \text{ die kleiner gleich } t \text{ sind}}$$

  Im diskreten Fall ist die Verteilungsfunktion $F_X(t)$ die Aufsummierung der sogenannten *Punktwahrscheinlichkeiten* $P(X = x_i)$, welche man auch als f_X (die **Wahrscheinlichkeitsfunktion**) bezeichnet, d.h.

  $$f_X = P(X = x).$$

- **X stetig verteilt**

  Die Verteilungsfunktion ist

  $$F_X(t) = \int_{-\infty}^{t} f_X(x)dx.$$

  Im stetigen Fall wird die sogenannte **Wahrscheinlichkeitsdichte** $f_X(x)$ bis t integriert um die Verteilungsfunktion $F_X(t) = \alpha$ zu erhalten. **Bemerkung:** das Integral der Wahrscheinlichkeitsdichte von $-\infty$ bis ∞ beträgt 1! Für reelle Zahlen $a \leq b$ gilt

  $$\begin{aligned} P(X \leq b) &= F(b) = \int_{-\infty}^{b} f(x)dx \\ P(a < X < b) &= F(b) - F(a) = \int_{a}^{b} f(x)dx \\ P(a < X) &= 1 - F(a) = \int_{a}^{\infty} f(x)dx. \end{aligned}$$

  Man beachte: Im Fall einer stetigen Variable gilt stets

  $$P(X = x) = F(x) - F(x) = 0,$$

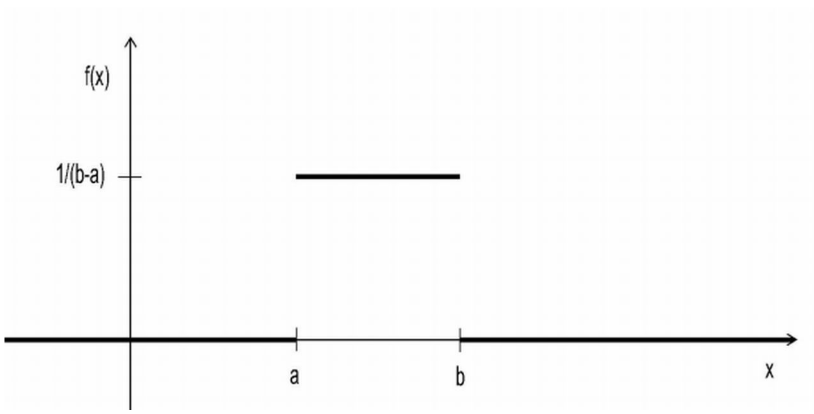
  d.h. für eine beliebige feste Zahl x nimmt X den Wert x mit Wahrscheinlichkeit 0 an, man sagt die "Punktwahrscheinlichkeit" ist stets gleich 0.

Beispiel:
Wir wollen nun die Verteilung der Zufallsvariable **Würfelwurf** beschreiben, dazu folgendes Beispiel Link zum Würfelwurf-Beispiel

Beispiel:
Ein Beispiel einer stetigen Verteilung ist die sogenannte **Gleichverteilung**, welche die folgende Wahrscheinlichkeitsfunktion besitzt:

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

Dazu eine Grafik:



Die Verteilungsfunktion erhält man durch Integration der Dichtefunktion, d.h. $F_X(t) =$

- für $x < a$:

$$\int_{-\infty}^{t} \underbrace{f(x)}_{=0 \text{ für } x<a} dx \int_{-\infty}^{t} 0dx = 0$$

- für $a \leq x \leq b$:

$$\begin{aligned} \int_{-\infty}^{t} f(x)dx &= \int_{-\infty}^{a} \underbrace{f(x)}_{=0 \text{ für } x<a} dx + \int_{a}^{t} \frac{1}{b-a}dx = 0 + \frac{x}{b-a}\Big|_{a}^{t} \\ &= \frac{t}{b-a} - \frac{a}{b-a} = \frac{t-a}{b-a} \end{aligned}$$

- für $x > b$

$$\begin{aligned}
\int\limits_{-\infty}^{t} f(x)dx &= \int\limits_{-\infty}^{a} \underbrace{f(x)}_{=0} dx + \int\limits_{a}^{b} f(x)dx + \int\limits_{b}^{t} \underbrace{f(x)}_{=0 \text{ für } x>b} dx \\
&= \left.\frac{x}{b-a}\right|_{a}^{b} = \frac{b-a}{b-a} = 1
\end{aligned}$$

d.h.

$$F_X(t) = \begin{cases} 0 & f\ddot{u}r\ x < a \\ \frac{t-a}{b-a} & \text{für } a \leq x \leq b \\ 1 & f\ddot{u}r\ x > b \end{cases}$$

Ein grafische Darstellung der Verteilungsfunktion der Gleichverteilung ist in folgendem Plot zu sehen: