1. Zufallsvariable und Verteilungsfunktion

1.1 Die Wahrscheinlichkeits- und Verteilungsfunktion

- **Notation:**

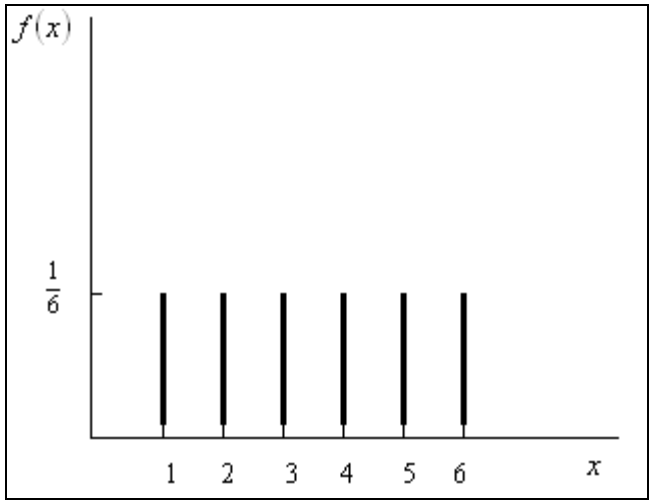
  X : Zufallsvariable, x_i : eine bestimmte Ausprägung von X

- **Wahrscheinlichkeitsfunktion einer diskreten Zufallsvariable:**

  Eine Funktion $f(x_i)$, die jeder Ausprägung x_i der (abzählbar unendlichen) Zufallsvariable X eine Wahrscheinlichkeit $P(X = x_i)$ zuordnet heißt Wahrscheinlichkeitsfunktion.

- **Eigenschaften:** $0 \leq f(x_i) \leq 1$ und $\sum_{i=1}^{n} f(x\ ) = 1$

Z.B. Würfel:



- **Verteilungsfunktion einer diskreten Zufallsvariablen:**

  Die Wahrscheinlichkeitsfunktion $F(x)$, welche die Wahrscheinlichkeit angibt, dass die Zufallsvariable X einen Wert kleiner oder

gleich x_i annimmt, $P(X \leq x_i)$, heißt Verteilungsfunktion:

$$F(x) = \sum_{i=1}^{x_i} f(x_i)$$

- **Eigenschaften:**

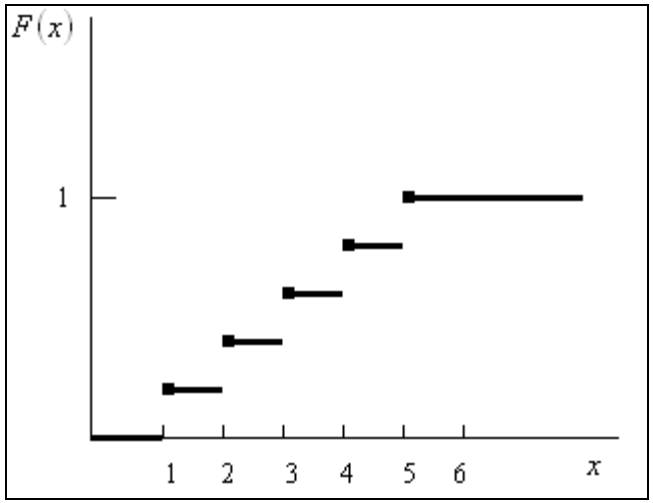
  (1) $F(x)$ ist monoton steigend:

  wenn $x_2 > x_1$ ist auch immer $F(x_2) > F(x_1)$

  (2) $F(x)$ ist rechtsseitig stetig

  (3) $\lim_{x \to -\infty} F(x) = 0$ und $\lim_{x \to +\infty} F(x) = 1$

Z.B. Würfel:



zu **(2)** rechtsseitig stetig: z.B. man nähert sich der 2 von rechts:

$$\lim_{x \to 2^+} = F(2) = \frac{2}{6}$$

man nähert sich der 2 von links:

$$\lim_{x \to 2^-} = F(1) = \frac{1}{6}$$

- **Wahrscheinlichkeitsintervall einer diskreten Zufallsvariable:**

$$P(x_1 < X \leq x_2) = P(X \leq x_2) - P(X \leq x_1) = F(x_2) - F(x_1)$$

1.2 Stetige Zufallsvariable

- **Dichtefunktion einer stetigen Zufallsvariable:**

  für den Fall einer überabzählbar unendlichen Zufallsvariable gilt für die Dichtefunktion:

  (1) $\int_{-\infty}^{+\infty} f(x)\,dx = 1$ und **(2)** $f(x) \geq 0$

- Die Dichtefunktion kann nur in einem unendlichen Intervall $[-\infty;+\infty]$ oder endlichen Intervall $[x_u; x_o]$ positive Werte annehmen (d.h. die Punktwahrscheinlichkeit ist 0), z.B.: $f(x) = \begin{cases} 0{,}5 & , 3 < x < 5 \\ 0 & , sonst \end{cases}$

- **Verteilungsfunktion einer stetigen Zufallsvariable:**

  Die Verteilungsfunktion stellt die Fläche unter der Dichtefunktion bis zur Stelle *x* dar: $F(x) = \int_{-\infty}^{x} f(u)du$

- **Eigenschaften:**

  (1) für $x < x_u$: $F(x) = 0$

  (2) für $x_u \leq x < x_o$: $F(x) = \int_{x_u}^{x} f(u)du$

  (3) für $x \geq x_o$: $F(x) = 1$

Die Dichtefunktion ist die erste Ableitung der Verteilungsfunktion, d.h.

$$\frac{dF(x)}{dx} = f(x)$$

- **Wahrscheinlichkeitsintervall einer stetigen Zufallsvariable:**

$$P(a < x < b) = \int_a^b f(x)dx = F(b) - F(a)$$

1.3 Erwartungswerte von Zufallsvariablen

- **Erwartungswert:**

  Der Erwartungswert der Zufallsvariable *X*, $E(x)$ oder auch μ, ist wie folgt definiert:

im diskreten Fall:

$$E(x) = \sum_{i=1}^{n} x_i \cdot f(x_i)$$

im stetigen Fall:

$$E(x) = \int_{-\infty}^{+\infty} x \cdot f(x) dx$$

1.4 Varianzen von Zufallsvariablen

- **Varianz und Standardabweichung:**

  Die Varianz der Zufallsvariable X, $Var(x)$ oder auch σ^2, ist wie folgt definiert:

  im diskreten Fall: $$\sigma^2 = \sum_{i=1}^{n} (x_i - E(X))^2 \cdot f(x_i) = \sum_{i=1}^{n} x_i^2 f(x_i) - E(X)^2$$

  Oft lässt sich die Varianz mit der zweiten Version dieser Gleichung einfacher berechnen.

im stetigen Fall: $$\sigma^2 = \int_{-\infty}^{+\infty} (x - E(X))^2 \cdot f(x) dx$$

Auch für diese Gleichung ist folgende Umformung oft nützlich:

$$\sigma^2 = \int_{-\infty}^{+\infty} x^2 \cdot f(x) dx - E(X)^2$$

- Die Standardabweichung ist in beiden Fällen:

$$\sigma = \sqrt{Var(X)} = \sqrt{\sigma^2}$$

1.5 Momenterzeugende Funktionen

- **Moment um Null**:

  Das *n*-te Moment um Null einer Zufallsvariable ist definiert als:

  diskreter Fall: $$E\left(X^n\right)=\sum_{i=1}^{n} x_i^n f\left(x_i\right)$$

  stetiger Fall: $$E\left(X^n\right)=\int_{-\infty}^{+\infty} x^n f\left(x\right)dx$$

- Beachte den Unterschied: $E\left(x^2\right)$ und $E\left(x\right)^2$ beim Berechnen der Varianz:

$$E\left(x^2\right)=\int_{-\infty}^{+\infty} x^2 \cdot f\left(x\right)dx$$ im Vergleich zu $$E\left(x\right)^2=\left(\int_{-\infty}^{+\infty} x \cdot f\left(x\right)dx\right)^2$$

Der erste Term ist also das zweite Moment um Null, während der zweite Term das erste Moment um Null (= Erwartungswert) zum Quadrat ist.

- **Zentrales Moment *n*-ter Ordnung:**

diskreter Fall: $$E\left(\left(X-E\left(X\right)\right)^n\right)=\sum_{i=1}^{n}\left(x_i-E\left(X\right)\right)^n \cdot f\left(x_i\right)$$

stetiger Fall: $$E\left(\left(X - E(X)\right)^n\right) = \int_{-\infty}^{+\infty} \left(x - E(X)\right)^n \cdot f(x)dx$$

Die Varianz ist also das zentrale Moment zweiter Ordnung.

1.6 Standardisieren

Eine Zufallsvariable X mit Erwartungswert μ und Standardabweichung σ lässt sich zur standardisierten Zufallsvariable Z transformieren: $Z = \frac{X - \mu}{\sigma}$

Eigenschaften:

$$E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}E(X) - \frac{1}{\sigma}E(\mu) = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0$$ und

$$Var(Z) = Var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}Var(X) = \frac{\sigma^2}{\sigma^2} = 1$$

Eine Standardisierte Zufallsvariable hat also einen Erwartungswert von 0 und eine Varianz von 1.

1.7 Tschebyscheffsche Ungleichung

Mit der Ungleichung von Tschebyscheff kann die Wahrscheinlichkeit geschätzt werden, dass eine Zufallsvariable in einem symmetrisch um den Erwartungswert gelegenen Intervall liegt. Je nach Verteilung wird die Wahrscheinlichkeit sehr ungenau. Vorteil ist aber, dass die Verteilung nicht bekannt sein muss:

$P(\mu - c \cdot \sigma < X < \mu + c \cdot \sigma)$ oder auch $P(|X - \mu| < c \cdot \sigma)$

Die Wahrscheinlichkeit, dass sich X um <u>wenigstens</u> $c\sigma$ von μ unterscheidet, lässt sich abschätzen als: $P\left(\left|X-\mu\right| \geq c\cdot\sigma\right) \leq \frac{1}{c^2}$.

Die Wahrscheinlichkeit, dass sich X um <u>höchstens</u> $c\sigma$ von μ unterscheidet, lässt sich abschätzen als: $P\left(\left|X-\mu\right| < c\cdot\sigma\right) > 1-\frac{1}{c^2}$.

1.8 Schiefe und Wölbung

Die Schiefe einer Verteilung zeigt an, wie stark die Verteilung nach rechts (negative Schiefe) oder nach links (positive Schiefe) geneigt ist. Die Wölbung zeigt an, wie hoch die Wahrscheinlichkeiten um den Mittelwert und die extremen Enden der Verteilung liegen.

- **Schiefe:** $\frac{E\left(\left(X - E(X)\right)^3\right)}{\sigma^3}$,

  d.h. die Schiefe ist das zentrale Moment dritter Ordnung bezogen auf die Standardabweichung in der dritten Potenz.

- **Kurtosis:** $\frac{E\left(\left(X - E(X)\right)^4\right)}{\sigma^4}$,

  d.h. die Kurtosis ist das zentrale Moment vierter Ordnung bezogen auf die Standardabweichung in der vierten Potenz.

  Eine Kurtosis >3 wird auch als Excess Kurtosis bezeichnet.

1.9 Median, Quantile und Modus

- **Median:**

  Der Median oder Zentralwert x_{Med} einer Zufallsvariablen X ist diejenige Zahl, für welche die Wahrscheinlichkeit, dass X grösser oder kleiner als x_{Med} ist, gleich groß ist: $P(X \leq x_{Med}) = F(x_{Med}) = 0{,}5$

  oder: $x_{Med} = F^{-1}(0{,}5)$

  Da sich der Median für viele diskrete Zufallsvariablen nicht eindeutig bestimmen lässt schreibt man auch:

und

$$\left.\begin{array}{l} P(X \le x_{Med}) \ge 0{,}5 \\ P((X \ge x_{Med}) \ge 0{,}5 \end{array}\right\}$$

- **Quantil:**

  Eine Zahl $x[q]$ mit $0 < q < 1$ heisst *q*-Quantil wenn gleichzeitig gilt:

und

$$\left.\begin{array}{l} P(X \le x[q]) \ge q \\ P((X \ge x[q]) \ge 1-q \end{array}\right\}$$

Der Median ist also das 0,5- oder 50%-Quantil und das *q*-Quantil eine Verallgemeinerung.

- **Modus:**

  der Modus einer Verteilung ist diejenige Zahl x_M , die im diskreten Fall die höchste Eintrittswahrscheinlichkeit oder im stetigen Fall die höchste Wahrscheinlichkeitsdichte aufweist:

$$f(x_M) > f(x) \text{ für alle } x \neq x_M$$