Statistik Grundlagen

Definition einer Zufallsvariable X
Formal:
Zurodnungsvorschrift, die jedem möglichen Ergebnis eine Zahl zuordnet.
Praktisch:
X ordnet dem Ausgang eines Zufallsexperiments einen Zahlenwert zu.
Klassifizierung von Zufallsvariablen

- Binäre Zufallsgröße: Wahr/Falsch (0/1)
- Qualitative Größen: Haarfarbe
- Quantitavie Größen:
  - diskrete Zufallsgröße (Zählraten)
  - kontinuierlich Zufallsgröße (Messreihen, z.B. Winkelverteilungen)

Statistik Grundlagen

Bsp:

Von 8 Personen werden 3 Merkmale erfasst

- a=Geschlecht (0=w, 1=m) binär
- b=Alter (in Jahren) quantitativ
- c= Haarfarbe (0=schwarz, 1=blond, ...) qualitativ

Was können wir im Bezug auf Zufallsgrößen aussagen?

In den meisten Fällen interessieren wir uns für Wahrscheinlichkeiten, dass

- ein Ereignis E eintritt
- eine Zufallsgröße einen bestimmten Wert hat
- Messungen in einem definierten Intervall liegen

Was ist ein Ereignis??
X ist das numerische Resultat eines Zufallsexperiments. Die Menge aller möglichen Werte dieser Variable heißt Ereignismenge (Stichprobenraum), häufig Ω abgekürzt.

Beispiel Ereignisse

- Roulette: mögliche Ergebnisse sind die Zahlen von 0 bis 36

$$\Omega = \{0, \dots, 36\}$$

- Dauer zwischen zwei radioaktiven Zerfällen

$$\Omega = \mathbb{R}_+$$

Ein Ereignis E ist eine Teilmenge von Ω. Ist $X \in \Omega$ dann tritt das Ereignis ein wenn $X \in E$.
Würfelwurf:

$$\Omega = \{0, 1, 2, \dots, 6\} \tag{1}$$

$$\text{Ereignis gerade Zahl } E = \{2, 4, 6\} \tag{2}$$

Rechenoperationen mit Ereignissen/Mengen

$$A \cup B \quad \text{Vereinigung} \quad A \text{ oder } B \tag{3}$$

$$A \cap B \quad \text{Durchschnitt} \quad A \text{ und } B \tag{4}$$

$$A' \text{ oder } \overline{A} \quad \text{Komplement} \quad \text{nicht } A \tag{5}$$

$$A \subseteq B \quad \text{Teilmenge} \quad \text{aus A folgt B } (\overline{A} \cup B) \tag{6}$$

Ereignisse lassen sich hierüber darstellen

$$E = A \cap B \cap C \tag{7}$$

E tritt genau dann ein, wenn A, B und C eintreten.
Besonders sind disjunkte Ereignisse

$$A \cap B = \emptyset \tag{8}$$

Wahrscheinlichkeit

Definition nach Kolmogorov:
Sei Σ die Ereignisalgebra über Ω, A und B Ereignisse in Σ und p eine Abb. von Σ nach $\mathbb{R}$. p heißt Wahrscheinlichkeit, wenn

- $p(a) \geq 0$, für Alle $A \in \Sigma$
- $A \cap B = \emptyset \Rightarrow p(A \cup B) = p(A) + p(B)$, für Alle $A, B \in \Sigma$
- Normierung $p(\Omega) = 1$
- Enthält Σ unendlich viele Ereignisse verlangt man zusätzlich

$$p(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} p(A_n)$$

, mit $A_n, n \in \mathbb{N}$ paarweise disjunkt

Weiterhin definiert man die bedingte Wahrscheinlichkeit

$$p(A|B) = \frac{p(A \cap B)}{p(B)} \tag{9}$$

d.h. Wahrscheinlichkeit für A wenn B eingetreten ist.
Rechenregeln

$$p(\emptyset) = 0 \tag{10}$$
$$p(A) \leq 1 \ \forall A \tag{11}$$
$$p(\overline{A}) = 1 - p(A) \tag{12}$$
$$A \subset B \Rightarrow p(A) \leq p(B) \tag{13}$$
$$p(A \cup B) = p(A) + p(B) - p(A \cap B) \tag{14}$$

Frequentistische Interpretation

Klassische Interpretation der Wahrscheinlichkeit:

$$p(E) = \lim_{N \to \infty} \frac{n(E)}{N} \tag{15}$$

als Häufigkeit eines Ereignisse bei einer großen Zahl an unabhängigen Wiederholungen des Experiments.
Häufig meinen wir diese Interpretation in der Physik.

Streng genommen existiert der Grenzwert nicht.

Bayesische Wahrscheinlichkeitsinterpretation

Seien $E_1, E_2, \dots, E_n$ Erieginisse und zerlgen Ω in folgender Weise:

- $E_i \cap E_j = \emptyset \forall i \neq j$
- $E_1 \cup E_2 \cup \dots \cup E_n = \Omega$

Dann gilt:

$$p(E_1) + p(E_2) + \dots = p(E_n) = p(\Omega) = 1$$

Weiterhin

$$p(A) = p(A|E_1)p(E_1) + \dots + p(A|E_n)p(E_n)$$

Dann gilt

$$p(E_i|A) = \frac{p(A|E_i)p(E_i)}{p(A)} \tag{16}$$

Beispiel

Sie stehen vor Gericht und werden der Fahrerflucht beschuldigt. Der Kronzeuge sagt aus es hätte sich ein blaues Auto vom Tatort entfernt (Ihres ist Blau).
Ihr Anwalt verteidigt Sie, indem er anführt, dass bei den zur Tatzeit vorliegenden Sichtverhältnissen die Farben der Autos in 10 % der Fälle falsch identifiziert werden.
Weiterhin gibt es in Ihrer Stadt 100 rote Autos und 10 blaue. Wie großist die Wahrscheinlichkeit, dass Sie zu unrecht beschuldigt werden?

Nach Bayes

$$p(f|b) = \frac{p(b|f)p(f)}{p(b)} \tag{17}$$

$$p(b) = \frac{19}{110} \tag{18}$$

$$p(f) = \frac{11}{110} = 0.1 \tag{19}$$

$$p(b|f) = \frac{10}{11} \tag{20}$$

$$p(f|b) = \frac{\frac{10}{11}\frac{11}{110}}{\frac{19}{110}} = \frac{10}{19} \approx 53\% \tag{21}$$

Sie haben ein gutes Argument gegen den Zeugen!

Bayesische Interpreation

Die Wahrscheinlichkeit eines Ereignisses entspricht dem Grad der Überzeugung, dass das Ereignis E eintritt, basierend auf Annahmen und Vorwissen (bedingte Wahrscheinlichkeiten).

Diese Interpretation ist anwendbar

- Einmalige Vorkomnisse, bzw. schwer reproduzierbare Ereignisse
- Katastrophenvorhersagen

Verteilungsfunktionen

Zu Wahrscheinlichkeit kann man Wahrscheinlichkeitsdichtefunktionen angeben mit

$$p_X(E) = \int_E f_X(x')dx' \tag{22}$$

wobei f_X die Dichtefunktion zu X ist. Wir unterscheiden diskrete Dichtefunktionen

$$f(n) \geq 0, \qquad \sum_n f(n) = 1 \tag{23}$$

$$p(n_1 \leq n \leq n_2) = \sum_{n_1}^{n_2} f(n) \tag{24}$$

Verteilungsfunktionen

Und kontinuierlich Dichtefunktionen

$$f(x) \geq 0, \qquad \int_{-\infty}^{\infty} f(x)dx = 1 \tag{25}$$

$$p(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} f(x)dx \tag{26}$$

Elementarereignisse haben Wahrscheinlichkeit Null. Die Wahrscheinlichkeit, dass eine Zufallsvariable unterhalb eines Wertes liegt wird kumulative Verteilungsfunktion genannt

$$p(x \leq y) = \int_{-\infty}^{y} f_X(x')dx' = F_X(y) \tag{27}$$

Eigenschaften der Verteilungsfunktionen

$$p(x_1 \leq x \leq x_2) = F_X(x2) - F_X(x_1) \tag{28}$$

Weiterhin gilt:

- $0 \leq F_X(x) \leq 1$ für ale $x \in \mathbb{R}$
- $x_1 \leq x_2 \Rightarrow F_X(x_1) \leq F_X(x_2)$
- $\lim_{x \to -\infty} F_X(x) = 0 \lim_{x \to \infty} F_X(x) = 1$

Definition der Quantilfunktion

$$F_X(x_\alpha) = \alpha \qquad \text{mit } 0 < \alpha < 1 \tag{29}$$

heißt das α-Quantil d. Verteilung. Die Funktion

$$x = F_X^{-1}(\alpha)$$

heißt Quantilfunktion.

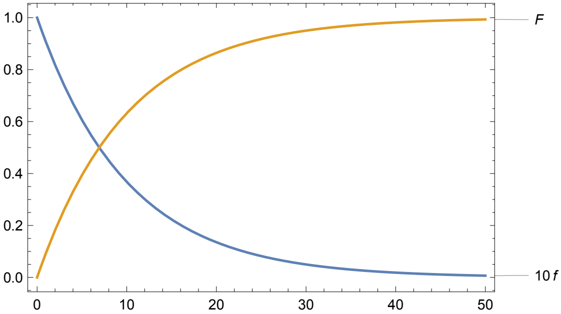
- Die Quantile zu den Niveaus $\alpha = 0.25, 0.6, 0.75$ heißen Quartile.
- Das Quantil zu $\alpha = 0.5$ heißt Median.

Beispiel radioaktiver Zerfall:

$$f(t) = \frac{1}{\tau} e^{-t/\tau} \tag{30}$$

$$F(t) = 1 - e^{-t/\tau} \tag{31}$$

f ist die Zerfallswahrscheinlichkeitsdichte und F die Zerfallswahrscheinlichtkeit.



$\tau = 10$ Median

$$t = \tau \ln 2$$

Erwartungswert

Definition:
Sei X eine Zufallsvariable mit Dichtefunktion $f_X(x)$. Weiter sei g(x) beliebige stetige Fkt.
Man definiert die Erwartung:

$$E[g(x)] = \int_{\mathbb{R}} g(x')f_X(x')dx' \tag{32}$$

Für $g(x) = x$ entspricht E dem Erwartungswert von x

$$E[x] = \overline{x} = \int x \; f_X(x)dx \tag{33}$$

Die Standardabweichung oder auch Varianz ist definiert als

$$\sigma^2 = \overline{(x - \overline{x})^2} \tag{34}$$

Das entspricht dem Mittelwert der quadratischen Abweichung zum Mittelwert

$$\sigma^2 = \int\limits_{-\infty}^{\infty} (x - \overline{x})^2 f(x) dx \tag{35}$$

$$= \int\limits_{-\infty}^{\infty} \left(x^2 - 2x\overline{x} + \overline{x}^2\right) f(x) dx \tag{36}$$

$$= \int\limits_{-\infty}^{\infty} x^2 f(x) dx - 2\overline{x} \int\limits_{-\infty}^{\infty} x f(x) dx + \overline{x}^2 \underbrace{\int\limits_{-\infty}^{\infty} f(x) dx}_{1} \tag{37}$$

$$= \overline{x^2} - 2\overline{x}\overline{x} + \overline{x}^2 = \overline{x^2} - \overline{x}^2 \tag{38}$$

Für diskrete Verteilungen ist die Varianz gegebene zu

$$\sigma^2 = \frac{1}{N}\left(\sum x^2 - \frac{\left(\sum x\right)^2}{N}\right) \tag{39}$$

Achtung:
Hier ist die Defintion der Varianz einer Variablen x. Um die Varianz aus einer Stichprobe abzuschätzen benötigt man eine erwartungstreue Schätzfunktion.

Momente, Schiefe und Wölbung einer Verteilung

Momente sind die Erwartungswerte von x^n, und $(x - \langle x \rangle)^n$, mit

- $\langle x^2 \rangle \equiv$ n-te Moment M_n
- $(x - \langle x \rangle)^n \equiv$ n-tes Zentralmoment μ_n

Schiefe:

$$v(x) = \frac{\mu_3}{\sigma^3} = \frac{\langle (x - \langle x \rangle)^3 \rangle}{\sigma^3} \tag{40}$$

Wölbung:

$$\beta_2 = \frac{\mu_4}{\sigma^4} \tag{41}$$

$$\gamma_2 = \beta_2 - 3 \tag{42}$$