

Statistik Theorie

Definitionen

Ω = Grundmenge = Ergebnismenge = Menge aller möglichen Ergebnisse

\mathcal{A} = Ereignisraum = σ -Algebra (Sigma-Algebra) = Menge aller messbaren Ergebnisse über eine definierte Grundmenge Ω

$\mathcal{P}(\Omega)$ = Potenzmenge von Ω (ist die Menge aller Teilmengen der Grundmenge)

$P(A)$ = Wahrscheinlichkeit eines Ereignisses

Axiome der Stochastik

1. $P(\Omega) = 1$

= Wahrscheinlichkeit des Ereignisses, das alle möglichen Versuchsausgänge umfasst ist 1

2. $P(\emptyset) = 0$

= Wahrscheinlichkeit, dass ein unmögliches Ereignis eintritt ist 0

3. $0 \leq P(A) \leq 1$

= alle Wahrscheinlichkeiten liegen zwischen einschließlich 0 und 1

4. $P(A) + P(\bar{A}) = 1$

= Die Wahrscheinlichkeit des Eintretens und des Nichteintretens addieren sich zu 1

5. $\sum_{i=1}^n P(A_i) = 1$

= In einem vollständigen System von Ereignissen A_i ist die Summe der Wahrscheinlichkeiten 1

Lageparameter

Mittelwert (Moment 1. Ordnung)

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$$

Empirische Varianz = Varianz für Stichproben (Moment 2. Ordnung)

Gibt an, wie sehr eine Zufallsvariable X von ihrem Erwartungswert $E(X)$ abweicht.

$$\sigma = s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2$$

$n - 1$ sind die Freiheitsgrade (= Anzahl der frei wählbaren Parameter). Das -1 fällt weg, wenn die Varianz der Grundgesamtheit berechnet wird.

Hat immer positive Werte, ist zum Rechnen leichter aber nicht sehr anschaulich.

Standardabweichung für Stichproben

Ist die Wurzel aus der Varianz, hat auch negative Werte. Ist schlechter zum Rechnen aber anschaulicher.

$$s_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2}$$

Schiefe (Moment 3. Ordnung)

$$g_1 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^3$$

Kurtosis (Moment 4. Ordnung)

$$g_2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^4$$

Freiheitsgrade für Stichproben und Grundgesamtheit gelten genauso wie bei der Varianz.

Wahrscheinlichkeiten

Messraum

Gibt einen messbaren Raum an, das heißt er definiert über welche Mengen eine Aussage getroffen werden kann.

Ein Tupel (Ω, \mathcal{A}) ist ein Messraum, wenn

- Ω Omega eine beliebige Grundmenge ist
- \mathcal{A} eine σ -Algebra ist

A ist eine messbare Menge wenn $A \in \mathcal{A}$ ist.

Wahrscheinlichkeitsraum

Ist ein Messraum der um ein Wahrscheinlichkeitsmaß P erweitert wird und damit aus drei Tupeln besteht:

W-Raum: (Ω, \mathcal{A}, P)

Das Wahrscheinlichkeitsmaß P ist dabei eine Funktion die jedem Ereignis eine Zahl zwischen 0 und 1 zuordnet.

Diese Funktion muss drei Eigenschaften erfüllen:

1. $P(A) \geq 0$ für alle $A \in \mathcal{A}$
= Die Wahrscheinlichkeit jedes Ereignisses in der σ -Algebra muss größer oder gleich 0 sein
2. $P(\Omega) = 1$
= Wahrscheinlichkeit des Ereignisses, das alle möglichen Versuchsausgänge umfasst ist 1
3. Es gilt für eine Folge $A_1, A_2 \dots$ von paarweise disjunkten Ereignissen die σ -Additivität:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

(paarweise disjunkt = $A_i \cap A_j = \emptyset$ für $i \neq j$ = die Menge haben keine gemeinsamen Werte)

Allgemeiner Additionssatz für Wahrscheinlichkeiten

$$P(A \cup B) = P(A) + P(B) + P(A \cap B)$$

Bedingte Wahrscheinlichkeiten

Die Wahrscheinlichkeit von A wenn bekannt ist, dass B bereits eingetreten ist:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Multiplikationstheorem

Erhält man durch Multiplizieren der bedingten Wahrscheinlichkeit mit $P(B)$:

$$P(A \cap B) = (A|B) P(B)$$

Durch Vertauschen von A und B ergibt sich:

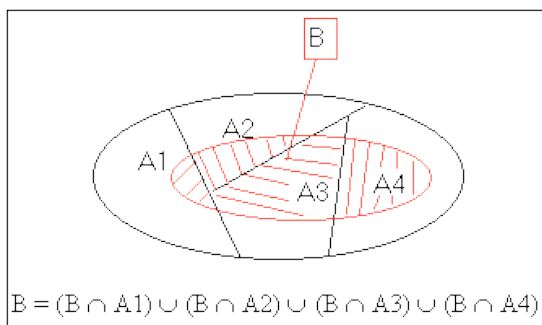
$$P(A \cap B) = (B|A) P(A)$$

Satz der vollständigen Wahrscheinlichkeiten

Wenn nur bedingte Wahrscheinlichkeiten für ein Ereignis bekannt sind, lässt sich daraus die totale Wahrscheinlichkeit als Summe aller bedingten Wahrscheinlichkeiten errechnen.

$$P(A) = \sum_{i=1}^{\infty} P(A|C_i)P(C_i)$$

Als Venn-Diagramm:



Zufallsvariable = Stochastische Größe

Wird als sG oder als X bezeichnet. Die Variable kann jeden beliebigen Wert annehmen. Der angenommene Wert ist vom Zufall abhängig und hängt von der jeweiligen Wahrscheinlichkeit ab.

Definition:

Eine Abbildung X von Ω nach \mathbb{R} , die jedem $\omega \in \Omega$ eine reelle Zahl $X(\omega) = x$ zuordnet.

Der Wertebereich / Merkmalsraum von X lautet $M_X = \{x \mid x = X(\omega), \omega \in \Omega\}$

Erwartungswert einer Zufallsvariable

Wird bezeichnet mit $E(X)$

Ist die Zahl, die eine Zufallsvariable im Mittel annimmt. Wenn man ein Experiment unendlich oft wiederholt, ist der Erwartungswert der Durchschnitt aller Ergebnisse.

Gesetz der großen Zahlen

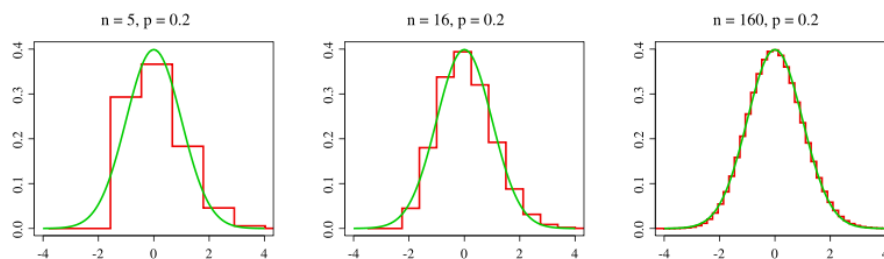
$$\lim_{n \rightarrow \infty} h_n(A) = : P(A)$$

Bedeutung: Bei n Wiederholungen eines Zufallsexperiments tritt das fragliche Ereignis $H_n(A)$ -mal auf, so dass sich die relative Häufigkeit von A einem Grenzwert, der theoretischen Wahrscheinlichkeit annähert.

Zentraler Grenzwertsatz

Besitzt die Verteilung der Grundgesamtheit eine endliche Varianz, so ist die Verteilung der arithmetischen Mittel von Zufallsstichproben approximativ normal, sofern der Stichprobenumfang genügend groß ist.

Bedeutung: Die Summe einer großen Zahl von unabhängigen Zufallsvariablen nähert sich einer stabilen Verteilung. Die Verteilungen konvergieren, bei sehr vielen n der Normalverteilung an.



Verteilungsfunktion

Ist eine reellwertige Funktion F_X , mit der man die Wahrscheinlichkeitsverteilung einer reellen Zufallsvariable X beschreibt.

Definition:

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

Die Wahrscheinlichkeit des Ereignisses $P(X \leq x)$ liegt immer zwischen 0 und 1.

X ist die Zufallsvariable = stochastische Größe

Die Verteilungsfunktion ist eine Funktion, mit der man die Wahrscheinlichkeitsverteilung einer reellen Zufallsvariable beschreibt.

Verteilungsfunktionen haben vier Eigenschaften:

1. $0 \leq F(x) \leq 1$ für $x \in \mathbb{R}$

= alle Werte der Verteilungsfunktion liegen zwischen 0 und 1

2. Aus $x < y$ folgt $F(x) \leq F(y)$

= Die Funktion ist monoton wachsend, wird also immer nur größer oder bleibt gleich, fällt aber nie.

$$3. \quad \lim_{x \rightarrow -\infty} F(x) = 0 \text{ und } \lim_{x \rightarrow \infty} F(x) = 1$$

= Die Funktion konvergiert bei minus Unendlich gegen 0 und bei plus Unendlich gegen 1

$$4. \quad \lim_{h \downarrow 0} F(x+h) = F(x) \quad \text{für } x \in \mathbb{R}$$

= Die Funktion ist rechtsstetig

Empirische Verteilungsfunktion

Ist die Verteilungsfunktion einer Stichprobe.

Definition:

$$\tilde{F}_n(x) = \begin{cases} 0 & \text{für } x < x_{(1)} \\ \frac{i}{n} & \text{für } x_{(i)} \leq x_{(i+1)}, \quad i = 1, 2, \dots, n-1 \\ 1 & \text{für } x_{(n)} < x_{(1)} \leq x \end{cases}$$

$\tilde{F}_n(x)$ ist eine Treppenfunktion mit Sprüngen an den Stellen $x_{(i)}$ der Höhe $\frac{1}{n}$.

Die Funktion ordnet jeder reellen Zahl x den Anteil der Stichprobenwerte zu, die kleiner oder gleich x sind.

Dichtefunktion

ist eine Funktion um eine stetige Wahrscheinlichkeitsverteilung zu beschreiben.

Eine stochastische Größe X hat eine stetige Verteilung wenn die Verteilungsfunktion $F_X(x)$ eine stetige Funktion auf \mathbb{R} ist. Die meisten stetigen stochastischen Größen sind absolut stetig, d. h. es gibt eine Funktion f_x für die gilt:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t) dt$$

Eine solche Funktion f_x nennt man Dichtefunktion von X .

Eigenschaften der Dichtefunktion

$$1. \quad f_X(x) \geq 0, \quad x \in \mathbb{R}$$

= Die Werte der Dichtefunktion sind immer größer oder gleich 0 (Wahrscheinlichkeiten)

$$2. \quad \int_{-\infty}^{\infty} f_X(t) dt = 1$$

= Die Fläche unterhalb der kompletten Dichtefunktion ist immer 1 (komplette Wahrscheinlichkeit)

Umrechnen zwischen Dichte- und Verteilungsfunktion

$$\int f_X(x) = F_X(x)$$

und

$$F'_X(x) = f_X(x)$$

Normalverteilung

Wird durch zwei Parameter charakterisiert:

1. Der Lageparameter $\mu \in \mathbb{R}$ (Mittelwert = Erwartungswert = das Zentrum der Kurve)
2. Der Skalierungsparameter $\sigma > 0$ (Standardabweichung = wie weit die Werte der Kurve streuen)

Schreibweise für die Normalverteilung: $X \sim N(\mu, \sigma^2)$.

Die Normalverteilung ist immer symmetrisch und nie schief, die Wölbung ist immer 3. Die Kurve kann unterschiedlich breit sein (hängt vom Parameter σ ab).

Exponentialverteilung

Die Exponentialverteilung wird durch den Skalierungsparameter $\tau > 0$ (Tau) definiert und in erster Linie für Zeitintervalle verwendet. Der Parameter $\lambda = 1/\tau$ gibt dabei die Zahl der erwarteten Ereignisse pro Einheitsintervall an.

Schreibweise für die Exponentialverteilung: $X \sim \text{Exp}(\lambda)$.

Dichte der Exponentialverteilung: $f(x) = \lambda e^{-\lambda x}$ für $x \geq 0$

Verteilungsfunktion der Exponentialverteilung: $F(x) = 1 - e^{-\lambda x}$ für $x \geq 0$

Erwartungswert $\mathbb{E}(X) = 1/\lambda$ und Varianz $\text{Var}(X) = 1/\lambda^2$

Binomialverteilung

Beschreibt die Anzahl der Erfolge in einer Serie von gleichartigen und unabhängigen Versuchen mit genau zwei möglichen Ausgängen.

Bernoulli-Experiment

= ein Experiment bei dem es nur zwei mögliche Ausgänge gibt.

n = Anzahl der Experimente, x = Anzahl der Erfolge und $n - x$ = Anzahl der Misserfolge

$\binom{n}{x} = \frac{n!}{x!(n-x)!}$ gibt die möglichen Positionen für die x Erfolge an.

Wahrscheinlichkeit für jede dieser Möglichkeiten ist binomialverteilt:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{für } x \in \{0, 1, 2, \dots, n\}$$

Schreibweise für die Binomialverteilung: $X \sim B(n, p)$.

Erwartungswert und Varianz: $\mathbb{E}(X) = np$ und Varianz $\text{Var}(X) = np(1-p)$

Poissonverteilung

Bei großen n lässt sich die Binomialverteilung gut durch die Poissonverteilung annähern: $\lambda = np$

$$p(x) = P(X = x) = \sum_{i=0}^{\infty} \frac{\lambda^i e^{-\lambda}}{x!} \quad \text{für } x \in \{0, 1, 2, \dots\}$$

Schreibweise für Poisson-Verteilung: $\mathbf{X} \sim \mathbf{P}(\lambda)$

Korrelationskoeffizient

Misst den Grad der linearen Assoziation zwischen zwei Merkmalen und gibt einen Wert zwischen -1 (vollständiger, negativer linearer Zusammenhang) und 1 (vollständiger, positiver linearer Zusammenhang) aus. Bei einem Wert um 0 herum sind die Stichproben unkorreliert.

Kompliziertere Assoziationen werden über den Korrelationskoeffizient nicht ausreichend erfasst, Korrelation sagt auch nichts über Kausalität aus.

Die Berechnung erfolgt durch zwei Standardabweichungen s_x und s_y von zwei Stichproben x und y .

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

Kovarianz

Misst den Grad des monotonen Zusammenhangs zwischen zwei Zufallsvariablen.

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Über die Kovarianz lässt sich der Korrelationskoeffizient auch berechnen:

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

Bivariat normalverteilt

Bei zwei Stichproben die bivariat Normalverteilt sind, besitzt Y eine Normalverteilung bei einem fixen Wert von X und umgekehrt.

Die Abhängigkeit zwischen X und Y wird durch die Kovarianz charakterisiert.

Statistische Tests

liefern nach bestimmten Regeln eine Entscheidung darüber, ob eine vorgegebene Hypothese über die untersuchte Grundgesamtheit anhand von Daten aus einer Stichprobe verworfen werden soll oder nicht.

H_0 = Nullhypothese (davon wird ausgegangen solange das Gegenteil nicht bewiesen wird)

H_1 = Alternativhypothese (soll über den Test bewiesen werden)

α = Signifikanzniveau (gibt Aussagekraft des Testergebnisses an)

Fehler

1. Fehler 1. Art = α -Fehler

Die Hypothese ist richtig und wird fälschlicherweise verworfen (die Wahrscheinlichkeit für so einen Fehler gibt das Signifikanzniveau α an)

2. Fehler 2. Art = β -Fehler

Die Hypothese ist falsch und wird trotzdem nicht verworfen.

Chiquadrat-Anpassungstest

Dient zur Überprüfung der Form einer Verteilung (zb. ob eine Stichprobe normalverteilt ist). Dabei wird eine Klasseneinteilung aller Werte getroffen und die empirischen (gemessenen) Häufigkeiten werden mit den theoretischen Werten verglichen.

Wenn die Werte zu weit voneinander abweichen, wird die Hypothese verworfen, sonst angenommen. Berechnen der Prüfgröße für den Test:

$$\chi^2 = \sum_{i=1}^k \frac{(X_i - np_i)^2}{np_i}$$

k = die Anzahl der Klassen

X_i = die absolute Häufigkeit (die Anzahl der Daten in der jeweiligen Klasse)

np_i = die erwartete relative Häufigkeit der Werte einer Klasse laut Hypothese

Konfidenzintervalle

Gibt die Präzision eines Schätzers mittels einer Überdeckungswahrscheinlichkeit an.

Die Präzision wird mit einer Wahrscheinlichkeit $1 - \alpha$ (Konfidenzkoeffizient) angegeben. Das Intervall ist der Wertebereich, welcher den wahren Wert des unbekannten Parameters θ (Theta) mit der vorgegebenen Wahrscheinlichkeit überdeckt.

$$P_{\theta}(T_1(X) < \theta < T_2(X)) \geq 1 - \alpha \quad \text{für alle } \theta \in \Theta$$

= Überdeckungswahrscheinlichkeit des Parameters θ zwischen dem Intervall T_1 und T_2 ist größer oder gleich $1 - \alpha$.

Beispiel: 95%iges Konfidenzintervall bedeutet, dass der interessierte Parameter mit 95%iger Wahrscheinlichkeit vom geschätzten Intervall überdeckt wird.

Schätzer

Wenn man viele Stichproben hat, möchte man die Verteilungsfunktion die dahinter liegt schätzen. Die Theorie dahinter ist definiert durch den Hauptsatz der Statistik.

Hauptsatz der Statistik (Gliwenko-Cantelli-Satz)

Eine empirische (gemessene) Verteilungsfunktion einer eindimensionalen Stichprobe konvergiert mit der Wahrscheinlichkeit 1 gegen die tatsächliche (theoretische) Verteilungsfunktion.

Das Theorem wird durch das Gesetz der großen Zahlen bewiesen.

Gütekriterien von Schätzern

1. Konsistenz: je größer der Stichprobenumfang, desto näher liegt das Ergebnis am wahren Wert

$$\widehat{\theta}_n \xrightarrow{P} \theta \quad \text{wenn } n \rightarrow \infty$$

2. Effizienz: je kleiner die Varianz desto besser die Effizienz

Der Schätzer $\hat{\theta}_1$ ist effizienter als $\hat{\theta}_2$ wenn $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$.

3. Erwartungstreue: gibt es keine Verzerrung (keinen systematischen Fehler) ist der Schätzer asymptotisch erwartungstreu bzw. unverzerrt.

Wenn $n \rightarrow \infty$ dann $\mathbb{E}_\theta(\overline{\theta_n}) \rightarrow \theta$ für alle $\theta \in \Theta$

4. Genauigkeit: wird durch den mittleren quadratischen Fehler angegeben, soll möglichst klein sein
5. Asymptotisch Normalverteilt: Schätzer konvergiert in Verteilung gegen eine normalverteilte stochastische Größe.

Schätzwerte

≠ Schätzer. Schätzwerte sind Parameter, die anhand von Stichproben geschätzt werden (zb. Mittelwert, Varianz,...), werden in Kleinbuchstaben angegeben.

Momentschätzer

= die unbekannten theoretischen Momente der Verteilung sollen den Stichprobenmomenten gleichgesetzt werden damit sie aus dieser Gleichung errechnet werden können.

Maximum Likelihood-Schätzer

Die Werte einer Stichprobe werden als Realisierung eines Experiments betrachtet, dass von einem unabhängigen Parameter abhängt. Bis auf diesen Parameter wird das Experiment eindeutig bestimmt und bekannt angesehen.

Likelihood-Prinzip: „Entscheide dich für jenes Modell, das die Daten mit höchster Wahrscheinlichkeit erzeugt hat, also für das plausibelste Verteilungsmodell.“

Definition der ML-Funktion einer Stichprobe:

$$L(\theta) = \prod_{i=1}^n p(x_i; \theta) \quad \text{für } \theta \in \Theta$$

Der ML-Schätzer von θ ist der Wert aus Θ , der die Funktion $L(\theta)$ maximiert. Damit wird die Wahrscheinlichkeit, die Ergebnisse aus der Stichprobe zu erhalten maximiert und der „plausibelste“ Wert für die vorhandene Stichprobe gefunden.

Eigenschaften des ML-Schätzers:

- Invariant
- Asymptotisch erwartungstreu
- Asymptotisch effizient
- Konsistent
- Asymptotisch normalverteilt