1 Grundbegriffe

Grundgesamtheit, Population
(gesamte Gruppe, Bsp. Schüler einer Schule)

1. Umfang

   - Stichprobe/Teilerhebung
     (nur kleiner Teil der gesamten Gruppe untersucht)

2 Grundlagen: Wahrscheinlichkeit

Wichtiges zur Wahrscheinlichkeitsrechnung

- $P(\cup_{i=1}^{n}) = \sum_{i=1}^{n} P(A_i)$ für paarweise disjunkte Ereignisse $A_1, A_2, ..., A_n \subset \Omega$
- $A \subseteq B \Rightarrow P(A) \leqslant P(B)$
- $P(\overline{A}) = 1 - P(A)$
- $P(A|\overline{B}) = 1 - P(\overline{A}|\overline{B})$
- $P(\overline{B} \cap A) = P(A|\overline{B}) \cdot P(\overline{B})$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ für beliebige A, B $\subset \Omega$
- $P(A \cup B) = P(A) + P(B)$ für disjunkte Ereignisse $A, B \subseteq \Omega$ (Additivität)

2.1 Bedingte Wahrscheinlichkeiten

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

Satz der totalen Wahrscheinlichkeit

$$P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)$$

Wichtiger Spezialfall: Multiplikationssatz

$$P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B})$$

2.2 Stochastische Unabhängigkeit

$$P(A|B) = P(A)\ bzw.\ P(B|A) = P(B)$$

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

2.3 Bedingte Unabhängigkeit

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

2.4 Satz von Bayes

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

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

2.5 Satz von Bayes II

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

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}$$

3 Zusammenhangsmaße diskreter Merkmale

X = Zeilen
Y = Spalten

3.1 Kontigenztafeln

- Absolute Häufigkeit: n = n
- Relative Häufigkeiten: n = 1

Bedingte Häufigkeiten

$$f_{Y|X}(b_1|a_1) = \frac{h_{i1}}{h_{i\cdot}}, ..., f_{Y|X}(b_m|a_i) = \frac{h_{im}}{hi\cdot}$$

$$f_{X|Y}(a_1|b_1) = \frac{h_{1j}}{h_{\cdot j}}, ..., f_{X|Y}(a_k|b_j) = \frac{h_{kj}}{h\cdot j}$$

3.2 Empirische Unabhängigkeit

$$f_{Y|X}(b_j|a_1) = f_Y(b_j|a_2) = ... = f_{Y|X}(b_j|a_k) \forall j = 1, ..., m$$

oder halt:

$$\forall i, j: h_{ij} = \tilde{h}_{ij}$$

3.3 χ^2- und Kontigenzkoeffizient

Berechnung von $\tilde{h}$

ist erwartete Häufigkeit in einer Kontingenztafel

$$\text{mit } \tilde{h}_{ij} = \frac{h_{i\cdot} h_{\cdot j}}{n}$$

χ^2-Koeffizient

$$\chi^2 := \sum_{i=1}^{k}\sum_{j=1}^{m} \frac{(h_{ij} - \tilde{h}_{ij})^2}{\tilde{h}_{ij}} = \sum_{i=1}^{k}\sum_{j=1}^{m} \frac{(h_{ij} - \frac{h_{i\cdot} h_{\cdot j}}{n})^2}{\frac{h_{i\cdot} h_{\cdot j}}{n}}$$

$$= n \sum_i \sum_j \frac{(f_{ij} - f_{i\cdot} f_{\cdot j})^2}{f_{i\cdot} f_{\cdot j}}$$

$$\text{mit } \tilde{h}_{ij} = \frac{h_{i\cdot} h_{\cdot j}}{n}$$

- $\chi^2 \in [0, n(min(k,m) - 1)]$
- $\chi^2 = 0 \Leftrightarrow$ X und Y empirisch unabhängig
- χ^2 groß, starker Zusammenhang
- χ^2 klein, schwacher Zusammenhang

Spezialfall für 2x2-Tafel

a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	

$$\chi^2 = \frac{n(ad - bc)^2}{(a+b)(a+c)(b+d)(c+d)}$$

Kontigenzkoeffizient

$$K := \sqrt{\frac{\chi^2}{n + \chi^2}},\ K \in \left[0, \sqrt{\frac{M-1}{M}}\right],\ M = min\{k, m\}$$

Korrigierter Koeffizient

$$K^* := \frac{K}{\sqrt{(M-1)/M}},\ K^* \in [0,1]$$

4 ZVs, Verteilungen und Häufigkeiten

4.1 Zufallsvariable

$$X : \Omega \to T_X,\ T_X \subseteq \mathbb{R}$$

- ordnet jedem $\omega \in \Omega$ genau ein $x \in T_X$ zu: $X(\omega) = x$
- Mehrere Elementarereignisse können selben Zahlenwert zugeordnet sein
- Ausprägungen/Relisierungen: $x = X(\omega)$ von X

4.2 Diskreter Erwartungswert

$$E(X) := \sum_{x \in T_X} x \cdot P(X = x) = \sum_{\omega \in \Omega} P(\{\omega\})X(\omega)$$

$$= \sum_{x \in T_X} x \cdot f_X(x)$$

4.3 Eigenschaften Erwartungswert

- X=a mit Wahrscheinlichkeit 1 (determinische ZV)

$$E(X) = a$$

- Linearität

$$E(a \cdot X + b \cdot Y) = a \cdot E(X) + b \cdot E(Y)$$

$$E(aX + b) = a \cdot E(X) + b$$

- symmetrisch um Punkt c: $f(c-x) = f(c+x) \forall x \in T_X$, dann $E(X) = c$
- für bel. $a_1, ..., a_n \in \mathbb{R}$ und bel. ZV $X_1, ..., X_n$

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i \cdot E(X_i)$$

4.4 Transformationsregel Erwartungswert

Für reelle Funktion Y=g(X):

$$E(Y) = E[g(X)] = \begin{cases} \sum_{x \in T} g(x) f(x) & \text{X diskret} \\ \int_{-\infty}^{\infty} g(x) f(x) dx & \text{X stetig} \end{cases}$$

4.5 Varianz ZV

anstatt $Var(X)$ geht auch σ^2

$$Var(X) = \sum_{i=1}^{n} (x_i - E(X))^2 P(X = x_i)$$

Eigenschaften Varianz

- Einfachere Berechnung mit Varianz:

$$Var(X) = E(X^2) - (E(X))^2$$

- $Var(aX + b) = a^2 Var(X) \forall a, b \in \mathbb{R}$
- unabhängige ZV X, Y: $Var(X + Y) = Var(X) + Var(Y)$

4.6 Standardabweichung

$$\sigma = +\sqrt{Var(X)}$$

5 Wichtige parametrische Verteilungen

5.0.1 Binomialverteilung

Modelliert Anzahl Erfolge mit fester Anzahl Versuchen
$E(X) = np$ und $Var(X) = np(1-p)$

$$P(X = x) = f(x) = \binom{n}{k} \cdot p^k (1-p)^{n-k},\ k \in T$$

$$X \sim \mathscr{B}(n, p),\ n \in \mathbb{N},\ p \in [0,1]$$

5.0.2 Negative Binomialverteilung

Verwendung: Wartezeiten/Misserfolge; älle Fehlerfolge vor Erfolgen"
$T_X = \{n, n+1, n+2, ...\}$ mit $n \in \mathbb{N}^+, \pi \in (0,1)$

$$f(x) = \binom{x-1}{n-1} \pi^n (1-\pi)^{x-n} |(x \geq n), \pi \in (0,1), T_X = \{n, n+1, ...\}\ n \in \mathbb{N}^+$$

$$X \sim \mathscr{NB}(n, \pi)$$

5.0.3 Geometrische Verteilung

Modelliert Anzahl Versuche bis A zum erster Erfolg eintritt
$E(X) = \frac{1}{p}$ und $Var(X) = \frac{1-p}{p^2}$

$$f(x) = \underbrace{(1-p)^{x-1}}_{\text{(x-1)-mal } \overline{A}} \cdot \underbrace{p}_{\text{1-Mal A}} |(x \in T), p \in (0,1)$$

$$X \sim \mathscr{G}(p)$$

5.0.4 Normalverteilung

modelliert kontinuierliche Werte um Mittelwert mit sym. Verteilung (Körpergröße, Gewicht)
Träger $T = \mathbb{R}$ mit Parametern $\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right), x \in \mathbb{R}$$

$\mu = 0$ und $\sigma^2 = 1$ ist standardnormalverteilt.

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

5.0.5 Poisson-Verteilung

Modelliert Anzahl Ereignisse mit festen Zeit/Raumbereich → konstante Rate, unabhängig (Anzahl Anrufe Callcenter)

$$f(x) = exp(-\lambda)\frac{\lambda^x}{x!} | (x \in T), t = \mathbb{N}^+$$

$$X \sim \mathcal{P}(\lambda)$$

Approximation:

$$\mathcal{B}(n, \pi) \approx \mathcal{P}(\lambda = n\pi)$$

6 Schätzung und Grenzwertsätze

6.1 Standadisierte ZV/Z-Score

$$\tilde{X} = \frac{X - \mu_X}{\sigma_X}$$

$$E(\tilde{X}) = (E(X) - \mu_X) = 0$$

$$Var(\tilde{X}) = \frac{1}{\sigma_X^2} Var(X) = 1$$

6.2 Standadisierung summierter ZVn

$$E(Y_n) = n \cdot \mu_X, \; Var(Y_n) = n \cdot \sigma_X^2$$

$$Z_n = \frac{Y_n - n\mu_X}{\sqrt{n}\sigma_X} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu_X}{\sigma_X}$$

6.3 Zentraler Grenzwertsatz

$$\mu_{\hat{p}} = p, \; SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

$$Z_n \sim \mathcal{N}(\mu = 0, \sigma^2 = 1) \text{ a :≅ asymptotisch}$$

Nett to know:

$$\mu_x = \mu, \; \sigma_x = \frac{\sigma}{\sqrt{n}}$$

6.4 Punktschätzung

1. Z-Score/Standadisieren
2. Tabellenwerk
3. Tabellenwert-(1-Tabellenwert)

6.5 Intervallschätzung

$$\hat{p} \pm 1.96 \cdot SE_{\hat{p}}$$