Vertiefung der Grundlagen der Computerlinguistik
- Zusammenfassung zum Bereich Statistik -

Klaus U. Schulz

22. Dezember 2016

Zusammenfassung

Dieses Handout fasst in kurzen Stichworten den Statistik-Teil des Kurses "Vertiefung der Grundlagen der Computerlinguistik" zusammen. Das Handout soll lediglich die Orientierung erleichtern. Zum echten Verständnis muss zusätzlich ein Buch oder Skript zur Statistik als Hauptleküre verwendet werden.

Inhaltsverzeichnis

Grundsätzlich sind zwei Teilgebiete der Statistik aus Vogelperspektive zu unterscheiden: Beschreibende versus analytische "schließende" (induktive, inferential) Statistik. Beispiele für Fragestellungen der analytischen Statistik?

1 Beschreibende Statistik

Gesamte Population versus Sample, Stichprobe.

1.1 Maße für die Größe eines mehrfach erhobenen Messwertes

- *Mittelwert* (engl. mean) $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ (= arithmetisches Mittel).
- *Median.* Um die Rolle von Ausreißern zu eliminieren. Beobachtete Werte x_i in aufsteigender Größe sortieren.
  - Falls ungerade Zahl von $2n + 1$ Beobachtungen, diejenige mit Rang $n + 1$.
  - Falls gerade Zahl von $2n$ Beobachtungen, der Mittelwert der zwei Werte mit Rang n, $n + 1$.
- *Getrimmter Mittelwert.* Bilde den Mittelwert ohne 10% größte und kleinste Werte.

- *Modus.* Wichtig bei Experimente, bei denen bestimmte Werte (Ausgänge, Beobachtungen) mehrfach auftreten, vor allem bei kategorialen Werten. Der Modus ist der am häufigsten aufgetretene Wert. Er ist eindeutig, wenn es ein eindeutiges Maximum der Frequenzverteilung gibt. Vgl. Balkendiagramme.

1.2 Maße für die Streuung eines mehrfach erhobenen Messwertes

Die Summe aller Abstände vom Mittelwert ist stets = 0, daher nicht informativ!

Empirische Varianz (n Werte $x_1, \ldots, x_n$):

$$\tilde{s}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

Empirische Standardabweichung: Wurzel aus der Varianz:

$$\tilde{s} = \sqrt{\tilde{s}^2}.$$

Stichprobenvarianz (sample variance) (n Werte):

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

Stichproben-Standardabweichung: Wurzel aus der Stichprobenvarianz:

$$s = \sqrt{s^2}.$$

Bei der induktiven Statistik wird oft Stichprobenvarianz genommen. Warum wird bei Stichprobenvarianz $n-1$ als Teiler verwendet: die Stichprobenwerte haben eine Tendenz, näher am Mittelwert der Stichprobe als am Mittelwert der Population zu liegen. Eigentlich ist man daran interessiert, die Streuung auf der Population zu schätzen. Der Teiler n würde als Ergebnis eine zu tiefe Schätzung liefern. Wird die Stichprobenvarianz gewählt, so wird als Standardabweichung $s = \sqrt{s^2}$ genommen.

Rechenregeln:

$$\begin{aligned} s^2 &= \frac{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}{n-1} \\ \tilde{s}^2 &= \frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2 \end{aligned}$$

Ist $y_i = ax_i + b$ $(1 \leq i \leq n)$, so gilt

$$\begin{aligned} \tilde{s}_y^2 &= a^2 \tilde{s}_x^2 \quad \text{und} \quad \tilde{s}_y = |a| \tilde{s}_x \\ s_y^2 &= a^2 s_x^2 \quad \text{und} \quad s_y = |a| s_x. \end{aligned}$$

1.3 Weitere Maße

Neben den Lage- und Streuungsmaßen kann man auch "Konzentrations und Dichtemaße" (Hintergrundfrage: liegen viele Werte ggfs. bei verschiedenen Punkten nahe beisammen?) angeben. Siehe Kap. 2 Fahrmeir et al. Interessant auch Beschreibung mehrdimensionaler Daten (Kapitel 3).

2 Analytische Statistik

2.1 Wahrscheinlichkeitsbegriff

Ein "Experiment" oder "Zufallsvorgang" ist ein wiederholbarer Prozess, der Beobachtungen (Ausgänge, Werte) erzeugt.

"Stichprobenraum", "Ergebnisraum" (Sample Space): Raum aller möglichen Beobachtungen (Ausgänge, Werte) eines Experiments.

"Ereignis" (Event): Eine Menge möglicher Beobachtungen (Ausgänge, Werte), Teilmenge des Stichprobenraums.
"Einfaches Ereignis" (Simple event): einelementige Menge.
"Zusammengesetzes Ereignis (compound event)": mehrelementige Menge.

Boolesche Operationen für Ereignisse: Vereinigung, Durchschnitt, Differenz, Komplement. Disjunkte Ereignisse.

Wahrscheinlichkeitsraum: Paar (S, P) wo S ein Stichprobenraum, P Abbildung von $2^S \to \mathbb{R}$ mit drei Eigenschaften:

1. $P(A) \geq 0$ für alle $A \subseteq S$,
2. $P(S) = 1$,
3. Additivität für endliche Menge disjunkter Ereignisse, Additivität für abzählbar unendliche Folge von Ereignissen.

Rechenregeln:

Boolesche Regeln für Mengen.
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Verallgemeinerung: Wahrscheinlichkeit der Vereinigung von drei (mehreren) beliebigen Ereignissen.

2.2 Zähltechniken

Falls alle Ausgänge eines Experiments gleichwahrscheinlich sind, so reduzieren sich Wahrscheinlichkeitsberechnungen auf geschicktes Zählen.

Produktregel. Es sollen nacheinander m Dinge gewählt werden. Hat man bei der i-ten Wahl n_i Möglichkeiten, so gibt es

$$\Pi_{i=1}^{m} n_i$$

Auswahlen (m-Tupel).

Geordnete Folgen (mittels Produktregel) $n!$ mögliche Anordnungen (geordnete Reihenfolgen) von n Elementen in einer Folge. (Bem. $0! = 1$) Wieviele geordnete Reihenfolgen von k Elementen aus einer Auswahl von n Elementen gibt es?

$$(n(n-1)(n-2)\dots(N-k) = \frac{n!}{(n-k)!}$$

Ungeordnete Auswahlen Wieviele Möglichkeiten gibt es, k Elemente aus Gesamtheit von n auszuwählen: Müssen $n!/((n-k)!)$ durch $k!$ teilen:

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

Hierbei $\binom{n}{n} = \binom{n}{0} = 1$.

2.3 Bedingte Wahrscheinlichkeit und Bayessches Gesetz

Bedingte Wahrscheinlichkeit. $P(A|B) = \frac{P(A \cap B)}{P(B)}$ damit

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

(Daten der rechten Seite oft leichter zu kriegen, Daten links interessant: Beispiel: Von vier Probanden habe einer Blutgruppe A-pos, es sei unklar, welcher. Wie groß ist W, dass man zumindest drei Probanden testen muß, bevor man den mit Blutgruppe A-pos erwischt?
B = der erste hat nicht A-pos,
A = der zweite hat nicht A-pos,
Gesuchte W ist $P(A \cap B) = P(A|B)P(B) = \frac{2}{3} \cdot \frac{3}{4} = 0,5$.

Bayessches Gesetz, einfache Form:

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

Anwendungsbeispiel. Eine von 1000 Personen hat seltene Krankheit: Es gibt einen Test:
- (i) falls Person Krankheit hat, in 99% aller Fälle positives Testergebnis.
- (ii) falls Person Krankheit nicht hat, in 2% aller Fälle positives Testergebnis.
Falls eine zufällig gewählte Person positiv getestet wird, wie groß ist W, dass die Person die Krankheit hat? (Entspricht Umdrehen! der Frage in (ii)!!). Modellierung:
A: Person hat Krankheit.
B: positives Testergebnis.
$P(B) = 0,001 \cdot 0,99 + 0,999 \cdot 0,02 = 0,02097$.
Gesucht ist $P(A|B) = P(B|A)P(A)/P(B) = 0,99 \cdot 0,001/0,02097 = 0,047$. D.h. die allermeisten Fälle sind "blinder Alarm"!! Liegt an der Seltenheit der Krankheit. (Je seltener eine Krankheit, desto anspruchsvoller ist es, einen Test zu entwickeln, der falsche Alarme weitgehend vermeidet.)

Bayessches Gesetz, allgemeine Form: Seien $A_1, \ldots, A_n$ disjunkte Ereignisse, $P(A_i) > 0$, Vereinigung sei Ereignisraum S. Dann gilt für jedes Event B mit $P(B) > 0$ und jedes k stets

$$P(A_k|B) = P(A_k \cap B)/P(B) = \frac{P(B|A_k)P(A_k)}{\sum_i P(B|A_i)P(A_i)}$$

Nenner: disjunkte Zerlegung von $P(B)$, da $P(B|A_i)P(A_i) = P(B \cap A_i)$.

Unabhängigkeit zweier Ereignisse. Ereignisse A, B heißen *unabhängig* gdw. $P(A|B) = P(A)$. Äquivalent: $P(A \cap B) = P(A) \cdot P(B)$.

Unabhängigkeit von n Ereignissen $A_1, \ldots, A_n$. Für jede Auswahl von $k \geq 2$ Ereignissen $A_{i_1}, \ldots, A_{i_k}$ gilt $P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdot \ldots \cdot P(A_{i_k})$.

Nicht gleichbedeutend mit paarweiser Unabhängigkeit: Beispiel: Drei Preise, Vier Zettel in Box: Preis 1, Preis 2, Preis 3, Preise 1,2,3 Einer wird gezogen. A_i ($i = 1, 2, 3$): Preis i steht auf gezogenem Zettel. Jeder Zettel hat W 1/4.

$$P(A_1) = P(A_2) = P(A_3) = 1/2.$$

$P(A_1 \cap A_2) = 1/4 = P(A_1) \cdot P(A_2)$, analog für Paare 1-3, 2-3. Also gilt paarweise Unabhängigkeit. Aber $P(A_1 \cap A_2 \cap A_3) = 1/4$, jedoch $P(A_1) \cdot P(A_2) \cdot P(A_3) = 1/8$. Also nicht alle n unabhängig.

3 Zufallsvariablen (random variables)

Ordnet man den Ergebnissen eines Zufallsexperiments reelle Zahlen zu, erhält man Zufallsvariable (ZV). "ZV" ist also Variable oder Merkmal X, dessen Werte

Ergebnisse eines Zufallexperiments sind. Man fragt dann nach der Wahrscheinlichkeit, einen bestimmten Zahlenwert zu erhalten: $P(X = x)$. Man unterscheidet eindimensionale (ein Wert) und mehrdimensionale (mehrere ggfs. voneinander abhängige Zahlenwerte) ZV. Diskrete und stetige ZV.

3.1 Diskrete ZV (discrete random variables)

ZV X heißt "diskret", falls X endlich viele oder abzählbar unendlich viele Werte $x_1, \ldots, x_k, \ldots$ annehmen kann. Menge $\{x_1, \ldots, x_k, \ldots\}$ heißt *Wertemenge* oder *Träger* von X.

"*Wahrscheinlichkeitsverteilung*" von X durch die Werte

$$p_i = P(X = x_i) \quad (i = 1, \ldots, k, \ldots)$$

gegeben. (Wahrscheinlichkeiten nur für Trägermenge) Ist A Zahlenmenge, so $P(X \in A) = \sum_{i:x_i \in A} p_i$.

Sonderfall binäre ZV X, einzige Werte $0, 1$. Zwei Werte

$$P(X = 1) = \pi, \quad P(X = 0) = 1 - \pi.$$

X heißt dann"Bernoulli-Variable", das Experiment ein "Bernoulli-Experiment". Oft wird $X = 1$ mit dem Eintritt eines "speziellen Ereignisses A" gleichgesetzt.

"*Wahrscheinlichkeitsfunktion*" $f(x)$ einer diskreten ZV X ist für $x \in IR$ definiert durch

$$f(x) = \begin{cases} p_i = P(X = x_i) & \text{falls } x = x_i \text{ zum Träger von } X \text{ gehört,} \\ 0 & \text{sonst.} \end{cases}$$

Oft wird nicht zwischen Wahrscheinlichkeitsverteilung und Wahrscheinlichkeitsfunktion unterschieden (formal ist nur der Definitionsbereich unterschiedlich). Im Englischen "probability distribution".

"*Verteilungsfunktion*" (engl. "Cumulative distribution function") $F(x)$ einer diskreten ZV ist für $x \in \mathbb{R}$ definiert durch

$$F(x) = P(X \leq x)$$

(Summe der Ws von Werten von X kleinergleich x). Treppenfunktion.

Beachte Unterschied $f(x)$, $F(x)$!!!

3.2 Beispiele diskreter ZV

Diskrete Gleichverteilung: Endlicher Träger $\{x_1, \ldots, x_k\}$, $P(X = x_i) = 1/k$.

Geometrische Verteilung: Wiederholen Bernoulli-Experiment mit Variable Y (Parameter $\pi = P(Y = 1)$) so oft (unabhängige Versuche), bis zum ersten Mal 1 (bzw. charakteristisches Ereignis A) auftritt. Setzen nun X = Zahl der Versuche, bis zum ersten Mal A eintritt. Träger ist unendlich $1, 2, 3, \ldots$.

$$p_k = P(X = k) = (1 - \pi)^{k-1}\pi$$

Wahrscheinlichkeiten p_k bilden eine geometrische Folge. ZV X heißt *geometrisch verteilt* mit Parameter π, man schreibt $X \sim G(\pi)$.

Anwendungen/Beispiele: Lebensdauer eines Geräts in Tagen, Dauer in Tagen, bis Kunde einer Versicherung den ersten Schaden meldet, ??Zahl der Wörter in einem Text, bis das erste Nomen (Verb) kommt (!Problem: aufeinanderfolgende Worttypen nicht unabhängig!)?

Binomialverteilung: Wiederholen ein Bernoulli-Experiment mit Parameter $P(A) = \pi$ genau n mal. Fragen, wie oft charakteristisches Ereignis A auftritt, diese Zahl ist ZV X. (Standardbeispiel: Ziehen von Kugeln aus Urne mit Zurücklegen. In der Urne sind N Kugeln, darunter M schwarze. A = es wird schwarze Kugel gezogen.)
A_i = beim i-ten Versuch wird eine schwarze Kugel gezogen. Träger von X ist $\{0, 1, \ldots, n\}$. EINE FESTE Folge von n Ergebnissen, wo genau x mal Ereignis X eintritt, hat W $\pi^x(1 - \pi)^{n-x}$. Die Zahl unterschiedlicher derartiger Folgen gleicht der Zahl der Möglichkeiten, x Element aus n auszuwählen: $\binom{n}{x}$. Also gilt

$$P(X = x) = \binom{n}{x}\pi^x(1 - \pi)^{n-x}$$

ZV X heißt dann *binomialverteilt mit Parameter n und π*, kurz $X \sim B(n, \pi)$.

Anwendungen/Beispiele: Zahl der Nieten, wenn man n Lose kauft. Wieviele muss man kaufen, dass man mit W 90% ein Gewinnlos zieht (Unterschiedliche Binomialverteilungen zu betrachten)? Wenn π Prozent Eier schlecht sind, wie ist W, dass in Schachtel mit 20 ein schlechtes ist? ??Wie groß ist W, daß Satz der Länge n mindestens zwei Verben enthält (!Problem: aufeinanderfolgende Worttypen nicht unabhängig!)?? !!Wie groß ist W, daß von 10 zufällig gewählten Wörtern aus einem deutschen Text 35 nicht flektierende sind?!

Fr die Werte der Binomialverteilung zu wichtigen Parametern siehe ausgeteilte Kopien. Als natürliche Verallgemeinerung der Binomialverteilung ergibt sich die Multinomialverteilung. (Bild der Urne mit Kugeln unterschiedlicher Farben - wie groß ist die Wahrscheinlichkeit, bei n Ziehungen mit Zurücklegen eine bestimmte Farbverteilung in der Ergebnismenge zu erhalten.)

Hypergeometrische Verteilung. Urnenmodell ohne Zurücklegen. N Einheiten (Kugeln), M haben eine Eigenschaft (sind schwarz), ziehen genau n mal.
X = Zahl der gezogenen Einheiten mit Eigenschaft.
Träger ist $max(0, n-(N-M), \ldots, min(n, M)$. Wahrscheinlichkeitsfunktion

$$f(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

für x aus Träger, 0 sonst. ZV X heißt *hypergeometrisch verteilt mit Parametern* n, N, M. Schreibweise $X \sim H(n, N, M)$. (Falls n/N klein, so ist X annähernd binomialverteilt, Parameter n und $\pi = M/N$.)

Anwendungen/Beispiele: In einer Klasse von 20 Schülern sind 10 gut, 10 schlecht. Wenn wir n zufällig auswählen, wie groß ist W, dass wir darunter genau x gute haben?

Poissonverteilung: ZV X mit Verteilung

$$f(x) = \begin{cases} P(X=x) = \frac{\lambda^x}{x!}e^{-\lambda} & \text{für } x = 0, 1, 2, \ldots \\ 0 & \text{sonst} \end{cases}$$

heißt *Poisson-verteilt mit Parameter* λ. Schreiben $X \sim Po(\lambda)$.

Hintergrund "Poissonprozess". Zu bestimmten, nicht vorhersagbaren Zeitpunkten soll ein bestimmtes Ereignis eintreten. Hierbei sollen folgende Annahmen gelten:

1. Zwei Ereignisse treten nie genau gleichzeitig auf.
2. Es gibt einen Parameter α, so daß die W, dass in sehr kleinem Zeitintervallen der Länge t genau ein Ereignis auftritt, annähernd αt ist. (D.h. α ist eine Art "Ereignisrate".)
3. Die W für das Eintreten einer bestimmten Zahl von Ereignissen in einem Teilintervall hängt nur von Länge l, aber nicht von der Lage des Intervalls ab.
4. Die Anzahlen von Ereignissen in disjunkten Teilintervallen sind unabhängig.

Dann liegt ein "Poissonprozess" vor und die W, x Ereignisse in Intervall der Länge t zu haben, ist

$$\frac{(\alpha \cdot t)^x e^{-\alpha \cdot t}}{x!}$$

Bezogen auf Interval der Länge t haben wir eine Poissonvariable mit Parameter $\lambda = \alpha \cdot t$.

Anwendungen:

- Zahl der radioaktiven Zerfälle in einer Minute bei gegebener Rate.
- Zahl der eingehenden Telefonanrufe pro Minute in einer großen Telefonzentrale.
- Zahl der Großschäden pro Monat bei einer Versicherung.
- ? Zahl der in einer Stunde (zwischen 8 und 18 Uhr) ankommenden Buße an einer Bushaltestelle??

3.3 Grenzverhalten

Bemerkung: Hatten gesagt, dass hypergeometrische Verteilung mit Parametern n (Zahl der Ziehungen), M (Zahl schwarze Kugeln), N (Kugeln) für n/N klein (Faustregel: nicht mehr als 5% der Kugeln werden gezogen) gegen $B(n,\pi) = B(n, M/N)$ geht.

Für sehr große Zahl von Ziehungen n und sehr kleinen Parameter π und $n \cdot \pi = \lambda$ konstant geht Binomialverteilung $B(n,\pi)$ gegen Poissonverteilung $Po(\lambda)$!

Unabhängigkeit

Oft erhlt man aus einem Experiment unterschiedliche Werte und Messungen. Hieraus ergeben sich mehrere Zufallsvariablen (siehe Abschnitt mehrdimensionale ZV unten). Die n diskreten ZV $X_1, \ldots, X_n$ heißen *unabhängig*, wenn für beliebige Elemente $x_1, \ldots, x_n$ der jeweiligen Träger gilt

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdot \ldots \cdot P(X_n = x_n).$$

3.4 Erwartungswert und Varianz diskreter ZV

Erwartungswert einer diskreten ZV. Sei X diskrete ZV mit Träger $x_1, \ldots, x_k, \ldots$ Der Erwartungswert von X ist

$$\begin{aligned} \mu = \mu_X = E(X) &= x_1 \cdot P(X = x_1) + \ldots + x_k \cdot P(X = x_k) + \ldots \\ &= \sum_{x_i \text{ im Träger}} x_i f(x_i) \end{aligned}$$

Beispiel Würfeln $\frac{1}{6}(1+2+3+4+5+6) = \frac{21}{6} = 3,5$.

Rechenregeln zum Erwartungswert:
- Ist $Y = aX + b$, so $E(Y) = aE(X) + b$.

- $E(X+Y) = E(X) + E(Y)$
- Sind X und Y unabhängig(!!), so $E(XY) = E(X)E(Y)$.

Varianz einer diskreten ZV Sei X diskrete ZV mit Träger $x_1, \ldots, x_k, \ldots$ Die Varianz von X ist

$$\begin{aligned} \sigma^2 &= Var(X) = (x_1 - \mu)^2 f(x_1) + \ldots + (x_k - \mu)^2 f(x_k) + \ldots . \\ &= \sum_{x_i \text{ im Träger}} (x_i - \mu)^2 f(x_i) \end{aligned}$$

Die Standardabweichung ist $\sigma = \sqrt{\sigma^2}$.

Damit $Var(X) = E(X - \mu)^2$. Regeln:
- $Var(X) = E(X^2) - E(X)^2$.
- Ist $Y = aX + b$, so $Var(Y) = a^2 Var(X)$ und $\sigma_Y = |a|\sigma_X$
- $Var(X + Y) = Var(X) + Var(Y)$

Aufgabe: Varianz beim Würfelexperiment.

Erwartungswert und Varianz bei bekannten diskreten ZV

Geometrische Verteilung: Falls $X \sim G(\pi)$, so $E(X) = 1/\pi = Var(X)$.
Binomialverteilung: Falls $X \sim B(n, \pi)$, so gilt $E(X) = n\pi$ und $Var(X) = n\pi(1 - \pi)$.
Hypergeometrische Verteilung: Falls $X \sim H(n, N, M)$, so $E(X) = nM/N$, $Var(X) = (1 - M/N)(N - n/N - 1)$.
Poissonverteilung: Falls $X \sim Po(\lambda)$, so $E(X) = Var(X) = \lambda$.

3.5 Stetige ZV (continuous random variables)

Definition: ZV X heißt "stetig", wenn es eine Funktion $f(x) \geq 0$ gibt , so dass für jedes Intervall $[a, b]$ stets gilt

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

(= Fläche unter dem Graph der Funktion f von Grenze a bis b)

Funktion f heißt "Wahrscheinlichkeitsdichte" oder "Dichte" oder Dichtefunktion" von X ("probability distribution" of X, "probability density function" of X). Es gilt: Gesamtfläche unter Graph von $f = \int_{-\infty}^{+\infty} f(x)dx = 1$.

Die "Verteilungsfunktion" ("cumulative distribution function") zu X ist

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$$

(Gesamtfläche unter dem Graph von f von $-\infty$ bis x)

Es gilt $(a < b)$ $P(a \leq x \leq b) = F(b) - F(a)$. An Stellen x, wo die Ableitung F' von F existiert, gilt $F'(x) = f(x)$.

Unabhängigkeit stetiger ZV. Stetige ZV X und Y heißen unabhängig, wenn für alle x und y gilt $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$.

3.6 Beispiele stetiger ZV

Stetige Gleichverteilung auf Interval $[a, b]$: $f(x) = 1/(b - a)$ für $a \leq x \leq b$, $f(x) = 0$ sonst.

Exponentialverteilung: Pendant zur geometrischen Verteilung, Wartezeit zum Eintreffen eines Ereignisses unter der Voraussetzung, dass die bereits verstrichene Zeit keinen Einfluß auf die Restwartezeit hat. Stetige ZV mit nichtnegativen Werten heißt exponentialverteilt mit Parameter λ, man schreibt $X \sim Ex(\lambda)$, wenn sie die Dichte

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x \geq 0, \\ 0 & \text{sonst} \end{cases}$$

besitzt. Verteilungsfunktion

$$F(x) = 1 - e^{-\lambda x}.$$

Enger Zusammenhang zur Poissonverteilung: bei Exponentialverteilung betrachten wir Wartezeit bis zum ersten Eintreffen. Die ZAHL von Ereignissen in Zeitintervall ist poissonverteilt mit Parameter λ genau dann, wenn die Zeitdauern ZWISCHEN aufeinanderfolgenden Ereignissen unabhängig und exponentialverteilt mit Parameter λ sind.

Normalverteilung. Wichtige Familie stetiger ZV, durch zwei Parameter μ (Mittelwert) und σ (Standardabweichung) bestimmt. Dichte ist durch

$$f(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2)$$

für $x \in \mathbb{R}$ gegeben. Glockenkurve. Bedeutung: Gutes Modell für Verteilung eine Variable, wenn diese durch Zusammenwirken einer größeren Zahl von zufälligen Einflüssen entsteht: Auftretende Messfehler, Abweichung von Sollwert bei Produktion etc., Punktezahlen in Tests,.. $f(x)$ Formel, Glockenkurve, Maximum

bei μ, Wendepunkte bei $\mu + -\sigma$ Verteilungsfunktion $F(x)$ nicht analytisch in geschlossener Form angebbar. Tabellen!

Standardnormalverteilung. Normalverteilung mit Parametern $\mu = 0$, $\sigma = 1$, Dichte

$$\phi(z) = \frac{1}{\sqrt{2\pi}} exp(-\frac{z^2}{2})$$

Berechnungen mit allgemeinen Normalverteilungen lassen sich auf Tabellen oder Berechnungsverfahren zur Standardnormalverteilung zurückführen, wenn man μ und σ kennt.

χ-Quadratverteilung. Sind $X_1, \ldots, X_n$ unabhängige und standardnormalverteilte ZV, dann heißt die Verteilung der ZV $Z = X_1^2 + \ldots + X_n^2$ Chi-Quadrat-Verteilung mit n Freiheitsgraden, oder $\chi^2(n)$-Verteilung, man schreibt $Z \sim \chi^2(n)$. Große Bedeutung in der Testtheorie.

t-Verteilung. Auch Student-Verteilung genannt. Ist X standardnormalverteilt, $Z \sim \chi^2(n)$ und sind X und Z unabhängig, so heißt die Variable

$$T := \frac{X}{\sqrt{Z/n}}$$

t-verteilt mit n Freiheitsgraden, $T \sim t(n)$. Die Verteilung wird als $t(n)$-Verteilung bezeichnet. Große Bedeutung bei Parametertests.

3.7 Erwartungswert und Varianz stetiger ZV

Der *Erwartungswert* einer stetigen ZV X mit Dichtefunktion $f(x)$ ist

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

Die *Varianz* eine stetigen ZV X ist mit Dichtefunktion $f(x)$ ist

$$\sigma^2 = \sigma_X^2 = Var(x) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx.$$

Die *Standardabweichung* ist $\sigma = \sqrt{Var(X)}$.

Es gibt viele Rechenregeln...s. Fahrmeir et al.

3.8 Quantile

Englisch quantil oder percentile: Sei p reelle Zahl in $[0, 1]$. Jeder Wert x_p, für den $P(X \leq x_p) = F(x_p) \geq p$ und $P(X \geq x_p) \geq 1 - p$ gilt, heißt *p-Quantil*

der ZV X. Bei diskreten ZV oft nicht eindeutig: falls es einen Wert x gibt, so dass $P(X < x) < p$ und $P(X \geq x) > p$ gilt, so ist dieser Wert x das in diesem Fall eindeutig bestimmte p-Quantil. Falls jedoch p als Wert einer Treppenstufe angenommen wird (d.h. es gibt ein x im Träger mit $P(X \leq x) = p$), so sind alle Werte zwischen x und dem nächsten Punkt im Träger p-Quantile. Bei stetigen ZV hingegen ist sind p-Quantile meist (nicht immer) eindeutig: das p-Qantil ist Wert x_p, für den $F(x_p) = p$ gilt. Bedeutung in Testtheorie: die Annahme, dass eine ZV eine bestimmte Verteilung hat, wird als wahrscheinlich falsch betrachtet, wenn Wert in einem Test (in mehreren Tests) unterhalb des $0,05$ $(0,025)$ oder oberhalb des $0,975$-Quantils liegen. Median ist 50% Quantil. (Unterschied zum Erwartungswert! Z.B: verschiebt man den oberen 51% Anteil der W-Masse nach rechts, wächst Erwartungswert, Median bleibt.)

4 Mehrdimensionale ZV

Idee: Erfassung unterschiedlicher numerischer Werte oder Merkmale bei derselben Art von Zufallsexperiment oder denselben Untersuchungseinheiten. Interessante Frage ist meist, ob und die die unterschiedlichen Werte zusammenhängen. Wie hängen Hubraum und Verbrauch bei zufällig gewählten Automobilen zusammen? Wie Wohnungsgröße (qm) und Mietpreis bei Wohnungen?

Oder Roulette: X Werte 1 (rot) 2 (schwarz) 3 (Zero) Y Werte 1 (gerade Zahl) 2 (ungerade Zahl) 3 (Zero)

Allgemeiner ZV $X_1, \ldots, X_n$. Fragen nach der gemeinsamen Verteilung

$$P(X_1 \text{ in Wertebereich } 1, \ldots, X_n \text{ in Wertebereich } n).$$

4.1 Zweidimensionale diskrete ZV

Nachfolgend $n = 2$: "bivariate ZV" (X, Y).
X Träger $x_1, x_2, \ldots$
Y Träger $y_1, y_2, \ldots$
W-Funktion

$$f(x,y) = \begin{cases} P(X = x, Y = y) & \text{für } (x,y) \text{ im gemeinsamen Träger} \\ 0 & \text{sonst.} \end{cases}$$

Bei endlichen Trägern beschreibbar durch Kontingenztabellen mit Einträgen $p_{i,j} = W(X = x_i, Y = y_j)$, vgl. Tabelle 1.

Durch Aufsummieren über Zeilen oder Spalten erhält man die Wahrscheinlichkeiten, die sich ergeben, wenn man eine der beiden Variablen "vergisst": Rand-

	y_1	y_2	$\cdots$	y_m	
x_1	$p_{1,1}$	$p_{1,2}$		$p_{1,m}$	$p_{1.}$
x_2	$p_{2,1}$	$p_{2,2}$		$p_{2,m}$	$p_{2.}$
$\cdots$					
x_k	$p_{k,1}$	$p_{k,2}$		$p_{k,m}$	$p_{k.}$
	$p_{.1}$	$p_{.2}$		$p_{.m}$	

Tabelle 1: Kontingenztabelle für zweidimensionale diskrete ZVs.

verteilungen
Randverteilung für X: $f_X(x) = P(X = x) = \sum_j f(x, y_j)$ (am rechten Tabellenrand)
Randverteilung für Y: $f_Y(y) = P(Y = y) = \sum_i f(x_i, y)$ (unterste Zeile).

4.2 Zweidimensionale stetige ZV

Die Zufallsvariablen X und Y sind gemeinsam stetig verteilt, wenn es eine zweidimensionale Dichtefunktion $f(x, y) \geq 0$ gibt, so dass $P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y)\, dy\, dx$ (Gebirge mit Gesamtvolumen über Ebene 1.) Man kann Randdichten (analog zu Randverteilungen oben) definieren, aber hier Funktionen: $f_X(x) = \int_{-\infty}^{+\infty} f(x, y)dy$ $f_Y(y) = \int_{-\infty}^{+\infty} f(x, y)dx$

4.3 Kovarianz und Korrelationskoeffizient

Kovarianz. Die Varianz bei einer eindimensionalen ZV gibt ein Maß an, wie stark die Werte um den Mittelwert streuen. Die Kovarianz einer bivariaten ZV (X, Y) erfasst, ob Abweichungen von X UEBER den Erwartungswert von X typischerweise gekoppelt sind mit Abweichungen von Y UEBER den Erwartungswert von Y (positive Kovarianz), oder Abweichungen von X UEBER den Erwartungswert von X typischerweise gekoppelt sind mit Abweichungen von Y UNTER den Erwartungswert von Y (negative Kovarianz).

$$Cov(X, Y) = E([X - E(X)][Y - E(Y)])$$

Beispiele: Stellen (X, Y) Gewicht und Körpergröße dar, erwartet man eine positive Kovarianz: Übergroße Personen haben tendentiell überdurchschnittliches Gewicht. Anders hingegen Alter und Körpergröße.

Unabhängige ZV X und Y haben Kovarianz 0.

Korrelationskoeffizient.

$$\begin{aligned}\rho(X,Y) &= Cov(X,Y)/(\sqrt{Var(X)}\sqrt{Var(y)}) \\ &= Cov(X,Y)/\sigma_X\sigma_Y\end{aligned}$$

Wertebereich zwischen -1 (negative Korrelation) und $+1$ (positive Korrelation).

Beispiel S 352.

Beispiel für empirische Kovarianz: Fehlerprofile (Reffle,Ringlstetter), siehe Kopie.

Eine positive Korrelation wird manchmal so ausgelegt, dass ein ursächliches Verhältnis (Kausalität, Merkmal X ist ausschlaggebend für Merkmal Y) vorliegt. Dies ist aber nicht immer gerechtfertigt. X und Y können gemeinsame Folgen einer ganz anderen Ursache sein. Der Verdacht einer Kausalität ergibt sich insbesondere, wenn eine zeitliche Verzögerung im Spiel ist. Leute, die als Kind ein Musikinstrument gespielt haben, haben später (im Durchschnitt) eine höhere Intelligenz. (Liegt das am Musizieren oder am besseren Elternhaus...?)

Sicher gilt aber: falls eine Kausalität (X beeinflusst wirklich direkt Y) gegeben ist, so sollte die sich empirisch in positiver Korrelation niederschlagen (wenn es keine anderen stärkeren Faktoren der Beeinflussung gibt)!

Siehe hierzu Regressionsanalyse!

5 Parameterschätzungen

Man möchte durch eine Stichprobe in der Regel auf Eigenschaften einer nicht direkt gegebenen W-Verteilung oder auf Eigenschaften einer großen Population schließen (Wieviele Leute wählen CDU? Wie groß ist die Wahrscheinlichkeit, daß ein Haus in Bayern massiv durch einen Sturm beschädigt wird?). Viele Schätzungen versuchen, einen bestimmten Wert konkret zu schätzen (*Punktschätzung*). Prinzipiell lassen sich Situationen unterscheiden, wo wir
(a) Keine Annahme machen, daß eine bestimmte Art von W-Verteilung (wie Normalverteilung) vorliegt, bzw.
(b) annehmen, dass eine bestimmte Art von Verteilung vorliegt (Familie klar, aber Parameter nicht).
Im Fall (a) möchte man allgemeine Parameter/Werte wie den Erwartungswert, die Varianz, den Median oder ein p-Quantil abschätzen. Im Fall (b) möchte man ggfs. zusätzlich die unbekannten Parameter abschätzen.

Da man auf Stichprobe angewiesen ist, ergeben sich viele Fragen. Wie (durch welche Funktion) erhalten wir aus den Daten für was eine Abschätzung? Wie

"zuverlässig" ist die Abschätzung? Mit welchem Risiko liegen wir in Abhängigkeit von der Stichprobengröße wie weit daneben?

Neben Punktschätzungen gibt es auch INTERVALLSCHAETZUNGEN. Hier gelangt man zu Aussagen der Form "mit Wahrscheinlichkeit/Konfidenz 98% liegt der wahre Wert im Intervall $[a, b]$."

5.1 Punktschätzungen

Ausgangspunkt n Stichprobenziehungen dargestellt durch "Stichprobenvariablen" $X_1, \ldots, X_n$ (Bild hier: Stichprobe noch nicht gezogen, Ziehen wäre wiederholbar.) Wenn wir die Ziehung explizit durchführen, erhalten wir Werte $x_1, \ldots, x_n$ Aus $x_1, \ldots, x_n$ soll auf einen bestimmten Parameter θ (s.o.) geschlossen werden.

SCHAETZFUNKTION (oder SCHÄTZSTATISTIK) $T = g(X_1, \ldots, X_n)$ ist Zufallsvariable T (denn: bei anderer Ziehung ergäbe sich anderer Wert), Funktion der Stichprobenvariablen. Der aus Realisierungen $x_1, \ldots, x_n$ erhaltene Wert ist der SCHAETZWERT.

Beispiele: "Natürliche" Schätzfunktionen erhält man aus bekannten empirischen Kennwerten:
Durchschnitt (arithm. Mittel) als Schätzung für Erwartungswert $g(x_1, \ldots, x_n) = 1/n(x_1 + \ldots + x_n)$.
$1/n$(Anteil der Treffer bei n Versuchen) als Schätzwert für π (Bernoulli-Experiment oder Binomialverteilung).
Empirische Varianz $(1/n)\sum_i (X_i - \bar{X})^2$ oder Stichprobenvarianz $(1/n-1)\sum_i (X_i - \bar{X})^2$ als Schätzung für Varianz σ^2.

Es gibt Eigenschaften, die Schätzfunktionen unbedingt immer haben sollten, und solche, die "idealerweise" vorliegen sollten.

Falls irgendwie möglich: "Erwartungstreue": Hat der zu schätzende Parameter den tatsächlichen Wert θ, so sollte der Erwartungswert von T auch θ sein! Abschwächung ist asymptotische Erwartungstreue für immer höheren Stichprobenumfang n.

Es gibt unterschiedliche Kriterien, wie man die Qualität zweier Schätzfunktionen für dieselbe Aufgabe vergleichen kann. Von zwei erwartungstreuen Schätzfunktionen ist die "besser", die die kleinere Varianz hat.

Maximum Likelihood Schätzung. Ein wichtiges allgemeines Prinzip, um Punktschätzungsfunktionen zu erhalten. Annahme ist, dass eine bekannte Familie von W-Verteilungen zu einem Parameter θ (es können auch mehrere Parameter

sein, dies ist komplexer..) vorliegt, man möchte den Parameter θ aus $x_1, \ldots, x_n$ schätzen. Wenn die $X_1, \ldots, X_n$ unabhängige Wiederholungen einer ZV mit Wahrscheinlichkeitsfunktion/Dichte $f(x)$ sind, gilt für die W-Werte/Dichten wenn man annimmt, das Wert θ der gesuchte Parameterwert ist

$$f(x_1, \ldots, x_n|\theta) = f(x_1|\theta) \cdot \ldots \cdot f(x_n|\theta)$$

Für feste $x_1, \ldots, x_n$ kann man dies aber auch als eine Funktion von θ auffassen. Für jeden Wert θ ergibt sich ein eigener Wert $f(x_1, \ldots, x_n|\theta)$. Diesen Wert (für festes $x_1, .., x_n$) kann man als Funktion

$$L(\theta) = f(x_1, \ldots, x_n|\theta).$$

schreiben. $L(\theta)$ heißt Likelihoodfunktion. Das Maximum Likelihood Prinzip sagt: wähle zu $x_1, \ldots, x_n$ als Schätzwert denjenigen Wert θ, für den die Likelihoodfunktion maximal ist.

Rechnerisch ergibt sich das Auffinden das maximalen Werts von L durch Nullsetzen der Ableitung. Meist enthält die Darstellung von L Produkte mit mehreren Faktoren (s.o.). Für die Ableitung ergeben sich dann komplexe Ausdrücke. $(fg)' = f'g + fg'$ Logarithmieren ist streng monoton, führt also auf dieselben Maxima, macht aus Produkten Summen. Hier Ableitung einfacher. $(f + g)' = f' + g'$. Log-Likelihood-Prinzip.

Beispiel: Buch Fink S. 50.

Maximum-a-posteriori-Schätzung. Wird verwendet, wenn man wenige Schätz-Daten hat. Wenn man über die Verteilung des Parameters θ Vorwissen hat, kann man die Wahrscheinlichkeit eines Wertes für θ in die Schätzung miteinfließen lassen. Siehe z.B. Buch von Fink 3.6.2.

5.2 Intervallschätzungen

Liefert als Ergebnis der Schätzung nicht einen Punkt, sondern ein Intervall, das "höchstwahrscheinlich" den wahren Parameter enthält. Die Wahrscheinlichkeit, dass das sich ergebende Intervall den wahren Wert erhält, wird "eingestellt" durch Konfidenzschranke. Üblich sind Konfidenzschranken von $0,9$ oder $0,95$ oder $0,99$. Dual ist die Irrtumswahrscheinlichkeit (W für Ereignis echter Wert nicht im Schätzintervall) α, typische Werte sind $0,1$ oder $0,05$ oder $0,001$. Standardsymbole: Irrtumswahrscheinlichkeit α, Konfidenzschranke $1 - \alpha$.

Man verwendet zwei Stichprobenfunktionen

$$G_{\text{unten}}(X_1, \ldots, X_n), \quad G_{\text{oben}}(X_1, \ldots, X_n)$$

Zu vorgegebener Konfidenzschranke $1 - \alpha$ liefern G_{unten} und G_{oben} zusammen ein $(1 - \alpha)$-Konfidenzintervall, falls gilt:

1. Die W, daß $G_{\text{unten}}(X_1, \ldots, X_n) \leq G_{\text{oben}}(X_1, \ldots, X_n)$ ist 1.
2. die W, daß das sich ergebende Intervall

$$[G_{\text{unten}}(X_1, \ldots, X_n), G_{\text{oben}}(X_1, \ldots, X_n)]$$

den wahren Wert θ enthält, ist $1 - \alpha$.

Manchmal möchte man für Aussagen der Form "θ ist fast sicher MINDESTENS..." Dann wählt man ein einseitiges Konfidenzinterval mit $G_{\text{oben}} = \infty$.

Manchmal möchte man für Aussagen der Form "θ ist fast sicher HOECHSTENS..." Dann wählt man ein einseitiges Konfidenzinterval mit $G_{\text{unten}} = \infty$.

Manchmal möchte man (mit Restrisiko α) sehr große und sehr kleine Werte für θ ausschließen, dann wählt man ein zweiseitiges Konfidenzinterval.

Beachte: der wahre Wert θ ist fest und nicht zufallsabhängig. Was durch den Zufall gesteuert wird, ist das Ziehen der Stichprobe und das sich hieraus ergebende Intervall. Es resultiert mit W $1 - \alpha$ ein Intervall, das θ enthält. Hat man ein konkretes Intervall festgelegt, so macht es keinen Sinn, von der Wahrscheinlichkeit zu reden, mit der θ in diesem Intervall liegt. Dies gilt entweder, oder es gilt nicht.

Allgemein gilt: Je höher die Konfidenz $1 - \alpha$, desto größer muß das Intervall bei der Schätzung gewählt werden. Zur Festlegung von Intervallgrenzen zu Stichproben verwendet man Quantile bekannter Verteilungen. Beispiele s. Fahrmeier Kap. 9.4

6 Testen von Hypothesen

Mit statistischen Tests prüft man, ob es ausreichend plausibel ist anzunehmen, dass eine bestimmte Hypothese für eine Verteilung oder mehrere Verteilungen zutrifft, oder nicht. Die Hypothese kann darin bestehen, daß überhaupt eine bestimmte Verteilung vorliegt, daß ein Parameter einen bestimmten Wert hat, dass zwei Verteilungen gleich sind, etc. etc. In der Praxis sind Hypothese und Gegenhypothese nicht als symmetrisch oder "gleichberechtigt" anzusehen.

Beispiel: In einem Test wird ein neues Medikament gegen ein altes und bewährtes getestet. Soll man das neue nehmen, wenn es etwas besser abschneidet? Die Einführung eines neuen Medikaments ist mit Risiken verbinden. Das Testergebnis könnte Zufall sein! Die Umstellung der Produktion und Vermarktung ist ein einschneidender Prozess. Es könnte neue Nebenwirkungen eintreten.. Man wird

sich nur für das neue Medikament entscheiden wollen, wenn es mit sehr hoher Konfidenz tatsächlich besser ist.

Beispiel: Man vermutet, dass in Deutschland mehr Jungen als Mädchen geboren werden. Wenn man nun 100 Geburten in einem Krankenhaus nimmt, und 55 Jungen sind, wie ist das Ergebnis zu werten?

Beim Testen hat man meist eine "kühne" Hypothese, an man nicht einfach beim ersten Anhaltspunkt glaubt, sondern die man erst als "bestätigt" betrachtet, wenn die Ergebnisse fast keinen anderen Schluss zulaßen. Als "Skeptiker" glaubt man zunächst an die sogenannte

"*Nullhypothese* H_0" (neues Medikament ist nicht besser, es gibt nicht mehr Jungen..) Die (interessantere) "*Alternativhypothese*" H_1 muss sich gegen die Nullhypothese durchsetzen, bevor man sie akzeptiert.

Beim Testen legt man ein Signifikanzniveau α fest (kleiner Wert, z.B. $\alpha = 0,01$). Wenn man eine Teststatistik (eine Art Zählverfahren) festgelegt hat, legt man die Schranke, ab der die Alternativhypothese angenommen wird, so fest, daß die Wahrscheinlichkeit eines "Fehlers 1. Art" kleinergleich α ist.

Fehler 1. Art: Obwohl die Nullhypothese richtig ist, wird die Alternativhypothese angenommen.
Fehler 2. Art: Obwohl die Alternativhypothese richtig ist, wird die Nullhypothese nicht verworfen.

In der Praxis ist der Fehler 1. Art sehr oft sehr viel schwerwiegender, daher legt man hier eine kleine obere Schranke α für die Fehlerwahrscheinlichkeit fest.

Beispiel für einen Test (Jungen versus Mädchengeburten) Falls die Nullhypothese H_0 (es gibt NICHT mehr Jungen als Mädchen) zutrifft, hat man eine Binomialverteilung mit π (Anteil Jungen) maximal $0,5$. Beobachtet man beim Testen n Geburten, so kann man VORHER eine (möglichst große) Menge möglicher Ergebnisse

Zahl der Jungen $= n$ von n,
Zahl der Jungen $= n - 1$ von n,
Zahl der Jungen $= n - 2$ von n,
...

so festlegen, dass deren Gesamtwahrscheinlichkeit für eine Binomialverteilung $B(n, 0, 5)$ die Schranke α nicht übersteigt. Dies seien die Ergebnisse $m, ..., n$. Man akzeptiert die Alternativhypothese H_1, wenn beim tatsächlichen Test dann eines dieser Ergebnisse eintritt. Klar: Je kleiner die Signifikanzschranke α, desto größer ist m.

Beispiel Fahrmeier: für $n = 10$ und Signifikanzniveau $0,1$ kann man die Nullhypothese verwerfen, wenn unter den 10 Geburten 8, 9 oder 10 Jungen sind.

Einseitige Test versus zweiseitige Tests

Eine andere Alternativhypothese H_1 wäre einfach nur, dass NICHT gleichviele Jungen wie Mädchen geboren werden. Dann kann die Nullhypothese sowohl für sehr kleine Zahl von Jungen $0, 1, ..$ wie auch für sehr große Werte $n, n-1, ..$ verworfen werden. Man geht wieder von $B(n, 0, 5)$ aus. Die Gesamtwahrscheinlichkeit der Ergebnisse
$0, 1, ..., m_u$
$n, n-1, .., m_o$
darf das Signifikanzniveau nicht übersteigen.

Probleme beim Testen

Man kann nicht immer Binomialverteilung zur Festlegung des Akzeptanzbereichs nehmen. Mit welcher Art von Statistik (Auswertung) und welcher Verteilung erhält man einen brauchbaren Test?

Anwendungen im linguistischen Bereich

Man untersucht mit derselben Art von Test viele linguistische Einheiten (Wörter, Phrasen,...). Die skeptische Annahme H_0 kann lauten, dass eine solche Einheit (für ein bestimmtes Thema) uninteressant ist, dass zwei Wörter nichts miteinander zu tun haben, etc. etc. Je klarer die Nullhypothese verworfen werden kann, desto interessanter "auffälliger" sind die Einheiten.

Tests - Beispiele

Nach Woods Fletcher Hughes: Statistics in language studies, S. 122-126.

Bei einer standardisierten Prüfung erreichen die Studenten erfahrungsgemäß im Durchschnitt 80 Punkte. In einem Jahr gibt es einen längeren Streik. Die Frage ist, ob sich dieser auf die Leistungsfähigkeit negativ niedergeschlagen hat. Soll durch einen Test geklärt werden.

Hierzu wird ein Sample von 10 Studenten zufällig gewählt. Die Probanden erreichen in der Prüfung in diesem Jahr eine durchschnittliche Punktezahl von $\bar{X} = 71, 5$ bei einer Standardabweichung von $s = 13, 18$.

Skeptische Nullhypothese H_0: Populations-Durchschnittswert ist $\mu = 80$.
Alternativhypothese H_1: $\mu < 80$.

Brauchen eine Teststatistik mit bekannter Verteilung, bei der wir die Quantile zur Grundlage der Entscheidung machen können. Es ist

$$\frac{\bar{X} - \mu}{s/\sqrt{n}}$$

t-verteilt (t-Verteilung) mit $n-1 = 9$ (10 Teilnehmer - 1) Freiheitsgraden. (Ähnlich zur Standardnormalverteilung, aber flacher). Im vorliegenden Fall erhalten wir als Wert der Statistik auf der Stichprobe

$$\begin{aligned} \frac{\bar{X} - \mu}{s/\sqrt{n}} &= (71,5 - 80)/(13,18/\sqrt{10}) \\ &= -2,04 \end{aligned}$$

Das linksseitige 5% Quantil der t-Verteilung mit 9 Freiheitsgraden ist beim Wert $-1,83$. Da der Testwert kleiner ist, kann die Nullhypothese mit Signifikanzlevel 5% verworfen werden.

Zweites Beispiel S. 139 bis S. 142 Kopien.

7 Markov-Modelle und Sprachmodelle

7.1 Markov-Modelle

7.2 Sprachmodelle

7.3 Glättung (smoothing)

8 Hidden-Markov-Modelle