Statistik-Skript

Hans-Christian Heinz
hheinz@imn.htwk-leipzig.de

WS2013/14

Teil I. Wahrscheinlichkeitsrechnung

1. Wahrscheinlichkeiten

1.1. Zufällige Ereignisse

Bedingungen für Experiment gegeben.
Versuch (Experiment): Realisierung der Bedingungen.
Ereignis: Ergebnis eines Experiments - soll beobachtet werden

- heißt notwending wenn es zwingend eintritt
- heißt zufällig, wenn es eintreten kann oder nicht

1.2. Relative Häufigkeiten, Wahrscheinlichkeitsbegriff

E - Menge von zufälligen Elementarereignissen
$e \in E$ - Menge von Bedingungen, die beim Versuch realisiert werden
Ereignis A: Teilmenge von E,$A \subseteq E$

Definition: *relative Häufigkeit*

A,B Ereignisse

n Anzahl ausgeführter Experimente

k Anzahl der Experimente bei denen A beobachtet wird

$\mathbf{A \cap B}$ das Ereignis, dass A und B beobachtet werden

$\mathbf{A \cup B}$ das Ereignis, dass A oder B beobachtet werden

relative Häufigkeit $h_n(A) = \frac{k}{n}$

Eigenschaften von h_n

1. $0 \leq h_n(A) \leq 1$
2. $h_n(E) = 1$
3. Es sei $A \cap B = \emptyset$

$$h_n(A \cup B) = h_n(A) + h_n(B)$$

Für $n \to \infty$: $h_n(A)$ stabilisiert sich.

W Formalisierung nach Kolmogorow

MA Sigma-Mengenalgebra von Teilmengen von E, d.h. $MA \subseteq 2^E$ = Menge aller Teilmengen von E
und $E \in MA$

$$A, B \in MA \to A \cup B \in MA \wedge A \setminus B \in MA$$

$$\forall : A_i \in MA \to \bigcup_{i=1}^{\infty} A_i \in MA$$

MA soll „die interessierenden" Ereignisse enthalten.
Bemerkung: $\bar{A}$ von A und $A \cap B$ sind dann auch in MA
Beweis: $\bar{A} = E \setminus A, A \cap B = \overline{\bar{A} \cup \bar{B}}$
Kolmogorow'sche Axiome (1933): $P : MA \to R$ mit:

1. $\forall A : A \in MA \to P(A) \geq 0$
2. $P(E) = 1$
3. $\{A_i\}$ geg. ($A_i \in MA$) mit $\forall i, j : i \neq j \to A_i \cap A_j = \emptyset$
   Dann $P(\bigcup_{i=1}^{\infty} A_i = \sum_{i=1}^{\infty} P(A_i)$

Definition: $[E, MA, P]$ heißt W-Raum

Beispiel:
$E = \{e_1, e_2, \dots, e_n\}$

$MA \subseteq 2^E = \{\underbrace{\{\}, \{e_1\}, \dots, \{e_n\}, \{e_1, e_2\}, \dots, \overbrace{\{e_1, \dots, e_n\}}^{E}}_{2^n \, Elemente}\}$

$p_1, p_2, \dots .p_n$ seien gegeben mit $\forall i : p_i \geq 0$ mit $\sum_{i=1}^{n} p_1 = 1$
A zufälliges Ereignis
$\hookrightarrow P(A) = \sum_{i:e_i \in A} p_i$

Wichtiger Spezialfall: Laplace'scher W-Raum
alle Versuchsausgänge gleichwertig, d.h. $\forall i : p_i = \frac{1}{n}$
$\hookrightarrow P(A) \frac{\text{Anzahl der } e_i \text{, die A bewirken}}{n}$
Mit diesen P heißt $[E, MA, P]$ Laplace'scher W-Raum.

Wiederholung:

Ereignis und Elementarereignis

Ereignis: Ergebnis eines Experiments (im Modell: Menge von Elementarereignissen)
Elementarereignis: Satz von Bedingungen (im Modell: nicht näher bestimmt)

Warum Sigma-Mengenalgebra Zur vollständigen Erfassung (mathematischen Modellierung) von Wahrscheinlichkeiten; um auch verm. uninteressante Ereignisse zu beinhalten; damit man z.B. das Ereignis „nicht 3“ berechnen kann, indem man $1 - p(3)$ bildet Vorlesung

1.3. Unabhängige Ereignisse

A,B (aus MA) heißen unabhängig, wenn

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

.
A = {Professer, 7:30, pünktlich zur Vorlesung}
B = {Morgen Milchreis in der Mensa}
C = {Morgen Gewitter}
D = {Stromausfall}
$rightarrow$ nicht modellierbar/entscheidbar Ein Ereignis ist immer stark abhängig von seinem Gegenteil!
Beispiel für Unabhängigkeit:
2 Würfel, A = {Gerade Augenzahl}, {B = Augenzahl≥2}

Insgesamt unabhängige Ereignisse
Ereignis $A :\in MA$ $(i = 1, \dots, n)$ heißen insgesamt unabhängig, wenn für jede Indexmenge $\{i_1, \dots, i_k\} \subseteq \{1, \dots, b\}$ gilt:

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) * \dots * P(A_{i_k})$$

Definition: $[E, MA, P]$ sei gegebener W-Raum
Es gelte $B \in MA$ und $P(B) > 0$
Bedingte W von A unter Bedingung B:

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

Bemerkung: A,B unabhängig

$$\hookrightarrow P(A|B) = \frac{P(A \cap B)}{P(B)} \stackrel{!}{=} \frac{P(A) * P(B)}{P(B)} = P(A)$$

Satz: $P(A|B)$ ist eine W, d.h. gibt

(a) $\forall A \quad A \in MA \rightarrow P(A|B) > 0$

(b) $P(E|B) = 1$

(c) $P(\bigcup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} P(A_i|B)$
bei $\forall i, j$ $j \neq i \rightarrow A_i \cap A_j \neq \emptyset$

Beweis:

(a) $A \cap B \in MA \rightarrow P(A \cap B) \geq 0 \rightarrow \frac{P(A \cap B)}{P(B)} \geq 0$

(b) $P(E|B) = \frac{P(E \cap B}{P(B)} = \frac{P(B)}{P(B)} = 1$

(c) $P(\bigcup_{i=1}^{\infty} A_i|B) = \frac{P([\bigcup_{i=1}^{\infty} A_i] \cap B)}{P(B)} = \frac{P([\bigcup_{i=1}^{\infty} A_i \cap B])}{P(B)}$
alle $A_i \cap B$ sind <u>durchschnittsfremd</u>
$= \frac{\sum_{i=1}^{\infty} P(A_i \cap B}{P(B)} = \sum_{i=1}^{\infty} P(A_i|B)$

Definition: $[E, MA, P_B]$ heißt bedingter W-Raum

1.4. Sätze zu bedingten Wahrscheinlichkeiten

Multiplikationssatz:

$A_1, \dots, A_n \in MA$
Dann gilt

$$P(\bigcap_{i=1}^{\infty} A_i) = P(A_1) * P(A_2|A_1) * \dots * P(A_n|\bigcap_{i=1}^{\infty} A_i)$$

Beweis: n=2, dann vollständige Induktion
Beispiel: Urnenmodell, Kugelnziehen ohne Zurücklegen
Urne enthalte 2 weiße, 3 rote und 5 schwarze Kugeln
$A_1 = \{1.\text{ Kugel ist weiß}\}$
$A_2 = \{2.\text{ Kugel ist weiß}\}$
$A_3 = \{3.\text{ Kugel ist schwarz}\}$
$P(A_1 \cap A_2 \cap A_3)$
$= P(A_1) * P(A_2|A_1) * P(A_3|A_1 \cap A_2)$
$= \frac{1}{5} * \frac{1}{9} * \frac{5}{8} = \frac{1}{72}$

Definition: Die (unendliche) Folge $\{A_i\}$ heißt vollständiges Ereignis, wenn $\forall i\ A_i \in MA$
$\forall i \forall j\ i \neq j \rightarrow A_i \cap A_j = \emptyset$
$\bigcup_{i=1}^{\infty} A_i = E$
E wird vollständig durch die Folge $\{A_i\}$ beschrieben

Satz von der totalen Wahrscheinlichkeit:
$A \in MA$
$\{B_i\}$ - vollständiges Ereignissystem $\forall i\ B \in MA; P(B_i) > 0$
Dann gilt:

$$P(A) = \sum_{i=1}^{\infty} P(A \cap B_i) = \sum_{i=1}^{\infty} P(A|B_i) * P(B_i)$$

Beweis:

$$A = A \cap E = A \cap (\bigcup_{i=1}^{\infty} B_i)$$

B_i durchschnittsfremd
$\hookrightarrow$ ($A \cap B_i$ auch durchschnittsfremd
Also:

$$P(A) = \sum_{i=1}^{\infty} P(A \cap B_i)$$

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

Bsp.: 3 Urnen U_i mit schwarzen, roten, weißen Kugeln
U_1: 3s, 3r, 2w   P(Aus U_1 wird gezogen) = 0.3
U_2: 1s, 2r, 1w   P(Aus U_2 wird gezogen) = 0.2
U_3: 4s, 0r, 6w   P(Aus U_3 wird gezogen) = 0.5
Wie groß ist W, dass eine rote Kugel gezogen wird?

$$P(r) = P(r|U_1) * P(U_1) + P(r|U_2) * P(U_2) + P(r|U_3) * P(U3)$$
$$P(r) = \frac{3}{8} * 0.3 + \frac{1}{2} * 0.2 + \frac{0}{10} * 0.5 = \frac{17}{80}$$

Ursachensatz (Satz von Bayes)

Voraussetzungen wie zuvor

$$P(B_i|A) = \frac{P(B_i) * P(A|B_i)}{\sum_{j=1}^{\infty} P(A|B_j) * P(B_j)} \leftarrow \text{Nenner } = P(A)$$

Beweis:

$$P(B_i \cap A) = P(A \cap B_i)$$

$$P(B_i|A) * P(A) = P(A|B_i) * P(B_i)$$

$$P(B_i|A) = \frac{P(A|B_i) * P(B_i)}{P(A)}$$

Urnenbeispiel wie zuvor:

$$P(U_i|r)\text{zu berechnen}$$

$$P(U_1|r) = \frac{P(U_1) * P(r|U_1)}{P(r)} = \frac{0.3 * \frac{3}{8}}{\frac{17}{80}} = \frac{9}{17}$$

Lösung Übungsaufgabe vom 18.10.2013

$$P(A_1 \cap A_2) = \frac{9}{36} = \frac{1}{4} = \frac{1}{2} * \frac{1}{2} = P(A_1) * P(A_2)$$

$$P(A_1 \cap A_3) = \frac{1}{36} = \frac{1}{2} * \frac{1}{18} = P(A_1) * P(A_3)$$

$$P(A_2 \cap A_3) \text{ analog}$$

$$P(A_1 \cap A_2 \cap A_3) = 0 \neq P(A_1) * P(A_2) * P(A_3)$$

Jeweils paarweise unabhängig, aber insgesamt nicht unabhängig

2. Zufallsgrößen (Zg)

2.1. Grundlegende Definitionen

Motivation:
X - Zg, d.h. ein Zahlenwert nach einem Experiment (noch zu präzisieren)

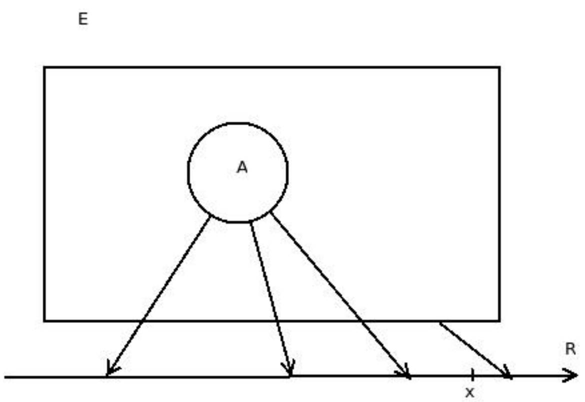
Von Interesse: Ereignisse
$\{e : X(e) \leq x\}$
$\{e : a < X(e) < b\}$

Definition: Mengensystem $\{(-\infty, x] | x \in \mathbb{R}\}$ wird betrachtet
BM sei die hierdurch erzeugte Sigma-Mengenalgebra (kleinste Sigma-Mengenalgebra, die obiges Mengensystem enthält). Die Elemente von BM heißen Borel-Mengen.

Beispiele von Borel-Mengen
$\{x: \ a < x \leq b\} = \{x: \ x \leq b\} \setminus \{x: \ x \leq a\}$
$\{b\}$, es sei $\{a_i\}$ monoton wachsend mit $\lim\limits_{i \to \infty} a_i = b$
$\{b\} = \bigcap_{i=1}^{\infty} \{x: \ a_i < x \leq b\}$
$\{x: \ a < x \leq b\} \cup \{x: \ c < x \leq d\}$

Zg:
$[E, MA, P]$ - W-Raum
R - Menge der reellen Zahlen
Es sei $B \in BM$
$X : E \to R$ mit $\{e : X(e) \in B\} \in MA$
bzw. $X^{-1}(B) \in MA$
Dann heißt X eine Zg.

Bemerkungen:
Die Abbildung X heißt dann [MA,BM] - messbar.
Die Messbarkeit von [E,MA,P] auf [R,BM] übertragen
Also: Eine Zufallsgröße ist eine borel-messbare Abbildung.



Verteilungsfunktion (Vf) zu einer Zg X:

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

$$\stackrel{kurz}{=} P(X \leq x)$$

Bemerkung: Wegen der Messbarkeit von X lässt sich zeigen, dass bei bel.$B \in BM$ gilt:

$$\{e \ X(e) \in B\} \in MA$$

Also existiert F(x).
Sprechweise: X ist verteilt nach F(x). Oder kurz: X $\sim$ F(x).
Es gilt immer

$$\underset{x \to -\infty}{F}(x) = 0$$

$$\underset{x \to \infty}{F}(x) = 1$$

F(x) ist monoton wachsend.

$F(x) = P(x \leq x)$ sei bekannt
$P(a < X \leq b)$ zu berechnen
$F(b) = P(e : X(e) \leq b)$
$= P(\{e : X(e) \leq a\} \cup \{e : a < X(e) \leq b\})$
$\underbrace{= P(X \leq a)}_{F(a)} + P(a < X \leq b)$
$\hookrightarrow P(a < X \leq b) = F(b) - F(a)$

Definition: Verteilungsfunktion F(x) gegeben, überall differenzierbar

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

heißt Dichtefunktion zur Zufallsgröße X.
$\hookrightarrow P(X \leq x) = \int_{-\infty}^{x} f(x)dx$

Definition: Zufallsgröße X,Y heißen mathematisch unabhängig, wenn

$$\forall x \forall y \; P(X \leq x \wedge Y \leq y) = P(X \leq x) * P(Y \leq y)$$

Bsp.: X,Y - Augenzahlen von 2 Würfeln
$Z = X + Y$, dann sind X und Z abhängig.
Körpergröße, Schuhgröße eines zufällig ausgewählten Menschen abhängig.
(Korrelation ist hinreichend, aber nicht notwendig!)

Diskrete Zufallsgröße X: X kann nur Werte annehmen aus einer Folge $\{\ldots, x_{-1}, x_0, x_1, x_2, x_3, \ldots\}$

Satz: Jede Verteilungsfunktion lässt sich in der Form

$$F(x) = a_1 F_1(x) + a_2 F_2(x) + a_3 F_3(x)$$

darstellen, wobei

(a) $a_i \geq 0,\ a_1 + a_2 + a_3 = 1$

(b) $F_1(x)$ ist absolutstetig, d.h. es gibt eine Funktion f(x) mit

$$F_1(x) = \int_{-\infty}^{x} f(y)dy \text{ (Lebesque Integral)}$$

Riemann-Integral ist Spezialfall des Lebesque-Integrals
d.h.

$$F'(x) = f(x) \text{ fast überall}$$

(abgesehen von einer Menge vom Lebesque-Maß 0)

(c) $F_2(x)$ eine diskrete Verteilungsfunktion it (reine Sprungfunktion)

(d) $F_3(x)$ singulär ist, d.h.
$F'(x) = 0$ fast überall und stetig.

2.2. Das Stieltjes-Integral

Motivation: Einf., um alle Verteilungsfunktionen einheitlich behandeln zu können. Verallgemeinerung des bekannten Riemann-Integrals.
Idee: Zu integrierende Funktion wird „gewichtet“ integriert (nicht alle x-Werte sind gleich wichtig)
Gegeben seien Funktionen g(x) und F(x) und ein Intervall [a,b]. F(x) sei stetig in a, ansonsten rechtsstetig.
g(x) sei stetig.

Definition: $\underbrace{Z(n,x)}_{\text{Vektor}(x_0 \ldots x_n)}$ sei eine Zerlegung von $[a,b]$
folgender Art::

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b$$

Es sei

$$\underbrace{V(n,x)}_{\text{abhängig von n und Vektor x}} = \sum_{i=1}^{n} |F(x_i) - F(x_{i-1})|$$

$$V = \sup_{Z(n,x)} \{V(n,x)\} = \text{kleinste obere Schranke bez. aller Zerlegungen von Z(n,x)}$$

V heißt totale Variation von F(x) in $[a,b]$
Bsp.:

$$F = \sin x \text{ in } [0, 2\pi]$$

$$V = 4$$

Definition: des Riemann-Stieltjes-Integrals
Es sei

$$S(n,x) = \sum_{i=1}^{n} g(x_i) * |F(x_i) - F(x_{i-1})|$$

$$\lim_{n\to\infty} S(n,x) = \int_a^b g(x)dF(x)$$

Satz: Grenzwert eindeutig bestimmt, unabhängig von Art der Zerlegung

Uneigentliches Stieltjes-Integral

$$\int_{-\infty}^{\infty} g(x)dF(x) = \lim_{a\to-\infty}\lim_{b\to\infty}\int_a^b \dots$$

falls Grenzwert existiert.

Satz: g(x) stetig und beschränkt auf $(-\infty,\infty)$
F(x) von beschränkten Totalvariation auf $(-\infty,\infty)$
$\hookrightarrow \int_{-\infty}^{\infty} \dots$ ex.

Rechenregeln:

(a) Es existiere $F'(x) = \frac{dF(x)}{dx} = f(x)$
Denn

$$\int_a^b g(x)dF(x) = \int_a^b g(x)\frac{dF(x)}{dx}dx$$
$$= \int_a^b g(x) * f(x)dx$$

(b)

$$\int_a^b [g_1(x) + g_2(x)]dF(x) = \int_a^b g_1(x)dF(x) + \int_a^b g_2(x)dF(x)$$

(c)

$$\int_a^b g(x)d(F_1+F_2) = \int_a^b g(x)dF_1 + \int_a^b g(x)dF_2$$

(d) k_1, k_2-Konst.

$$\int_a^b [k_1 * g]d[k_2 * F] = k_1 * k_2 \int_a^b gdF$$

(e)

$$\int_a^b gdF = \int_a^c \dots + \int_c^b \dots$$

(f) Partielle Integration

$$\int_a^b g(x)dF(x) = g(b) * F(b) - \int_a^b F(x)dg(x)$$

Ausblick
g(x) kann abzählbar unendlich viele Unstetigkeitsstellen haben, keine fällt mir einer Unstetigkeitsstelle von F(x) zusammen.
Zusammenfallen einiger Unstetigkeitsstellen → Lebesque-Stieltjes-Integral.

2.3. Kennwerte von Zufallsgrößen

X sei Zufallsgröße, $X \sim F(X)$

Erwartungswert:

$$EX = \int_{-\infty}^{\infty} xdF(x) \overset{\text{Nur bei Ex. von f(x)}}{=} \int_{-\infty}^{\infty} x * f(x)dx$$

Interpolation: Mittelwert der Versuchsausgänge
Bsp.: Würfel Verteilungsfunktion F(x)

$$EX = \int_{-\infty}^{\infty} xdF(x) = 1 * \frac{1}{6} + 2 * \frac{1}{6} + \cdots + 6 * \frac{1}{6} = 3.5$$

Andere Schreibweise

$$EX = \int_{-\infty}^{\infty} xdF(x) = \int_E X(e)dP(de)$$

Integration über die Menge der Elementarereignisse
Regel: $a, b \in \mathbb{R} : X, Y$ Zufallsgröße

$$E(a * X + b * Y) = a * EX + b * EY$$

Varianz:

$$\begin{gathered} VarX = E(X - EX)^2 \\ = E(X^2 - 2 * X + EX + (EX^2)) \\ = EX^2 - (EX)^2 \\ = \int_{-\infty}^{\infty} (x - EX)^2 dF(x) \end{gathered}$$

a sei Konstante

$$Var(a * X) = E(a * X - a * EX)^2 = a^2 E(X - EX)^2 = a^2 * VarX$$

Streuung

$$StrX = \sqrt{VarX}$$

Beispiel: $F(x) = 0$ für $x \leq 0; x$ für $x \in (0,1]; 1$ für $x > 1$
EX, VarX berechnen

Linearer Zusammenhang von Zufallsgrößen

Kovarianz:

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

Bemerkung 1: $Cov(X,X) = Var(X)$

Bemerkung 2: $Cov(X,Y) = E(X * Y) - EX * EY$

Korrelation:

Idee: 1-Normierung

$$Korr(X,Y) = \frac{E[(X - EX) * (Y - EY)]}{\sqrt{E(X - EX)^2 * E(Y - EY)^2}}$$

Kovarianz im Zähler, Streuung im Nenner

Satz: $-1 \leq Korr(X,Y) \leq 1$

Es gilt

$$Var(X + Y) = E(X + Y - EX - EY)^2$$
$$= E((X - EX) + (Y - EY))^2$$
$$= E[(X - EX)^2 + 2 * (X - EX) * (Y - EY) + (Y - EY)^2]$$
$$= VarX + VarY + 2 * Cov(X,Y)$$

X,Y unabhängig $\rightarrow Cov(X,Y) = 0, Korr(X,Y) = 0$
Umkehrung gilt im Allgemeinen nicht
Ausnahme: X,Y normalverteilt

Tschebyschew'sche Ungleichung

$$P(|X - EX| \geq x * StrX) \leq \frac{1}{x^2}$$

2.4. Beispiele für Verteilungsfunktion

(a) Diskrete Verteilungen

Einpunkt-Verteilungen F(X)
Es existiert Zahl a mit

$$P(X = a) = 1$$

$$\int_{a-0}^{a} xdF(x) \underset{\text{Intergrand an Sprungstelle mal Sprunghöhe}}{=} a$$

$$\int_{a-0}^{a} (x-a)dF(x)$$

$$= (a-a) * 1 = 0$$

Gleichmäßige Verteilung
n mögliche Versuchsausgänge $\{1, \ldots, n\}$ alle gleichwahrscheinlich, d.h.

$$P(X = k) = \frac{1}{n} \qquad k = 1, \ldots, n$$

$$P(X \leq k) = \sum_{i=1}^{k} P(X = k) = \frac{k}{n}$$

$$EX = \frac{n+1}{2}$$

Binomialverteilung
Ausgangspunkt: Benoulli'sches Schema
A gegeben $(A \in MA) \quad p = P(A) > 0$
X - Anzahl des Eintretens von A bei n unabhängigen Versuchen

$$B(n,k))P(e : X = k) = \binom{n}{k} * p^k * (1-p)^{n-k}$$

$$P(X \leq x) = \sum_{k:k \leq x} B(n,k)$$

$$EX = n + p \qquad VarX = n * p * (1-p)$$

Geometrische Verteilung
$A \in MA$ - interessierendes Ereignis
X - Anzahl von unabhängigen Versuchen bis erstmals A eintritt
{X=k} bedeuetet: (k-1) Fehlversuche A im k-ten Versuch
Geg.: $p = P(A) > 0$

$$P(X = k) = (1-p)^k * p \text{ mit } k = 1, 2, \ldots$$

$$\hookrightarrow EX = \frac{1}{p} \qquad VarX = \frac{1-p}{p^2}$$

(b) Stetige Verteilungen

Gleichverteilung

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{für } x \leq a \\ \frac{x-a}{b-a} & \text{für } x \in (a,b] \\ 1 & \text{für } x > b \end{cases}$$

$$F'(x) = f(x) = \frac{1}{b-a} \text{ für } (a,b)$$

$$EX = \frac{a+b}{2} \quad VarX = \frac{(b-a)^2}{12}$$

Wichtiger Spezialfall: $a = 0, b = 1$

Normalverteilung

Zufallsgröße X heit normalverteilt mit Parametern μ und σ, wenn x folg. Dichte hat:

$$nv(x,\mu,\sigma) = \frac{1}{\sigma * \sqrt{2 * \pi}} * e^{-\frac{(x-\mu)^2}{2*\sigma^2}}$$

$$\int_{-\infty}^{\infty} nvdx = 1$$

$$NV(x,\mu,\sigma) = \int_{-\infty}^{x} nv(x,\mu,\sigma)dx$$

$$EX = \mu \quad VarX = \sigma^2$$

Wichtiger Spezialfall:

$$\mu = 0, \quad \sigma = 1 \quad \text{Standard-Normalverteilung}$$

Student-Verteilung (W.S. Gosset, engl. Statistiker, Pseudonym: Student)

Vorbereitung: Gamma-Funktion (Euler 1730)

$$\Gamma(x) = \int_{0}^{\infty} y^{x-1} * e^{-y} dy$$

Bemerkenswert: Es gilt fü $x > 1$ die Funktionalgleichung

$$\Gamma(x) = (x-1) * \Gamma(x-1)$$

$\Gamma(1) = 1$, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Für natürliches x ist das die Fakultätsfunktion $(x-1)!$

Dichte der Student-Verteilung für $x \in \mathbb{R}$ (mit n Freiheitsgraden)

$$stu(x,n) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi * n}} * (1 + \frac{x^2}{n})^{-\frac{n+1}{2}}$$

symmetrisch bezüglich y-Achse
$Stu(x.n)$ zugehörige Verteilungsfunktion

Satz: $\lim\limits_{n\to\infty} Stu(x,n) = NV(x,0,1)$
$X \sim Stu$
$n \geq 2 \to EX = 0 \qquad n \geq 3 \to VarX = \frac{n}{n-2}$

χ^2-**Verteilung**
Dichte

$$chiQ(x,n) = \begin{cases} \frac{1}{2^{\frac{n}{2}} * \Gamma(\frac{n}{2})} * x^{\frac{n}{2}-1} * e^{-\frac{x}{2}} & \text{, für } x > 0 \\ 0 & \text{, für } x \leq 0 \end{cases}$$

Zugehörige Verteilungsfunktion ChiQ(x.n)
Bem. Es gibt auch χ-Verteilungen. Falls X eine χ-Verteilung hat, hat X^2 eine χ^2-Verteilung

2.5. Simulation von Zufallsgrößen

$F(x) = P(X \leq x)$ sei bekannt.

Aufgabe: konkrete Stichprobe $x_1, \dots, x_n$ zu erzeugen (n unabhängige Experimente mit X)
Vorüberlegung:
Y sei auf $[0,1)$ gleichverteilte Zufallsgröße, d.h.

$$G(x) = P(Y \leq x) = x \text{ für } x \in [0,1)$$

Satz: F(x) sei stetig und streng monoton wachsend. Dann gilt für die Zufallsgröße $F^{-1}(Y)$

$$P(F^{-1}(Y) \leq x) = F(x)$$

Beweis:

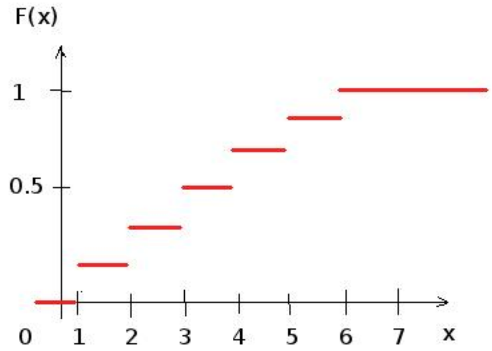
$$P(F^{-1}(Y) \leq x) = P(Y \leq F(x)) = F(x)$$

Realisierungen y_i von Y sind per Zufallsgenerator erzeugbar
$F^{-1}(y_i)$ ist die i-te Realisierung von X.
Lösung obiger Aufgabe:
y_i sei erzeugt

(a) Setze $x_i = F^{-1}(y_i)$ falls die rechte Seite existiert und eine Zahl ist.

(b) Setze $x_i = a$, falls $F^{-1}(y_i)$ ein Intervall $[0,1)$ ist

(c) Setze $x_i = \min\{x : F(x) > y_i\}$, falls $F^{-1}(y_i)$ nicht existiert.



Beispiel: Würfel-Verteilungsfunktion

(b) Es sei $y_i = 0.5$
$F^{-1}(y_i) = F^{-1}(0.5) = [3, 4)$
$\hookrightarrow x_i = 3$

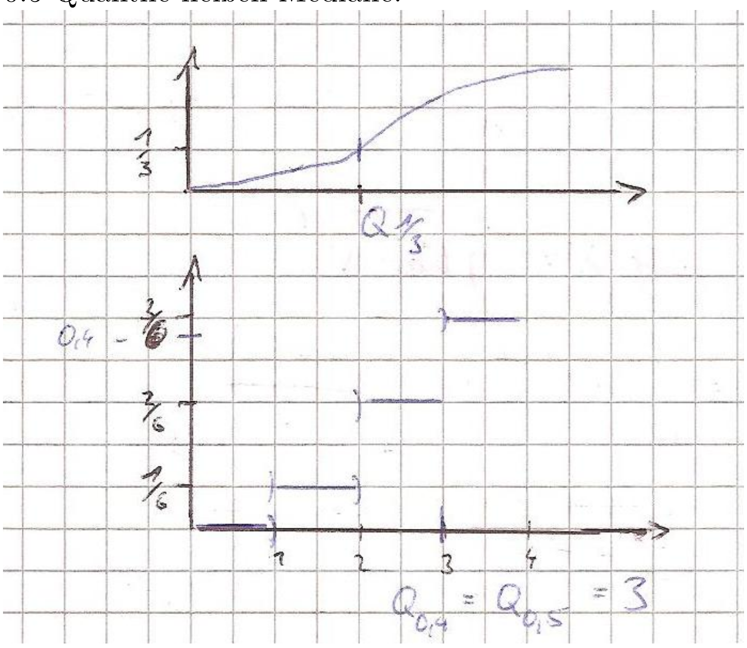
(c) Es sei $y_i = 0.51$
$\min x : F(x) > 0.51 = 4$

2.6. Quantile

Zufallsgröße X, $x \sim F(x)$
Es sei $0 \leq p \leq 1$. Jede Zahle Q_p, die der Bedingung

$$F(Q_p - 0) \leq p \leq F(Q_p)$$

genügt, heißt p-Quantil von F(x).
0.5-Quantile heißen Mediane.



Für Würfel-Verteilungsfunktion: 0.4-Quantil ist genau die 3, 0.45-Quantil auch.

Jede Zahl aus $[3, 4]$ ist Median, insbesondere auch die 4.
Bemerkung: F(X) ist streng monoton → alle Quantile eindeutig.

3. Funktionen von Zufallsgrößen und Zufallsvektoren

3.1. Funktionen einer einzelnen Zufallsgröße

Gegeben: W-Raum $[E, MA, P]$, Zg X, Funktion g(x)
Bilde g(X). Laut Zufallsgrößen-Definition müsste g(X(e)) borel-messbar sein. X selbst ist borel-messbar. X bildet Ereignisse $A \in MA$ in Sigma-Mengenalgebra BM ab.

Definition: g(x) heißt borel-messbar, wenn für jedes $B \in BM$ gilt:

$$g^{-1}(B) \in BM$$

Satz: Ist die Funktion g(x) stückweise stetig, dann ist sie borel-messbar.

Beispiel: Quadrat einer Zufallsgröße
X sei positive Zufallsgröße $X \sim F(x)$

$$P(X^2 \leq x) = P(X \leq \sqrt{x}) = F(\sqrt{x})$$

3.2. Zufallsvektoren

Borel-Messbarkeit ist auf Funktionen mehrerer Variabler übertragbar.
Folgende Funktionen sind borel-messbar:

- Vielfache
- Summen
- Produkte
- Quotienten (falls def.)
- Maximum, Minima
- Infikatorfunktionen
- Grenzwerte von Folgen einer Zufallsgröße

Definition: $X_1, \ldots, X_n$-Zufallsgröße über $[E, MA, P]$

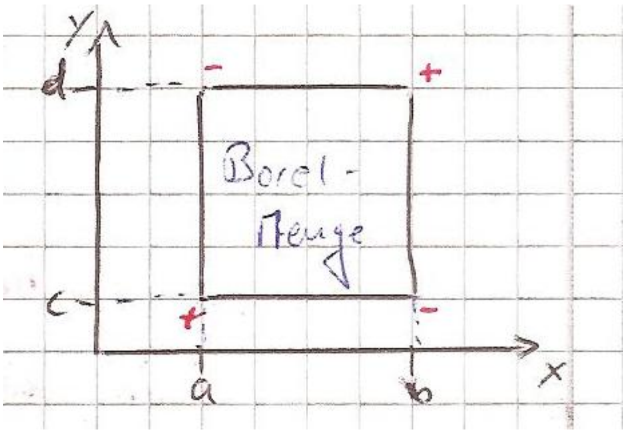
$$F(x_1, \ldots, x_n) = P(X_1 \ \ldots, X_n \leq x_n)$$

heißt gemeinsame Verteilungsfunktion der Zufallsgröße.
Die $P(X_i \leq x_i)$ heißen Randverteilungen.

F monoton wachsend bezüglich jedes x:
$n = 2$ $\quad P(X \leq x, Y \leq Y) \rightarrow P(X \leq x) = \lim_{y \to \infty} P(X \leq x, Y \leq y)$

Berechnung von

$$P(a < X \leq b, c < Y \leq d)$$
$$= \underbrace{F(b,d) - F(b,c) - F(a,d) + F(a,c)}_{\geq 0 \text{ , sonst ist F keine Verteilungsfunktion}}$$



Definition:

$$f(x_1, \ldots, x_n) = \frac{\delta^n F(x_1, \ldots, x_n)}{\delta x_1 \ldots \delta x_n}$$

heißt Dichtefunktion von $\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$, wenn die Ableitungen existieren und stetig sind (Voraussetzung des Satzes von Schwarz).

Definition: (Fall n=2, leicht verallgemeinerbar)
Gegeben: Zufallsgrößen X,Y: $X \sim F_X(x)$ $Y \sim F_Y(x)$

$$F_X(x|y) = P(X \leq x|Y = y) = \lim_{h \to 0} \frac{P(X \leq x, y - h < Y \leq y + h)}{P(y - h < Y \leq y + h}$$

heißt bedingte Verteilungsfunktion von X unter der Bedingung Y=y, falls der Limes existiert.

Definition:

$$E(X|Y=y) = \int_{-\infty}^{\infty} x dF_X(x)$$

heißt bedingter Erwartungswert unter Bedingung Y=y, falls Integral existiert.
Formeln für bedingte Dichten/bedingte Wahrscheinlichkeits-Verteilungen (diskreter Fall) $\rightarrow$ Literatur.

n-dimensionale Normalverteilung:

Zufallsvektor $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ gegeben.
X_i nicht einpunktverteilt
Erwartungsvektor $EX = \begin{pmatrix} EX_1 \\ \vdots \\ EX_n \end{pmatrix}$ und die Kovarianzmatrix C seien bekannt.

$$EX = \begin{pmatrix} EX_1 \\ \vdots \\ EX_n \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$$

$$C = \begin{pmatrix} Var(X_1) & Cov(X_1, X_2) & \dots & Cov(X_1, X_n) \\ Cov(X_1, X_2) & Var(X_2) & \dots & Cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_1, X_n) & \dots & \dots & Var(X_n) \end{pmatrix}$$

C ist symmetrisch und positiv definit $\rightarrow C^{-1}$ existiert

X heißt normalverteilter Zufallsvektor, wenn er folgende Dichtefunktion hat:

$$nv(x, \mu, X) = \frac{1}{\sqrt{(2\pi)^n * det(c)}} * e^{-\frac{(x-\mu)^T * C^{-1} * (x-\mu)}{2}}$$

3.3. Funktionen von Zufallsvektoren

Beispiel: Summe zweier Zufallsgrößen X und Y.
$\{y_i\}$ sei eine Zerlegung der y-Achse

$$\dots y_{-1} < y_{-1} < y_0 < y_1 < y_2 < \dots$$

$$P(X+Y \leq y) = \sum_{i=-\infty}^{\infty} P(X+Y|Y \in (y_i, y_{i+1}]) * \overbrace{P(Y \in (y_i, y_{i+1}]}^{F_Y(y_{i+1})+F_Y(y_i)}$$

Zerlegungsverfeinerung, Integral-Längen gegen 0:

$$= \int_{-\infty}^{\infty} P(X \leq x-y|Y=y) dP(Y \leq y)$$

falls X,Y unabhängig:

$$= \int_{-\infty}^{\infty} F_X(x-y) dF_Y(y)$$

Faltungsintegral $\overset{\text{falls Existenz der Dichten}}{\rightarrow}$ $f_{x+y}(x) = \int_{-\infty}^{\infty} f_x(x-y) * f_y(y) dy$

4. Grenzwertsätze - die Brücke zur Statistik

4.1. Konvergenzbegriffe für Folgen von Zufallsgrößen

W-Raum $[E, MA, P]$ gegeben; $X, Y, X_1, X_2, \dots$ - Zufallsgrößen
E - sicheres Ereignis
$\emptyset$ - unmögliches Ereignis: $P(\emptyset) = 0$
$A \in MA$ heißt fast sicher, wenn $P(A) = 1$
MA heißt fast unmöglich, wenn $P(A) = 0$
X,Y heißen gleich, wenn $\forall e \quad e \in E \rightarrow X(e) = Y(e)$
X,Y heißen fast sicher gleich, wenn $P(e : X(e) = Y(e)) = 1$
Stochastische Konvergenz (auch „Konvergenz in Wahrscheinlichkeit“): Eine Folge $\{X_i\}$ von Zufallsgrößen konvergiert stochastisch gegen die Zufallsgröße X, wenn

$$\forall \epsilon_1 \forall \epsilon_2 \exists n_o(\epsilon_1, \epsilon_2) \forall n (\epsilon_1 > 0 \wedge \epsilon_2 > 0 \wedge n \geq n_o \rightarrow P(e : |X_n(e) - X(e)| \geq \epsilon_1) \leq \epsilon_2)$$

kurz:

$$\epsilon \quad \epsilon > 0 \rightarrow \lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$$

$$\{x_n\}^P \rightarrow X$$

Fast sichere Konvergenz

$$P(e : \lim_{n\to\infty} X_n(e) = X(e)) = 1$$

kurz:

$$\{X_n\} \overset{\text{f.s.}}{\rightarrow} X$$

Interpretation: Es ist fast sicher, dasss für wachsendes n X_n ein letztes Mal stark vor X abweicht (mehr als ϵ), danach nie wieder.
(Es ist im Grunde klassische Konvergenz für die meisten ϵ - aber eben nicht für alle.)

Konvergenz im quadratischen Mittel
EX^2 möge existieren

$$\lim_{n\to\infty} E(X_n - X)^2 = 0$$

Konvergiert den Verteilungen nach (schwache Konvergenz)

$$X \sim F(x), X_n \sim F_n(x)$$

$\{x_n\}$ konvergiert schwach gegen X, wenn

$$\forall x (\text{F stetig an Stelle x}) \to \lim_{n\to\infty} F_n(x) = F(x)$$

Satz: Fast sichere Konvergenz $\to$ stochastische Konvergenz
stochastische Konvergenz $\to$ schwache Konvergenz

4.2. Gesetze der großen Zahlen

Zum Begriff
$\{x_i\}$ - Folge von Zufallsgrößen. Es sei

$$z_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Gegenstand ist das Konvergenzverhalten der Filge $\{z_n\}$

Definition: $\{x_i\}$ genügt den schwachen GGZ, wenn Konstante a existiert mit:

$$\{z_n\} \xrightarrow{P} a$$

Definition: $\{x_i\}$ genügt den starken GGZ, wenn Konstante a existiert mit:

$$\{z_n\} \xrightarrow{\text{f.s}} a$$

Die GGZ sind Grundlage vieler Resultate der Statistik.
Gefragt sind notwendige und hinreichende Bedingungen, insbesondere für starkes GGZ.
Spezialfall:
Alle X_i insgesamt unabhängig und identisch verteilt. $\to$ wichtige Resultate bekannt.
Relative Häufigkeit
n - Anzahl von Experimenten
k - Anzahl der Experimente, bei denen A beobachtet wurde
$h_n(A)$ wird mit Hilfe von Zufallsgrößen definiert:

$$X_i = \begin{cases} 0 & \overline{A} \text{ im i-ten Versuch} \\ 1 & A \text{ im i-ten Versuch} \end{cases} \quad h_n(A) = \frac{\sum_{i=1}^{n} X_i}{n}$$

Satz von Bernoulli (Jakob, 1654-1705)
$\epsilon > 0$ beliebig, p = P(A). Dann gilt

$$\lim_{n\to\infty} P(|h_n(A) - p| < \epsilon) = 1$$

D.h. $\{h_n(A)\} \xrightarrow{P} p$ (stochastische Konvergenz)
$\sim$ genügt dem schwachen GGZ.
Satz von Borel(1909)

$$\{h_n(A)\}^{\text{f.s.}}_{p}$$

Beweis mit Hilfe des Satzes von Kolmogorow

$$\forall i \quad EX_i = P(A) * 1 + (1 - P(A)) * 0 = P(A) = p$$

d.h. für alle Summanden in $h_n(A)$

$$\hookrightarrow Eh_n(A) = \frac{n * p}{n} = p$$

4.3. Der Zentrale Grenzwertsatz (ZGS)

Definition: Wenn eine Summe von Zufallsgrößen (z.B. arithmetische Mittel) asymptotisch normalverteilt ist, sagt man: „Es gilt der ZGS“. (Das ist in vielen Fällen so.)
Satz von Lindeberg-Lévy
$\{x_i\}$ - Folge von insgesamt unabhängigen, identisch verteilten Zufallsgrößen mit endlicher Varianz.
Es sei

$$Y_n = \sum_{i=1}^{n} X_i \text{ und } \mu = EX_1 \quad \sigma^2 = VarX_1 \neq 0$$

Es gilt $EY_n = n * \mu$ und wegen der Unabhängigkeit $VarY_n = n * \sigma^2$
Für die Zufallsgröße

$$z_n = \frac{Y_n - n * \mu}{\sigma * \sqrt{n}} = \frac{\frac{Y_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$\frac{Y_n}{n}$ - arithmetisches Mittel
μ - $E(\frac{Y_n}{n})$
$\frac{\sigma}{\sqrt{n}}$ - $Str(\frac{Y_n}{n})$
Mit der Verteilungsfunktion $F_n(x) = P(Z_n \leq x)$ gilt dann $EZ_n = 0$

Satz von Lindeberg-Lévy: Es gilt der ZGS, insbesondere

$$\lim_{n\to\infty} F_n(x) = \int_{-\infty}^{x} nv(y, 0, 1)dy = \frac{1}{2 * \pi} * \int_{0}^{x} e^{-\frac{y^2}{2}} dy$$

Sprechweise: $\{Z_n\}$ ist asymptotisch normalverteilt.
Zusatz:
Existiert $E|X_i|^3$, ist die Konvergenz sogar gleichmäßig.
Folgerung: n sei groß $\rightarrow P(a < Z_n \leq b) \approx \frac{1}{2*\pi} * \int_a^b e^{-\frac{y^2}{2}} dy$
Bemerkung: Y_n ist als Summe darstellbar. A sei interessierendes Ereignis mit $P(A) = p$.
X sei folgende Zufallsgröße:

$$X = \begin{cases} 1 & \text{falls A eintritt} \\ 0 & \text{sonst} \end{cases}$$

$\hookrightarrow EX = p, VarX = EX^2 - (EX)^2$
$= p - p^2 = p(1-p)$
Mit X werden n Versuche gemacht $\rightarrow x_1, \ldots, x_n$
Dann ist $Y_1 = \sum\limits_{1=i}^{n} X_i$

Für die Zufallsgröße

$$Z_n = \frac{\overbrace{Y_1 - n * p}^{EY_n}}{\underbrace{\sqrt{n * p * (1-p)}}_{StrY_n}}$$

gilt dann wieder $EZ_n = 0, VarZ_n = 1$.

Satz von Moivre-Laplace:
Die Zufallsgröße Z_n sind asymptotisch standardnormalverteilt:

$$\lim_{n \to \infty} P(Z_n \leq x) = NV(x, 0, 1)$$

Für große n gilt

$$P(a < Z_n \leq b) \approx \frac{1}{2\pi} \int_a^b e^{-\frac{y^2}{2}} dy$$

$$= P(a * \sqrt{n * p * (1-p)} + n * p < Y_n \leq b * \sqrt{n * p * (1-p)} + n * p)$$

Für ausreichend gute Näherungen sollte gelten

$$n * p * (1-p) > 9$$

d.h.

$$n > 36 \text{ bei } p = 0.5$$

Beispiel:
Integralberechnung nach Monte-Carlo-Methode
Motivation: Methode für zahlreiche best. Integrale anwendbar.
Aufgabe:

$$I = \int_0^\pi \sin x dx \quad (= 2, \text{ Erg. bekannt})$$

Ansatz: Der Kurvenverlauf wird in ein Rechteck eingehüllt. Fläche des Rechtecks $= \pi$.
Es werden n über Rechteck gleichverteilte Punkte (X_i, Y_i) in das Rechteck „geworfen“.
Voraussetzung:
Versuche insgesamt unabhängig,
$\forall i: \quad X_i, Y_i$ unabhängig,
$X_i \sim [0, \pi]$ - Gleichverteilung
$Y_I \sim [0, 1]$ - Gleichverteilung
Es sei $A = \{e : Y_i \leq \sin(X_i)\}$

$$p = P(A) = \int_0^\pi \int_0^{\sin x} \frac{1}{\pi} dy dx = \frac{2}{\pi} = 0.636620$$

$a_n(A)$ - absolute Häufigkeit der Punkte, die nicht oberhalb des Graphen von sin x in $[0, \pi]$ liegen.
$h_n(A) = \frac{a_n(A)}{n}$ - relative Häufigkeit p wird geschätzt durch $h_n(A)$. Beispiel-Daten:
$n = 1000$, davon
$a_n(A) = 6377$ Punkte oberhalb von $\sin x$
$h_n(A) = 0.6377$
I wird geschätzt als Teilfläche 63.77% der Gesamtfläche vom Rechteck der Größe π:
$I \approx 0.6377 * \pi = 2.000252$

Problem: Abweichung des „wahren“ p-Wertes von der zufälligen Realisierung $h_n(A) = 0.6377$.
Abschätzung der Genauigkeit
Für die absolute Häufigkeit $a_n(A) = h_n(A) * n$ gilt nach dem Satz von Moivre-Laplace, dass

$$Z_n = \frac{a_n(A) - n * p}{\sqrt{n * p * (1 - p)}} = \frac{(h_n(A) - p)\sqrt{n}}{\sqrt{p * (1 - p)}}$$

asymptotisch normalverteilt ist.
Für n=10000 liegt NV praktisch vor.

Zufallsgröße $Z \sim NV(x, 0, 1); \quad t > 0$ wird so bestimmt, dass

$$P(-t < Z \leq t) = \frac{1}{\sqrt{2\pi}} \int_{-t}^{t} e^{-\frac{y^2}{2}} dy = 0.95$$

Lösung: $t = 1.96$ (geringfügig zu groß)
Es gilt sogar $P(-1.96 < Z \leq 1.96) > 0.95$
$\underset{\text{n groß}}{\hookrightarrow} P(-1.96 < Z_n \leq 1.96) > 0.95$

Frage 1: Wie genau wird p durch $h_n(A)$ approximiert?

$$0.95 < P(-t < Z_n \leq t)$$

$$= P(|h_n(A) - p| \leq t * \frac{\sqrt{p(1-p)}}{\sqrt{n}})$$

$$< P(|h_n(A) - p| \leq \frac{1}{2 * \sqrt{n}}) \qquad |t = 1.96$$
$$= P(\sim \leq \frac{0.98}{\sqrt{n}})$$
$$< P(|h_n - p| \leq \frac{1}{\sqrt{n}})$$

Frage 2: Welche Werte kommen für p in Frage, so dass $-t < Z_n \leq t$ gilt?
Dazu: Lösung der Gleichungen $Z_n = \pm t$ (für $a_n(A) = 6377$)

$$\frac{a_n(A) - n * p}{\sqrt{n * p * (1-p)}} = \pm t$$

$$(a_n(A) - n * p)^2 = n * p * (1-p) * t^2$$
$$\hookrightarrow p^2 - \frac{2 * a_n(A) + t^2}{n + t^2} * p + \frac{a_n(A)^2}{n^2 + n * t^2} = 0$$
$$p_1 = \frac{2 * a_n(A) + t^2}{2 * (n + t^2} \sqrt{\frac{2a_n(A) + t^2}{2 * (n + t^2} - \frac{a_n(A)^2}{n^2 + n * t^2}}$$
$$p_2 = \sim + \sim$$

Einsetzen obiger Beispiel-Daten:
$n = 10000$
$a_n(A) = 6377$
$t = 1.96$
$\hookrightarrow p_1 = 0.628228$
$\hookrightarrow p_2 = 0.647066$
Für p-Werte aus $[p_1, p_2]$ ist das Versuchsergebnis plausibel
$\hookrightarrow 1.973636 = \pi * p_1 < \int_0^\pi \sin x dx \leq \pi * p_2 = 2.032817$
verlässlich mit 0.95 Sicherheit.

4.4. Die empirische Verteilungsfunktion

Gegeben: die Zugfallsgröße $X \approx F(x)$, $F(x)$ stetig
Ausführung von n insgesamt unabhängigen Versuchen mit X
$\hookrightarrow X_1, \ldots, X_n \quad X_i \approx F(x)$

$$I(X_i, x) = \begin{cases} 1 & \text{, wenn } X_i > x \\ 0 & \text{sonst} \end{cases}$$

Empirische Verteilungsfunktion zu F(x)

$$Emp_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i, x)$$

Satz von Gliwenko (Hauptsatz der mathematischen Statistik)
$\epsilon > 0$ beliebig gegeben. Dann gilt für die Zufallsgrößen-Folge $\{D_n\}$ mit

$$D_n = \underset{x\in\mathbb{R}}{sup}|Emp_n(x) - F(x)|$$

$$P(\lim_{n\to\infty} D_n \leq \epsilon) = 1$$

Bemerkung: Satz rechtfertigt, $Emp_n(x)$ als Ersatz für F(x) zu benutzen, falls F(x) unbekannt (n möglichst groß)

Satz von Kolmogorow:
Bei stetigem F(x) gilt:

$$K(x) = \lim_{n\to\infty} P(\sqrt{n} * D_n \leq x) = \begin{cases} 0 & \text{für } x \leq 0 \\ 1 + 2 * \sum_{i=1}^{\infty}(-1)^i * e^{-2i^2x^2} & \text{für } x > 0 \end{cases}$$

Bemerkung: Satz ist theoretische Grundlage für Kolmogorow-Anpassungstest

Teil II. Statistik

1. Grundlegende Begriffe der Statistik

Gegenstand: Wahrscheinlichkeits-Aussagen

- ohne Kenntnisse der Verteilungsfunktion
- mit Kenntnissen von Stichproben

Grundgesamtheit:
Gegeben:

- W-Raum $[E, MA, P]$
- der messbare Raum $[R, BM]$
- Zufallsgröße X: $E \to R$
- $F_X(x) = P(e : X(e) \leq x)$
- W-Raum $[R, BM, F_X(x)]$

$[R, BM, F_X]$ heißt Grundgesamtheit mit Verteilung F_X

X - Zufallsgröße, $F_X(x)$ unbekannt (im Allgemeinen)
n unabhängige Experimente mit X liefern Realisierungen $x_1, \dots, x_n$
$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ heißt konkrete Stichprobe von X aus der Grundgesamtheit
$X_1, \dots, X_n$ - insgesamt unabhängige Zufallsgrößen mit Verteilungsfunktion $F_X(x)$ (Vorstellung: n gedachte Experimente)
$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ heißt mathematische Stichprobe

Beschreibende Statistik:
Wiedergabe vorhandener Daten durch Auswahl und geeignete Darstellung mit Bestimmung interessierender Kenngrößen
(Diagramme, Häufigkeitstabellen, Mittelwerte,...)

Schließende Statistik:
Mathematische Auswertung gezielt erhobener Stichproben, Rückschlüsse auf die Grundgesamtheit

Typische Aufgabenstellungen
Punktschätzungen:

- W
- EX
- StrX, VarX
- Parameter von F(x)
- ...

Aussagen über Verlässlichkeit
Berechnung von Irrtumswahrscheinlichkeiten

Intervallschätzungen:
Schätzung eines Intervalls in das die Punktschätzung fällt mit Wahrscheinlichkeit, dass das auch geschieht.

Hypothesentests:

Ist Stichprobe mit Hypothese über Grundgesamtheit verträglich?

Stichproben-Funktion:
ist eine borel-messbare Funktion $g(x_1, \ldots, x_n)$
$\hookrightarrow (X_1, \ldots, X_n)$ ist dann Zufallsgröße

2. Schätztheorie

2.1. Punktschätzungen - grundlegende Definitionen

Zwei einführende Beispiele: Zufallsgröße $X \sim \underset{\text{i.A. unbekannt}}{F(x)}$ gegeben
EX wird geschätzt durch die Zugfallsgröße

$$M_n)\frac{1}{n}\sum_{i=1}^{n} X_i$$

X_i: Ergebnis von n Experimenten mit X

Einsetzen einer konkreten Stichprobe $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

$$\hookrightarrow m_n = \frac{1}{n}\sum_{i=1}^{n} x_i \approx EX$$

VarX wird geschätzt durch

$$V_n = \frac{1}{n-1}\sum_{i=1}^{n} \underset{\text{empirische Varianz}}{(X_i - M_n)^2}$$

v_n - Realisierung von V_n:

$$v_n = \frac{1}{n-1}(x_i - m_n)^2 \approx VarX$$

Interessant: Untersuchung der Verteilungsfunktion von M_n und V_n
Bemerkung: $M_n \sim F^{n*}(n * x)$
F^{n*} ist n-fache Faltung von F(x) mit sich selbst.
P(A) wird geschätzt durhc $\underset{\text{Spezialfall von } m_n}{h_n(A)}$

Verallgemeinerung
t sei ein unbekannter Parameter der Verteilungsfunktion $F_X(x)$, also $t = t(F_X)$
t sei zu schätzen: $t \in D$ (Wertebereich)
Bemerkung: t kann ein Vektor sein
Die Menge der in Frage kommenden Verteilungsfunktionen kann eingeschränkt sein (Beispiel: „Familie“ der Normalverteilungen; $t = \binom{\mu}{\sigma}$)

Häufige Voraussetzungen:
$X_1, \dots, X_n$ - insgesamt unabhängig Exemplare von X
$(x_1, \dots, x_n)^T$ - konkrete Stichprobe
g - borel-messbare Funktion: n-stellig
Die Stichprobenfunktion $g(X_1, \dots, X_n)$ heißt Schätzfunktion bzw. eine Punktschätzung für („den Punkt“) t.

$g(x_1, \dots, x_n)$ wird als Ersatz für t benutzt.
Aber mit welchem Grund?
g sollte wünschenswerte Eigenschaften haben:

Erwartungstreue: g heißt erwartungstreu, wenn

$$\forall F_X \quad Eg(X_1, \dots, X_n) = t(F_X) \quad \text{mit } \forall i\, X_i \sim F_x$$

Eine Folge$\{g_n\}$ von Schätzfunktionen heißt asymptotisch erwartungstreu, wenn

$$\forall F_X \lim_{n\to\infty} Eg_n = t(F_X)$$

Systematischer Fehler (Synonyme: Verzerrung, Bias)

$$Eg(X_1, \dots, X_n) - t(F_X) \text{ (i.a. ein Vektor)}$$

Meist systematischer Fehler von g gegenüber t. Bemerkung; Erwartungstreue Schätzfunktionen sind biasfrei

Konsistenz: $\{g_n\}$ heißt schwach/stark konsistent, wenn $\{g_n\} \overset{\text{P/f.s.}}{\to} t(F_X)$
Bessere Schätzfunktion
g_1 heißt besser als g_2, wenn für alle $\epsilon > 0$ und alle $t \in D$ gilt:

$$P(|g_1(X_1, \dots, X_n) - t| \leq \epsilon) > P(|g_2(\dots) - t| \leq \epsilon)$$

Wirksamere Schätzfunktion:
g_1, g_2 - erwartungstreu bezüglich t
g_1 heißt wirksamer als g_2, wenn für $t \in D$

$$Var[g_1(X_1, \dots, X_n)] < Var[g_2(X_1, \dots, X_n)]$$

Wirksamste Funktion ...

2.2. Punktschätzungen - wichtige Sätze

Satz: Hinreichend für schwache Konsistenz sind

- asymptotische Erwartungstreue
- $\lim_{n\to\infty} Var[g_n(X_1, \dots, X_n)] = 0$

Satz: M_n ust erwartungstreu für EX. Beweis trivial.

Satz: X, $X_i \sim F(X)$ X_i insgesamt unabhängig

$$V_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M_n)^2 \quad S_n = \sqrt{V_n}(\approx StrX)$$

Behauptung: V_n schätzt VarX erwartungstreu.
Beweis: Zuerst Umschreiben von V_n - Klammern ausmultiplizieren

$$V_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i^2 - 2 * X_i * M_n + M_n^2)$$

$$V_n = \frac{1}{n-1}\sum_{i=1}^{n}X_i^2 + \frac{1}{n-1}(-2 * M_n * \sum_{i=1}^{n}X_i + \sum_{i=1}^{n}M_n^2)$$

$$V_n = \frac{1}{n-1}\sum_{i=1}^{n}X_i^2 + \frac{1}{n-1}(-2 * M_n * n * M_n + n * M_n^2)$$

$$V_n = \frac{1}{n-1}\sum_{i=1}^{n}X_i^2 - \frac{n}{n-1}M_n^2$$

Erwartungswertbildung (X_i durch X ersetzbar, $\sum$ entfällt)

$$EV_n = \frac{n}{n-1}(EX^2 - EM_n^2)$$

Hier fehlt ein Stück!

Damit folgt:

$$\begin{aligned} EV_n &= \frac{n}{n-1} * [E(X^2) - \frac{1}{n}E(X^2) - \frac{n-1}{n}(EX)^2] \\ &= \frac{n}{n-1} * (\frac{n-1}{n}E(X^2) - \frac{n-1}{n}(EX)^2) \\ &= EX^2 - (EX)^2 = VarX \end{aligned}$$

Bemerkungen

(a) Wäre EX bekannt, könnte VarX durch

$$W_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - EX)^2$$

erwartungstreu geschätzt werden

(b) Mit

$$V_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - M_n)^2$$

Wird VarX nur asymptotisch erwartungstreu geschätzt

Weitere Resultate für den Fall $X \approx NV(x, \mu, \sigma)$
$\hookrightarrow M_n \approx NV(x, \mu, \frac{\sigma}{n})$
$\hookrightarrow V_n$ asymptotisch normalverteilt
$\hookrightarrow M_n$ und V_n sind unabhängig Es sei $S_n = \sqrt{V_n}$
$\hookrightarrow T_n = \frac{M_n - \mu}{S_n} * \sqrt{n} \approx Stu(x, n-1)$
$\hookrightarrow \frac{n-1*S_n^2}{\sigma^2} \approx ChiQ(x, n-1)$

Schätzung einer Kovarianz
Kovarianz zweier Zufallsgrößen X und Y

$$Cov(X,Y) = E((X - EX) * (Y - EY))$$

$((x_1, y_1), \ldots, (x_n, y_n))$ sei konkrete Stichprobe zu (X,Y)
Schätzwert für Cov: $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$

$$cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}[(x_i - \underbrace{m_n(x)}_{\text{emp Mittel zu x}}) * (y_i - m_n(y))]$$

Schätzung der Korrelation

$$korr(x,y) = \frac{cov(x,y)}{\sqrt{\sum_{i=1}^{n}(x_i - m_n(x))^2 * \sum_{i=1}^{n}(y_i - m_n(y))^2}}$$

Schätzung einer Kovarianzmatrix zu einem Zufallsvektor $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$:

Voraussetzung: hohe Anzahl von Realisierungen
Ergebnis: $\begin{pmatrix} cov(x_1, x_1) & cov(x_1, x_2) & \ldots & cov(x_1, x_n) \\ \ldots & & & \\ cov(x_1, x_n) & & \ldots & cov(x_n, x_n) \end{pmatrix}$ Vorraussetzung: Stichprobenumfang >> Dimension des Zufallsvektors

2.3. Intervallschätzungen

Synonym: Bereichs-Schätzungen
Aufgabenstellung: Berechnung von P(—Parameterwert-Schätzwert—∈Intervall)

Wichtige Voraussetzung (für die folgende Aufgabe)
$X \approx NV(x, \mu, \sigma)$
$(X_1, \ldots, X_n)$ sei mathematische Stichprobe zu X

Aufgabe 1:
$\mu = EX$ unbekannt und zu schätzen
$\sigma^2 = VarX$ bekannt

$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i \approx NV(x, \mu, \frac{\sigma}{\sqrt{n}}$
Nach ZGS gilt:

$$0.95 = P(-1.96 < \frac{M_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96)$$
$$= P(M_n - \frac{1.96}{\sqrt{n}} * \sigma < \mu \leq M_n + \frac{1.96}{\sqrt{n} * \sigma)}$$

M_n ist Zufallsgröße $\rightarrow$ zufällige Intervallgrenzen
Interpretation:
Man kann mit W von 0.95 (Konfidenzniveau) darauf vertrauen, dass μ im angegebenen Intervall liegt.
Das gilt für jede Realisierung m_n von M_n. Also: Man berechne die Intervallgrenzen (mit m_n statt M_n)
μ liegt dann im Intervall mti W 0.95
Bemerkung:

(a) Falls X nicht normalverteilt ist, ist M_n dennoch asymptotisch normalverteilt

(b) Ist σ nicht bekannt, kann (notfalls) V_n verwendet werden

Aufgabe 2:
wie Aufgabe 1, aber jetzt sei σ unbekannt
Nitzung der Student'schen T-Funktion

$$V_n' = \frac{1}{n}\sum_{i=1}^{n}(X_i - M_n)^2 \quad S_n' = \sqrt{V_n^2}$$

$$T_n = \frac{M_n - \mu}{S_n'} * \sqrt{n-1}$$

Satz: $T_n \approx Stu(x, n-1)$
Konstruktion von Intervallgrenzen T_l, T_r
Vorgabe eines Konfidenzniveaus α (meist 0.95)

Bestimmung von t' si, dass $P(-t' < T_n \leq t') = \alpha$
Berechnung der Zufallsgrößen

$$T_l = -\frac{s'_n * t'}{\sqrt{n-1}} + M_n \quad T_r = \frac{s'_n * t'}{\sqrt{n-1}} + M_n$$

Dann gilt

$$\alpha = 0.95 = P(-t' < T \leq t')$$
$$P(-t' < \frac{M_n - \mu}{s'_n} * \sqrt{n-1} \leq t'$$
$$P(-\frac{s'_n * t'}{\sqrt{n-1}} + M_n < \mu \leq \frac{s'_n * t'}{\sqrt{n-1}} * M_n)$$
$$= P(T_l < \mu \leq T_r)$$

Einsetzung konkreter Realisierung x_i für $X_i \rightarrow$ konkrete Intervallgrenzen

Aufgabe 3:
2-seitiges Konfidenzintervall für σ gesucht Es sei $\mu = EX$ unbekannt

$$V_n = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \underbrace{M_n}_{\text{emp. Mittelwert}})^2$$

$$\chi^2 = \frac{n-1}{\sigma^2} * V_n^2 = \frac{1}{\sigma^2} * \sum_{i=1}^{n} (X_i - M_n)^2$$

Satz:

$$\chi^2 \approx ChiQ(x, n)$$

Vergabe eines Konfidenzniveaus α (meist 0.95)
Bestimmung von t_l, so dass $P(\chi^2 \leq t_l) = \frac{\alpha}{2}$
Bestimmung von t_r, so dass $P(\chi^2 > t_l) = \frac{\alpha}{2}$
$\hookrightarrow P(t_l < \chi^2 \leq t_r) = \alpha$ Zum testen sei Hypothese $H : \sigma = \sigma_0$

(a) Einsetzen von σ_0 in χ^2-Formel

(b) t_l, t_r berechnen

(c) Konkrete Stichprobe in χ^2-Formel einsetzen - Konkreter χ^2-Wert!

$$\text{Test: } \chi^2 \in (t_l, t_r]?$$

- wenn ja: Es gibt keinen Grund, H abzulehnen.

- wenn nein: H ist abzulehnen, weil bei Richtigkeit ein unwahrscheinliches Ereignis mit W $1-\alpha$ eingetreten wäre, was unglaubhaft ist.

P(Fehler 1. Art) = P(Ablehnung zu Unrecht) = $1-\alpha$
P(Fehler 2. Art) = P(Nicht-Ablehnung zu Unrecht) = keine Aussage möglich (ohne weiteres

Aufgabe 4: wie Aufgabe 3, aber $\mu = EX$ bekannt
V_n' in χ^2-Formel verwenden (Dividieren durch n, μ statt M_n)
ChiQ(x,n-1) verwenden
Ansonsten analog vorgehen

2.4. Maximum-Likelihood-Schätzmethode

$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ - konkrete Stichprobe aus Grundgesamtheit

$$[R, BM, F_X(x,t)]$$

$$t = \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} \text{ unbekannter Punktvektor, } t \in D \text{ möglicherweise bekannter Bereich}$$

(a) X hat diskrete Verteilung; k Werte möglich: $a_1, dots, a_k$; $p_i = P(X = a_i)$
$p_i(a_i, t_1, \ldots, t_n)$ - Funktion des Vektors t
Likelihood-Funktion:

$$L(x_1, \ldots, x_n, t_1, \ldots, t_m) = \prod_{i=1}^{n} p_i(x_i, t_1, \ldots, t_m)$$

(b) Stetiger Fall
$F_X(x,t)$ sei nach x differenzierbar $\to f(x) = \frac{d(F_X(x,t)}{dx}$
Likelihood-Funktion:

$$L(x_1, \ldots, x_n, t_1, \ldots, t_m) = \prod_{i=1}^{n} f(x_i, t_1, \ldots, t_m)$$

ML-Methode zum Finden einer Punktabschätzung für t
Vorschrift: Der Punkt $t*$ wird als Schätzwert verwendet, für den die Likelihood-Funktion bei vorliegender Stichprobe x maximal wird.
formal: $L(x, t*) = \underbrace{\max_{t \in D} L(x,t)}_{\text{Existenz, Eindeutigkeit nicht sicher}}$
Mögliche Methoden $t*$ zu finden:

L nach allen t_i differenzieren, Ableitung $= 0$ setzen
$\hookrightarrow$ Gleichungssystem (m Gleichungen) lösen
besser: ln(L) ableiten
$\hookrightarrow ln(\prod \dots)$ ist eine Summe
Bemerkung: Unter bestimmten Bedingungen ist eine ML-Schätzung

- erwartungstreu
- konsistent
- asymptotisch normalverteilt
- asymptotisch effektiv
- erschöpfend

ML-Schätzung einer Streuung
$X \approx NV(x, 0, \sigma)$ mit $\mu = 0$ und σ unbekannt
$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ sei konkrete Stichprobe vom Umfang $n \geq 2$ mit $x_1 = -2, x_2 = 1$
σ ist mit ML-Methode zu schätzen

$$L(-2, 1, \sigma) = \frac{1}{2\pi\sigma^2} * e^{-\frac{-2*-2}{2\sigma^2}} * e^{-\frac{1*1}{2\sigma^2}}$$

$$= \frac{1}{2\pi\sigma^2} * e^{-\frac{5}{2\sigma^2}}$$

$$L'(-2, 1, \sigma) = \frac{-2}{2\pi\sigma^3} * e^{-\frac{5}{2\sigma^2}} + \frac{1}{2\pi\sigma^2} * \frac{2 * 5}{2\pi\sigma^3} * e^{-\frac{5}{2\sigma^2}}$$

$$= \frac{-1}{\pi\sigma^3} * e^{\cdots} + \frac{5}{2\pi\sigma^5} * e^{\cdots} \overset{!}{=} 0$$

$$\hookrightarrow 2\sigma^2 = 5$$

$$\hookrightarrow \sigma = \sqrt{\frac{5}{2}}$$

L'' wäre zu untersuchen für den Maximumsnachweis
Kürzer: ϵ klein, L' programmieren (Matlab)

$$L'(2, 1, \sqrt{\frac{5}{2}} - \epsilon) < 0$$

$$L'(2, 1, \sqrt{\frac{5}{2}}) = 0$$

$$L'(2, 1, \sqrt{\frac{5}{2}} + \epsilon) > 0$$

$\rightarrow L'$ monoton wachsend
Vergleich mit Ergebnis der bekannten Schätzformel

$$V_2' = \frac{1}{2}\sum_{i=1}^{2}(x_i - 0)^2$$

bekannter Wert $EX = 0$, nicht M_2

$$\frac{1}{2} * (4+1) = \frac{5}{2} \leftarrow \text{Schätzwert für } \sigma^2$$

$$\sqrt{\frac{5}{2}} \leftarrow \text{Schätzwert für} \sigma$$

Vorlesung KW56.1 fehlt

3. Signifikanztests

Begriff:
Es geht nur eine Hypothese über eine Grundgesamtheit
Gegeben: Hypothese H, konkrete Stichprobe
Gesucht: Aussage, ob Stichprobe gegen H spricht
Prinzip

- Annahmen über Grundgesamtheit
- Aufstellen von H
- Vorgabe eines Konfidenzniveaus α (meist 0.95) bzw. Irrtums-Wahrscheinlichkeit $1 - \alpha$
- Definition einer Stichprobenfunktion, Ermittlung über Verteilung
- Benennen eines kritischen Bereichs

Unterscheidung

- verteilungsabhängiger Fall: Ausnahmen über Verteilung der Grundgesamtheit (z.B. NV)
- verteilungsunabhängige Tests: keine solchen Annahmen

3.1. Student-Test

Begriff:
Bezeichnung für Tests, wo Testgröße Student-verteilt ist

3.1.1. Student-Test I

(zum Test eines Mittelwerts) Problematik: Zg X $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ konkrete Stichprobe

μ_0 sei gegeben

H: $EX = \mu_0$ Daher gibt es signifikanten Unterschied? Wichtige Voraussetzung!!!

$$X \approx NV(x, \mu, \sigma) \quad \sigma > 0 \text{ (i.a. unbekannt)}, EX = \mu$$

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad V_n' = \frac{1}{n}\sum (X_1 - M_n)^2$$

Stichprobenfunktion:

$$T_n = \frac{\mu_n - \mu_0}{\sqrt{V_n'}} * \sqrt{n-1} \approx Stu(x, n-1)$$

α vorgegeben
Intervall I gesucht mit $P(T_n \in I) = \alpha$

Bestimmung der linken, bzw. rechten Grenze t_l, bzw. t_r aus

$$\frac{1-\alpha}{2} = \int_{-\infty}^{t_l} stu(x, n-1)dx = \int_{t_r}^{\infty} stu(x, n-1)dx$$

Symmetrie von $stu(x, \dots)$ bezüglich $0 \to -t_l = t_r$

Definition: $I = [t_l, t_r]$ heißt Konfidenzintervall für T_n zum Konfidenzniveau α.
Dann gilt $P(|T_n| \leq t_r) = \alpha$
Vorgehen für Student-I-Test:

(a) $H : EX = \mu_0$

(b) α festlegen

(c) Testgröße t_n berechnen (t_n Realisierung von T_n)

(d) t_l, t_r ermitteln, so dass $P(|T_n| > t_r) < 1 - \alpha$

(e) Falls $t_n > t_r$ $\rightarrow$H ablehnen
Falls $t_n \leq t_r$, spricht nicht gegen H

Fehler 1. Art: Ablehnung einer zutreffenden Hypothese (W dafür $1 - \alpha$)
Fehler 2. Art: Nichtablehnung einer falschen Hypothese

3.1.2. Student-Test II

(Vergleich zweier Mittelwerte)
Synonym: doppelter t-Test
Problem: Zg X,Y unabhängig
2 konkrete Stichproben:

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \text{ und } \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

Gilt $EX = EY$?
Voraussetzung: X und Y normalverteilt mit $VarX = VarY$ (i.a. unbekannt)
Beispiel: Reißfestigkeit von Stahlseilen, die nach verschiedenen Verfahren hergestellt werden
Berechnung von $M_M(\mathbf{X}), V'_m(\mathbf{X}), M_n(\mathbf{Y}), V'_n(\mathbb{Y})$

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix}$$

Stichprobenfunktion: Es sei $k = m + n$

$$T_k = \frac{M_m(\mathbf{X}) - M_n(\mathbf{Y})}{\sqrt{(m-1)V'_m(\mathbf{X}) + (n-1)V'_n(\mathbf{Y})}} * \sqrt{\frac{m * n * (m+n-2)}{m+n}}$$

Satz: Falls $EX = EY$, ist T_k student-verteilt mit k Freiheitsgraden
Vorgehen:

(a) H: $EX = EY$

(b) α festlegen

(c) t_k berechnen (konkrete Stichprobe in T_k-Formel einsetzen)

(d) t_r bestimmen wie zuvor

(e) Falls $|t_k| > t_r$, H ablehnen, sonst nicht ablehnen

Fehler 1./2. Art möglich

3.2. Anpassungstests

syn.: Nichtparametrische Tests
Zg $X \approx F(x), F(x)$ unbekannt
Nichtparametrische Hypothese: Hypothese über Gestalt von $F(x)$
Anpassungstest: Test einer solchen Hypothese (Es geht nicht um Parameter)

Konkrete Stichprobe $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ muss vorliegen
$Emp_n(x)$ weicht nur wenig von F(x) ab.
Vorgehen:

- Hypothese bzgl. F(x) aufstellen
- geeignete Stichprobenfunktion definieren
- Test, ob Funktionswert in bestimmten Intervall liegt

3.2.1. Kolmogorow-Test

Vorgehen:

1. empirische Verteilungsfunktion $emp_n(x)$ berechnen aus Stichprobe
2. d_n berechnen (Realisierung von D_n)
3. Konfidenzniveau α wählen
4. α-Quantil der Kolmogorow-Verteilungsfunktion K(x) berechnen
   z.B. $K(1.359) > 0.95 = \alpha$
5. Gilt $\sqrt{n}d_n \leq 1.359$?
   Ja, dann besteht kein Grund die Hypothese abzulehnen.
   Nein, dann ablehnen.

Fehler 1. und 2. Art möglich.

3.2.2. Kolmogorow-Smirnow-Test

Problem: 2 Stichproben gegeben
Haben wir dieselbe Verteilungsfunktion?
(Wird hier nicht behandelt, nur der Vollständigkeit halber)

3.2.3. Der χ^2-Test

Allgemein: ein Test, bei dem Testgröße (annähernd) χ^2-verteilt ist.
Bsp.:
χ^2-Streuungstest zum Test einer Hypothese bezüglich σ bei Normalverteilung
χ^2-Anpassungstest (Darum geht es hier)
X - Zg $\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$-math. Stichprobe

$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$-konkrete Stichprobe
Hypothese H: $X \approx F(X)$
x-Achse wird in k paarweise disjunkte Mengen M_i zerlegt
$M_i = \{x : t_i < x \leq t_{i+1}\}$ $i = 1, \ldots, k$
Setze:

$$p_i = F(t_{i+1} - F(t_i)$$

A_i - absolute Zahl der Werte aus der math. Stichprobe, die in das Intervall M_i reinfallen.
a_i - absolute Zahl der Werte aus der konkreten Stichprobe, die in das Intervall M_i gefallen sind
$n * p_i$ - entsprechende „theoretische" Häufigkeiten
Empfehlung: $t_1, \ldots, t_{k+1}$ so wählen, dass $\forall i : a_i \geq 10$
Pearson'sche Stichprobenfunktion $rchi^2$

$$\chi^2 = \sum_{i=1}^{k} \frac{(A_i - n * p_i)^2}{n * p_i}$$

Es sei $F_n(x) = P(\chi^2 \leq x)$

Satz: $\lim_{n \to \infty} F_n(x) = ChiQ(x, k-1)$ (siehe WR 2.3)
Vorgehen beim Test:

1. Hypothese H: $X \approx F(x)$
2. k festlegen: M_i definieren (d.h. t_i)
3. χ^2-Wert aus konkreter Stichprobe berechnen
4. Sicherheits-Wert α(=0.95) festlegen
5. Grenzwert g berechnen aus Gleichung

$$\alpha = ChiQ(g, k-1)$$

6. Test $\chi^2 \leq g$?
   Ja $\to$ kein Grund zur Ablehnung von H
   Nein $\to$ H wird abgelehnt

Problem: Vernachlässigung der Abhängigkeit von n.
Test ist verlässlich, wenn n groß ist.

Beispiel zum χ^2-Anpassungstest
Aus einer sehr großen Anzahl von Werkstücken werden n=100 Stücke entnommen. Es

finden sich 22 fehlerhafte Stücke.
Hypothese H: P(Zufallswerkstück fehlerhaft) = 0.2
Problem: Ist H mit Stichprobe vereinbar?
Feststellung: t=22 ist die Realisierung einer nach B(k,100) verteilten Zufallsgröße mit p=0.2 (, falls H wahr ist)
Aufteilung der Stichprobe:

- 22 fehlerhafte Stücke
- 78 fehlerlose Stücke

Nebenüberlegung:

$$\begin{pmatrix} X_1 \\ \vdots \\ X_{100} \end{pmatrix} X_i = \begin{cases} -1 & \leftrightarrow \text{i-tes Stück fehlerhaft} \\ 1 & \leftrightarrow \text{i-tes Stück fehlerlos} \end{cases}$$

Zerlegung der x-Achse

$$M_1 = (-2, 0] \quad M_2 = (0, 2]$$

$$\hookrightarrow a_1 = 22 \quad a_2 = 78$$

$\rightarrow \chi^2$-Test anwendbar

$$\chi^2 = \frac{(a_1 - n * p)^2}{n * p} + \frac{(a_2 - n * p)^2}{n * p}$$

$$= \frac{(22 - 100 * 0.2)^2}{100 * 0.2} + \frac{(78 - 100 * (0.8))^2}{100 * 0.8}$$

$$= \frac{4}{20} + \frac{4}{80} = \frac{1}{4}$$

Gewünscht sei Sicherheits-Wahrscheinlichkeit $\alpha = 0.99$
g sei Lösung der Gleichung $\alpha = 0.99 = ChiQ(g, 1)$ $(k = 2, k - 1 = 1)$
Es ergibt sich $g = 6.6$,d.h. $P(\chi^2 \geq 6.6) = 0.01 = 1 - \alpha$
$0.25 < 6.6 \rightarrow$ kein Grund zur Ablehnung von H.

4. Korrelations- und Regressionsanalyse

Gegenstand: Abhängigkeiten zwischen Merkmalen
Korrellations-Analyse: Klärung, ob linearer Zusammenhang besteht.
Regressions-Analyse: Untersuchung der Form des Zusammenhangs

4.1. Korrelations-Analyse

Siehe 2.2 Punktschätzungen
korr(x,y) schätzt Korr(X,Y) - i.a. nicht erwartungstreu.
x,y - Stichproben (Vektoren), X,Y - 2 Zufallsgrößen Falls Korr(X,Y)=0, ist die Schätzung doch erwartungstreu
Zusätzliche Voraussetzung:
$\begin{pmatrix} x \\ y \end{pmatrix}$ habe eine 2D-NV
Unabhängigkeit$\leftrightarrow$Unkorreliertheit
Ziel: Unabhängigkeits-Test

$$T = \frac{Korr(X,Y) * \sqrt{n-2}}{\sqrt{1 - Korr(X,Y)^2}}$$

Hypothese H: Korr(X,Y) = 0

Satz: T student-verteilt mit n-2 Freiheitsgraden (Stu(x,n-2))
Realisierung von T:

$$t = \frac{korr(x,y) * \sqrt{n-2}}{\sqrt{1 - korr(x,y)^2}}$$

α wählen
t_l, t_r bestimmen, so dass

$$P(T \in (t_l, t_r]) = \alpha \text{ (unter H)}$$

$t \in (t_l, t_r)$?
Ja $\rightarrow$ nichts spricht gegen Unabhängigkeit ($\overset{\text{hier}}{=}$ Unkorreliertheit)
Nein $\rightarrow$ H verwerfen

4.2. Regressionsanalyse

$\begin{pmatrix} x \\ y \end{pmatrix}$ - Zufallsvektor
von Interesse: $y(x) = E(Y|X = x)$
Graph heißt Regressionskurve.
linearer Fall: Regressionsgerade
Regression 1. Art: Ermittlung der R-Kurve
Regression 2. Art: a,b so bestimmen (im lin. Fall)

$$E(Y - aX - b)^2$$

minimal wird.