Stochastik für Studierende der Informatik

Wiederholungs-/Fragestunde

Peter Czuppon

Uni Freiburg, 05. September 2016

Diese Zusammenfassung wurde mit Hilfe des Skriptes von Prof. Dr. Pfaffelhuber aus dem Sommersemester 2016 erstellt. Ferner deckt die vorliegende Version nicht den gesamten Inhalt der Vorlesung ab, sondern liegt meiner persönlichen Priosierung zugrunde. Außerdem sind Fehler keinesfalls ausgeschlossen.

Kapitel 1: Grundlagen

Gegeben: $x = (x_1, \dots, x_n)$ Vektor mit paarweise verschiedenen Elemente, d.h. $\forall i, j \in \{1, \dots, n\}$ mit $i \neq j$ gilt $x_i \neq x_j$.

Formeln der Kombinatorik / Urnenmodelle

Gesucht: Anzahl an k-elementige Teilmengen mit $k \leq n$ in verschiedenen Situationen.

	ohne Zurücklegen	mit Zurücklegen
ohne Reihenfolge	$\{x_{i_1}, \dots, x_{i_k}\}, i_l \neq i_m$	$\{x_{i_1}, \dots, x_{i_k}\}$
mit Reihenfolge	$(x_{i_1}, \dots, x_{i_k}), i_l \neq i_m$	$(x_{i_1}, \dots, x_{i_k})$

Kapitel 1: Grundlagen

Gegeben: $x = (x_1, ..., x_n)$ Vektor mit paarweise verschiedenen Elemente, d.h. $\forall i, j \in \{1, ..., n\}$ mit $i \neq j$ gilt $x_i \neq x_j$.

Formeln der Kombinatorik / Urnenmodelle

Gesucht: Anzahl an k-elementige Teilmengen mit $k \leq n$ in verschiedenen Situationen.

	ohne Zurücklegen	mit Zurücklegen
ohne Reihenfolge	$\binom{n}{k}$	$\binom{n+k-1}{k}$
mit Reihenfolge	$n(n-1)\cdots(n-k+1)$	n^k

Kapitel 1: Grundlagen

Gegeben: $x = (x_1, ..., x_n)$ Vektor mit r verschiedenen Zahlen, wobei $i-$te Zahl k_i mal vorkommt (also $k_1 + ... + k_r = n$).

Permutationen

Anzahl an verschiedenen Vektoren $(x_{i_1}, ..., x_{i_n})$:

$$\frac{n!}{k_1! \cdots k_r!}$$

Beachte: gilt $k_i = 1$ für alle i, so ist dies gerade Ziehen ohne Zurücklegen mit Reihefolge für $k = n$.

Aufgaben 4,6 auf Zusatzblatt.

Kapitel 1: Grundlagen

Rekursionen

Lösen von Rekursionen mit Hilfe von erzeugenden Funktionen.
Allgemeines Vorgehen:

1. Betrachte Rekursion $x_n = g(x_0, ..., x_{n-1})$, g beliebig
2. Setze an mit $f(z) = \sum_{n=0}^{\infty} x_n z^n$
3. Ersetze x_n durch rekursive Darstellung
   $\Rightarrow f(z) = \sum_{n=0}^{\infty} g(x_0, ..., x_{n-1}) z^n$
4. Finde explizite Reihendarstellung von f (bspw. mit Partialbruchzerlegung (siehe Aufgabe 3, Blatt 1) oder Nullstellen bestimmen (siehe Beispiel 1.11, Vorlesung))
5. Koeffizientenvergleich von expliziter Darstellung von f mit Ansatz

Kapitel 1: Grundlagen

Definitionen

X Zufallsvariable.

	diskret	stetig
Grundraum	E diskrete Menge	$\mathbb{R}$
Verteilung	$\mathbf{P}(X \in A) = \sum_{i \in A} \mathbf{P}(X = i)$	$\mathbf{P}(X \in A) = \int_A f(x)dx$
Normiertheit	$\sum_{i \in E} \mathbf{P}(X = i) = 1$	$\int_{\mathbf{R}} f(x)dx = 1$

Notation: $\mathbf{P}(A) := \mathbf{P}(X \in A)$.

Kapitel 1: Grundlagen

Eigenschaften von Verteilungen (Aufgabe 1, Zusatzblatt)

- Komplement: $\mathbf{P}(X \in A^c) = 1 - \mathbf{P}(X \in A)$
- Additivität: $(A_i)_{i \in \mathbf{N}}$ paarweise disjunkt ($A_i \cap A_j = \emptyset$, für $i \neq j$), dann:
$$\mathbf{P}\left(X \in \bigcup_{n \in \mathbf{N}} A_n\right) = \sum_{n=1}^{\infty} \mathbf{P}(X \in A_n)$$
- Siebformel:
$$\mathbf{P}(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \mathbf{P}(A_i) - \sum_{1 \leq i < j \leq n} \mathbf{P}(A_i \cap A_j) + \cdots \pm \mathbf{P}(A_1 \cap \cdots \cap A_n)$$

Kapitel 1: Grundlagen

Gemeinsame Verteilung

$X = (X_1, ..., X_n)$ höherdimensionale ZV.

- Randverteilung: $A_i \mapsto \mathbf{P}(X_i \in A_i)$
- Gemeinsame Verteilung:
  $\mathbf{P}(X \in A_1 \times \cdots \times A_n) = \int_{A_1} \cdots \int_{A_n} f(x_1, ..., x_n) dx_1 \cdots dx_n$
- Dichte: $\mathbf{P}(X \in dx_1 \cdots dx_n) = f(x_1, ..., x_n) dx_1 \cdots dx_n$

Unabhängigkeit (Aufgabe 5, Zusatzblatt)

$X_1, ..., X_n$ heißen *unabhängig*, falls

$$\mathbf{P}(X_1 \in A_1, ..., X_n \in A_n) = \mathbf{P}(X_1 \in A_1) \cdots \mathbf{P}(X_n \in A_n)$$

Kapitel 2: Wichtige (diskrete) Verteilungen

Diskrete Gleichverteilung

E endliche Menge, dann ist $X \sim \mathcal{U}(E)$, falls $\mathbf{P}(X \in A) = \frac{|A|}{|E|}$.

Binomialverteilung

$E = \{0, ..., n\}$, $0 \leq p \leq 1$. Dann ist $X \sim B(n, p)$, falls

$$\mathbf{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Interpretation: k Erfolge bei n Versuchen.

Weitere diskrete Verteilungen: Bernoulli ($= B(1, p)$), Poisson, geometrisch

Kapitel 2: Wichtige (stetige) Verteilungen

Stetige Gleichverteilung

$E = [a, b]$ Intervall, dann ist $X \sim \mathcal{U}([a, b])$, falls

$$\mathbf{P}(X \in (c, d)) = \frac{d-c}{b-a}.$$

Die zugehörige Dichte ist gegeben durch $f(x) = \frac{1}{b-a}\mathbb{1}_{x\in[a,b]}$.

Normalverteilung

$\mu \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}$. Dann ist $X \sim \mathcal{N}(\mu, \sigma^2)$, falls

$$\mathbf{P}(X \in dx) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx.$$

Weitere stetige Verteilungen: Exponential, Beta, Gamma

Kapitel 3: Kenngrößen von Zufallsvariablen

X E-wertige ZV.

Erwartungswert

- Diskreter Fall: $\mathbf{E}[X] := \sum_{k \in E} k\mathbf{P}(X = k)$.
- Stetiger Fall: $\mathbf{E}[X] := \int_E xf(x)dx$.

Eigenschaften:

- Monotonie: $X \leq Y \Rightarrow \mathbf{E}[X] \leq \mathbf{E}[Y]$
- Linearität: $\mathbf{E}[aX + bY] = a\mathbf{E}[X] + b\mathbf{E}[Y]$

Kapitel 3: Kenngrößen von Zufallsvariablen

Example

- Stetige Gleichverteilung: $X \sim \mathcal{U}([0,1])$

$$\mathbf{E}[X] = \frac{1}{1-0}\int_0^1 x dx = (\frac{x^2}{2})_{x=0}^{x=1} = \frac{1}{2}$$

- Poissonverteilung: $X \sim \mathrm{Poi}(\lambda)$

$$\mathbf{E}[X] = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda}\lambda e^{\lambda} = \lambda$$

Kapitel 3: Kenngrößen von Zufallsvariablen

Average-Case-Analyse

Allgemeines Vorgehen:

1. Definiere Anzahl an Schritten als Summe von Indikatorfunktion geeigneter Mengen
2. Bestimme die Wahrscheinlichkeit für das Eintreffen dieser Mengen
3. Berechne den Erwartungswert mit Hilfe der Linearität des Erwartungswertes

Aufgabe 7, Zusatzblatt.

Kapitel 3: Kenngrößen von Zufallsvariablen

X, Y E-wertige ZV, $\mu_X := \mathbf{E}[X]$, $\mu_Y := \mathbf{E}[Y]$.

Varianz und Kovarianz

Varianz von X ist gegeben durch

$$\sigma^2 := \mathbf{Var}[X] = \mathbf{E}[(X-\mu)^2] = \mathbf{E}[X^2] - \mathbf{E}[X]^2.$$

σ heißt Standardabweichung.
Die Kovarianz von X und Y ist gegeben durch

$$\mathbf{Cov}[X, Y] := \mathbf{E}[(X-\mu_X)(Y-\mu_Y)] = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y].$$

Für $\mathbf{Cov}[X, Y] = 0$ sind X, Y unkorreliert ($\neq$ unabhängig!).

Bemerkung: X, Y unabhängig $\Rightarrow$ X, Y unkorreliert (Beweis?)

Kapitel 3: Kenngrößen von Zufallsvariablen

Example

- Stetige Gleichverteilung: $X \sim \mathcal{U}([0,1])$

$$\mathbf{Var}[X] = \frac{1}{1-0}\int_0^1 x^2 dx - \frac{1}{4} = (\frac{x^3}{3})_{x=0}^{x=1} - \frac{1}{4} = \frac{1}{12}$$

- Poissonverteilung: $X \sim \text{Poi}(\lambda)$

$$\mathbf{Var}[X] = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} - \lambda^2$$

$$= e^{-\lambda}\lambda\left(\sum_{k=1}^{\infty}(k-1)\frac{\lambda^{k-1}}{(k-1)!} + e^{\lambda}\right) - \lambda^2 = \lambda$$

Kapitel 3: Kenngrößen von Zufallsvariablen

Charakteristische Funktion

X ZV. Dann ist die charakteristische Funktion gegeben durch:

$$\Psi_X(t) := \mathbf{E}[e^{itX}].$$

Example

$X \sim \mathrm{Exp}(\lambda)$. Dann gilt

$$\Psi_X(t) = \mathbf{E}[e^{itX}] = \int_0^\infty e^{itx} \lambda e^{-\lambda x} dx$$
$$= \lambda \left(\frac{1}{it - \lambda} e^{(it-\lambda)x} \right)_{x=0}^{x=\infty} = \frac{\lambda}{\lambda - it}$$

Kapitel 3: Kenngrößen von Zufallsvariablen

Charakteristische Funktion der Normalverteilung

Für $X \sim \mathcal{N}(\mu, \sigma^2)$ ist die charakteristische Funktion gegeben durch:

$$\Psi_X(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}.$$

Lineartransformationen von Normalverteilungen

Mit Hilfe der charakteristischen Funktion sieht man, dass für $X \sim \mathcal{N}(\mu, \sigma^2)$ auch $a + bX \sim \mathcal{N}(a + b\mu, (b\sigma)^2)$ für $a, b \in \mathbb{R}$ gilt.

Unabhängige normalverteilte Zufallsvariablen

Sind X, Y normalverteilte ZV, so gilt:
X, Y unabhängig genau dann wenn X, Y unkorreliert.

Kapitel 3: Kenngrößen von Zufallsvariablen

Simulationslemma

X $\mathbb{R}$-wertige ZV mit Verteilungsfunktion $F(x) := \mathbf{P}(X \leq x)$. Ferner sei $U \sim \mathcal{U}([0,1])$. Dann ist $F^{-1}(U)$ genauso verteilt wie X, wobei

$$F^{-1}(x) := \inf\{q : F(q) \geq x\}$$

Example

Sei $X \sim \text{Exp}(\lambda)$. Dann ist $-\frac{\ln(U)}{\lambda}$ exponentialverteilt zu Parameter λ (vgl. Aufgabe 15, Blatt 4).

Kapitel 4: Approximationssätze

Summen von Zufallsvariablen

$X_1, \ldots, X_n$ unabhängig, identisch verteilte ZV. $Y_n := X_1 + \ldots + X_n$.

$$\mathbf{E}[Y_n] = \mathbf{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{E}[X_i] = n\mathbf{E}[X_1]$$

$$\begin{aligned}\mathbf{Var}[Y_n] &= \mathbf{Var}\left[\sum_{i=1}^{n} X_i\right] \\ &= \sum_{i=1}^{n} \mathbf{Var}[X_i] + 2 \sum_{1 \le i < j \le n} \mathbf{Cov}[X_i, X_j] = n\mathbf{Var}[X_1]\end{aligned}$$

Kapitel 4: Approximationssätze

Schwaches Gesetz großer Zahlen

$X_1, X_2, \ldots$ $\mathbb{R}$-wertige, unabhängige, identisch verteilte ZV mit endlichem Erwartungswert $\mu := \mathbf{E}[X_1]$ und endlicher Varianz. Dann gilt für alle $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbf{P}\left(\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right) = 0.$$

Aufgabe 2, Zusatzblatt

Kapitel 4: Approximationssätze

Satz von de-Moivre-Laplace

Sei $Z \sim \mathcal{N}(0,1)$. Seien $X_1, X_2, \ldots$ binomialverteilte ZVen mit $\mathbf{Var}[X_n] \to \infty$. Dann ist für $-\infty \leq c < d \leq \infty$

$$\lim_{n\to\infty} \mathbf{P}\left(c \leq \frac{X_n - \mathbf{E}[X_n]}{\sqrt{\mathbf{Var}[X_n]}} \leq d\right) = \mathbf{P}(c \leq Z \leq d).$$

Aufgabe 9, Zusatzblatt

Kapitel 5: Bedingte Wahrscheinlichkeit

Definitionen

X, Y ZVen.

- Bedingte Wahrscheinlichkeit: $\mathbf{P}(Y \in B | X \in A) = \frac{\mathbf{P}(X \in A, Y \in B)}{\mathbf{P}(X \in A)}$
- Bedingte Verteilung: X diskret, dann ist die bedingte Verteilung gegeben durch

$$\mathbf{P}(Y \in B | X) : x \mapsto \mathbf{P}(Y \in B | X = x).$$

(X, Y) hat gemeinsame Dichte $f(x, y)dxdy$, dann ist die bedingte Verteilung gegeben durch

$$\mathbf{P}(Y \in B | X) : x \mapsto \frac{\int_B f(x, y)dy}{\int_{\mathbb{R}} f(x, z)dz}$$

Kapitel 5: Bedingte Wahrscheinlichkeit

Example

$U, V \sim \mathcal{U}([0,1])$ mit $X := \min(U,V)$ und $Y := \max(U,V)$ hat gemeinsame Dichte $f(x,y) = 2\mathbb{1}_{\{0\leq x\leq y\leq 1\}}$. Wir wollen $f_{X|Y=y}$ bestimmen. Dafür benötigen wir

$$f_Y(y) = \int_0^1 f(x,y)dx = 2\int_0^1 \mathbb{1}_{\{0\leq x\leq y\leq 1\}}dx = 2\int_0^y dx = 2y.$$

Nun verwenden wir die Formel von oben und erhalten (vgl. Beispiel 5.3, Vorlesung):

$$f_{X|Y=y}(x) = \frac{f(x,y)dx}{f_Y(y)} = \frac{2\cdot\mathbb{1}_{\{0\leq x\leq y\}}}{2y} = \frac{1}{y}\mathbb{1}_{\{0\leq x\leq y\}}$$

Aufgabe 12, Zusatzblatt

Kapitel 5: Bedingte Wahrscheinlichkeit

Formel von Bayes

X, Y diskrete ZVen. Dann gilt:

$$\begin{aligned}\mathbf{P}(X \in A | Y \in B) &= \frac{\mathbf{P}(Y \in B | X \in A)}{\sum_x \mathbf{P}(Y \in B | X = x)\mathbf{P}(X = x)} \\ &= \frac{\mathbf{P}(Y \in B | X \in A)}{\mathbf{P}(Y \in B)}.\end{aligned}$$

Die letzte Gleichheit ist die Formel für die totale Wahrscheinlichkeit.

Aufgabe 3, Zusatzblatt

Kapitel 5: Bedingte Wahrscheinlichkeit

Bedingte Erwartung

X, Y ZVen mit bedingter Verteilung $\mathbf{P}(Y \in B|X)$. Wir definieren die bedingte Erwartung wie folgt:

$$\mathbf{E}[Y|X] = \sum_y y\mathbf{P}[Y = y|X] \quad \text{bzw.} \quad \mathbf{E}[Y|X] = \frac{\int yf(X,y)dy}{\int f(X,y)dy}$$

Ferner gilt die *Turmeigenschaft*:

$$\mathbf{E}[\mathbf{E}[Y|X]] = \mathbf{E}[Y].$$

Kapitel 5: Bedingte Wahrscheinlichkeit

Bayes'sche Statistik

Fragestellung: Beobachtungen X mit gegebener Verteilung, die von einem (zufälligen) Parameter P abhängt (bspw. $X \sim B(n, P)$). Bestimme anhand der Beobachtungen die Verteilung von diesem Parameter (bei gegebener Ausgangsverteilung für P, z.B. $P \sim \mathcal{U}([0, 1])$). D.h. wie sieht die *bedingte Verteilung von P gegeben X* aus?
Allgemeines Vorgehen: Verwendung von Formel von Bayes liefert die konjugierte Verteilung (zumindest bis auf Proportionalitätskonstanten)

Vgl. Aufgabe 38, Blatt 10

Kapitel 5: Bedingte Wahrscheinlichkeit

Markovkette

Sei I eine Indexmenge (z.B. $I = \mathbb{N}$) und X_t eine E-wertige ZV für alle $t \in I$.

- Wir nennen $(X_t)_{t\in I}$ einen *stochastischen Prozess*.
- $(X_t)_{t\in I}$ ist eine *Markovkette*, falls $\mathbf{P}(X_{t+1} = i | X_0, ..., X_t) = \mathbf{P}(X_{t+1} = i | X_t)$.
- Die Wahrscheinlichkeiten $P_{ij}(t) := \mathbf{P}(X_t = j | X_{t-1} = i)$ definieren eine *Übergangsmatrix*. Falls die Wahrscheinlichkeiten unabhängig von t sind, heißt die Markovkette *homogen*.

Aufgabe 11, Zusatzblatt

Kapitel 6: Statistik

Schätzproblem (vereinfacht)

Gegeben: E-wertige Daten x, Menge Θ an möglichen Parametern, die die Verteilung von x bestimmen $((P_\vartheta)_{\vartheta\in\Theta})$.
Gesucht: Schätzstatistik $m : E \to \Theta$, also Funktion, die aus den Daten einen Parameter aus Θ konstruiert.

Erwartungstreue, Konsistenz

- Ein Schätzer $\hat\vartheta$ heißt *erwartungstreu*, falls $\mathbf{E}_\vartheta[\hat\vartheta] = \vartheta$ für alle $\vartheta \in \Theta$ gilt.
- Eine Folge von Schätzern $\hat\vartheta_1, \hat\vartheta_2, ..$ heißt *konsistent*, falls $\mathbf{P}^n_\vartheta(|\hat\vartheta_n - \vartheta| \geq \varepsilon) \to 0$ für $n \to \infty$.

Aufgabe 8, Zusatzblatt

Kapitel 6: Statistik

Maximum-Likelihood-Schätzer

Für gegebene Daten $x := (x_i)_{i \in I}$, die unabhängig voneinander sind und einen Parameter $\vartheta \in \Theta$ mit Zufallsvariable $X \sim \mathbf{P}_\vartheta$ ist die *Likelihood*-Funktion definiert durch:

$$\mathcal{L}(x, \vartheta) = \mathbf{P}_\vartheta(X \in dx) = \prod_{i \in I} \mathbf{P}_\vartheta(X_i \in dx_i).$$

Der *Maximum-Likelihood-Schätzer* ist dann gegeben durch $\hat{\vartheta}_{ML} := h(x)$ mit

$$\mathcal{L}(x, h(x)) = \max_{\vartheta \in \Theta} \mathcal{L}(x, \vartheta).$$

Idee: Wahrscheinlichkeit bei gegebenem Parameter ϑ die gegebenen Daten zu beobachten.

Kapitel 6: Statistik

Example

Gesucht: Schätzer des Erwartungswertes μ in einer Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ zu Daten $x := (x_i)_{i=1,\dots,n}$.

$$\mathcal{L}(x, \mu) = \prod_{i=1}^{n} f(x_i|\mu) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{\sum(x_i - \mu)^2}{2\sigma^2}\right)$$

Kapitel 6: Statistik

Example

Gesucht: Schätzer des Erwartungswertes μ in einer Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ zu Daten $x := (x_i)_{i=1,\dots,n}$.

$$\log(\mathcal{L}(x,\mu)) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

Ableiten liefert:

$$\frac{\partial}{\partial\mu}\log(\mathcal{L}(x,\mu)) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

Maximum bestimmen: $\hat{\mu} = \frac{\sum x_i}{n}$ und testen auf Maximum:

$$\frac{\partial^2}{\partial^2\mu}\log(\mathcal{L}(x,\mu)) = -\frac{1}{\sigma^2}n < 0.$$

Kapitel 6: Statistik

Mittelwert und empirische Varianz

Zu gegebenen Daten $x := (x_i)_{i=1,\ldots,n}$ definieren wir den *Mittelwert* durch

$$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} x_i$$

und die empirische Varianz durch

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

Beide Schätzer sind unverzerrt und konsistent (vgl. Theorem 6.9, Vorlesung und Aufgabe 44, Blatt 11).

Kapitel 6: Statistik

Testproblem (vereinfacht)

Gegeben: Beobachtungen x, Verteilungsraum $(\mathbf{P}_\vartheta)_{\vartheta\in\Theta}$,
Nullhypothese $H_0 : \vartheta \in \Theta_0$, Alternativhypothese $H_A : \vartheta \in \Theta_A$,
Signifikanzniveau $\alpha \in [0, 1]$
Gesucht: Ablehnungsbereich C, sodass H_0 abgelehnt wird, falls $x \in C$ und

$$\sup_{\vartheta\in\Theta_0} \mathbf{P}_\vartheta(x \in C) \leq \alpha$$

Der Test heißt *unverfälscht*, falls

$$\mathbf{P}_{\vartheta_0}(x \in C) \leq \mathbb{P}_{\vartheta_A}(x \in C)$$

für alle $\vartheta_0 \in \Theta_0$ und $\vartheta_A \in \Theta_A$ gilt.

Kapitel 6: Statistik

Binomialtest

Statistisches Modell mit Daten x, die nach $B(n,p)$ verteilt sind.

Hypothesen: (a) $H_0 : p \in \{p^*\}$ gegen $H_A : p \notin \{p^*\}$
(b) $H_0 : p \in [0, p^*]$ gegen $H_A : p \in (p^*, 1]$
(c) $H_0 : p \in [p^*, 1]$ gegen $H_A : p \in [0, p^*)$

Teststatistik: x ist unter $\mathbf{P}_p$ nach $B(n,p)$ verteilt.

Ablehnungsbereich:

(a) $\{0, ..., k, l, ..., n\}$ mit $\mathbf{P}_{p^*}(x \leq k), \mathbf{P}_{p^*}(x \geq l) \leq \alpha/2$
(b) $\{l, ..., n\}$ mit $\mathbf{P}_{p^*}(x \geq l) \leq \alpha$
(c) $\{0, ..., k\}$ mit $\mathbf{P}_{p^*}(x \leq k) \leq \alpha$

Beispiel: In n Versuchen werden Erfolge untersucht, bspw. positive Proben.

Kapitel 6: Statistik

Einfacher t-Test

Statistisches Modell mit Daten x, die nach $\mathcal{N}(\mu, \sigma^2)$ verteilt sind.

Hypothesen: (a) $H_0 : \mu \in \{\mu^*\}$ gegen $H_A : \mu \notin \{\mu^*\}$
(b) $H_0 : \mu \in [0, \mu^*]$ gegen $H_A : \mu \in (\mu^*, 1]$
(c) $H_0 : \mu \in [\mu^*, 1]$ gegen $H_A : \mu \in [0, \mu^*)$

Teststatistik: $T = \frac{\overline{X} - \mu^*}{\sqrt{s^2(X)/n}}$ unter $\mathbf{P}_{(\mu^*, \sigma^2)}$ nach $t(n-1)$ verteilt.

Ablehnungsbereich:

(a) $(-\infty, t_{n-1,\alpha/2}) \cup (t_{n-1,1-\alpha/2}, \infty)$
(b) $(t_{n-1,1-\alpha}, \infty)$
(c) $(-\infty, t_{n-1,\alpha})$

Student-t-Verteilung $t(n-1)$ mit $n-1$ Freiheitsgraden; Aufgabe 10, Zusatzblatt

Kapitel 6: Statistik

Fishers exakter Test

Grundidee: Prüfen auf Unabhängigkeit von zwei Ausprägungen in einem Datensatz.

Kontingenztabelle	Merkmal 2 Möglichkeit 1	Merkmal 2 Möglichkeit 2	Σ
Merkmal 1, Möglichkeit 1	S_{11}	S_{12}	$S_{1\bullet}$
Merkmal 1, Möglichkeit 2	S_{21}	S_{22}	$S_{2\bullet}$
Σ	$S_{\bullet 1}$	$S_{\bullet 2}$	S

Kapitel 6: Statistik

Fishers exakter Test

Hypothesen: H_0: Merkmale unabhängig
H_A: Merkmale nicht unabhängig

Teststatistik: S_{11} ist gegeben $S_{1\bullet}, S_{2\bullet}, S_{\bullet 1}, S_{\bullet 2}$ hypergeometrisch verteilt nach $H(S_{1\bullet}, S, S_{\bullet 1})$.

Ablehnungsbereich: $S_{11} \in C := \{0, ..., k, l, ..., S_{1\bullet} \wedge S_{\bullet 1}\}$, falls $H(S_{1\bullet}, S, S_{\bullet 1})(C) \leq \alpha$.

Hypergeometrische Verteilung

$$H(S_{1\bullet}, S, S_{\bullet 1})(k) = \frac{\binom{S_{\bullet 1}}{k}\binom{S_{\bullet 2}}{S_{1\bullet}-k}}{\binom{S}{S_{1\bullet}}}$$

Interpretation: S Gesamtmenge, $S_{1\bullet}$ gezogene Elemente, $S_{\bullet 1}$ Anzahl an möglichen Erfolgen.

Hilfreiche Aussagen

Nützliche Ungleichungen

- Cauchy-Schwartz-Ungleichung:

$$\mathbf{Cov}[X, Y]^2 \leq \mathbf{Var}[X]\mathbf{Var}[Y]$$

- Markov-Ungleichung: X $\mathbb{R}_+$-wertige ZV. Für alle $\varepsilon > 0$:

$$\mathbf{P}(X \geq \varepsilon) \leq \frac{\mathbf{E}[X]}{\varepsilon}$$

- Chebychev-Ungleichung: X $\mathbb{R}$-wertige ZV mit $\mu := \mathbf{E}[X] < \infty$. Für alle $\varepsilon > 0$:

$$\mathbf{P}(|X - \mu| \geq \varepsilon) \leq \frac{\mathbf{Var}[X]}{\varepsilon^2}$$

Hilfreiche Aussagen

Nützliche Gleichungen

- Geometrische Reihe: Für $0 \leq p < 1$ ist $\sum_{k=0}^{\infty} p^k = \frac{1}{1-p}$.
- Exponentialreihe: $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$.