Merkzettel für Statistik

erstellt aus dem Vorlesungsskript
"Wahrscheinlichkeitstheorie & Statistik"
von Prof. Leitner

§0 Anmerkung:

Der Merkzettel liefert nur eine Auswahl der wichtigsten Definitionen und Sätze aus dem Skript der Statistik Vorlesung des 4. Semesters, und erhebt keinen Anspruch auf Vollständigkeit.

§1 Einführung:

Definition
Sei (Ω, P) ein Wahrscheinlichkeitsraum (W - Raum) mit $\Omega = \{w_1, w_2, ..., w_n\}$.
Wird jedem Elementereignis $\{w_i\}$ die gleiche Wahrscheinlichkeit zugeordnet, also
$P(\{w_i\}) = \frac{1}{n}$, so spricht man von einem *Laplace-Experiment*.
Für ein beliebiges A gilt : $P(A) = \frac{\text{Anzahl der günstigen Ergebnisse}}{\text{Anzahl aller Ergebnisse}}$.

§2 Kombinatorik

Satz (2.1)
Die Anzahl der k - Tupel aus einer n - elementigen Menge ist n^k.

Satz (2.2)
Die Anzahl der Permutationen einer n - elementigen Menge ist $n!$

Satz (2.3)
Die Anzahl der k - Permutationen aus einer n - elementigen Menge ist $\frac{n!}{(n-k)!}$; $k \leq n$

Satz (2.4)
Die Anzahl der k - Teilmengen aus einer n - elementigen Menge ist $\frac{n!}{(n-k)!k!} = \binom{n}{k}$

Satz (2.5)
In einer Urne mit N Kugeln seien S schwarze und $N - S$ weisse Kugeln.
Sei X die Anzahl der gezogenen schwarzen Kugeln
a) Ziehen von n Kugeln ohne zurücklegen $(n \leq N)$

$$P(X = s) = \frac{\binom{S}{s} \cdot \binom{N-S}{n-s}}{\binom{N}{n}}$$

b) Ziehen von n Kugeln mit Zurücklegen

$$P(X = s) = \binom{n}{s} \cdot \left(\frac{S}{N}\right)^s (1 - \frac{S}{N})^{n-s}$$

§3 Die bedingte Wahrscheinlichkeit:

Definition
Sei (Ω, P) ein W - Raum, $A \subset \Omega$, $P(A) \neq 0$,
dann heißt $P(B \mid A) = P_A(B) = \dfrac{P(A \cap B)}{P(A)}$
die bedingte Wahrscheinlichkeit von B unter der Bedingung A

Satz (3.1) Produktregel
$P(A \cap B) = P(A) \cdot P(B \mid A) = P(B) \cdot P(A \mid B)$

Satz (3.2) Formel von Bayes
Es gilt :

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

§4 Unabhängigkeit:

Definition
Zwei Ereignisse heißen *unabhängig*, wenn gilt
$P(A \cap B) = P(A) \cdot P(B)$.
Andernfalls heißen sie *abhängig.*

Satz (4.1)
Sind A, B unabhängig, so auch $\overline{A}, B$ und $A, \overline{B}$ und $\overline{A}, \overline{B}$.

§5 Zufallsgröße:

Definition
Zwei Zufallsgrößen X und Y heißen unabhängig, wenn gilt :
$P(X = x \text{ und } Y = y) = P(X = x) \cdot P(Y = y)$ für alle $x, y \in R$

§6 Erwartungswert und Varianz:

Definition
Erwartungswert: Schwerpunkt der Wahrscheinlichkeitsverteilung

Sei X eine Zufallsgröße auf einen Wahrscheinlichkeitsraum (Ω, P).
Der Wert, wobei X die Werte $x_1, x_2, ..., x_n$ annimmt,
$E(X) = EX = x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) + ... + x_n \cdot P(X = x_n)$
heißt *Erwartungswert* von X.

Definition
Sei X eine Zufallsgröße mit den Werten $x_1, x_2, ..., x_n$ mit dem Erwartungswert EX = μ

$$VarX = \sum_{i=1}^{k} (x_1 - \mu)^2 \cdot P(X = x_i) = E((X - \mu)^2)$$

heißt *Varianz* von X.
Die Größe $\sqrt{VarX}$ heißt Standardabweichung, und wird auch mit σ oder σ(X) bezeichnet.
Also $\sigma^2 = VarX$.

Satz (6.1)
$VarX = E(X^2) - (EX)^2$

Satz (6.2) Rechenregeln für Erwartungswerte
Seien X, Y Zufallsvariablen, dann gilt
(i) $E(X + Y) = EX + EY$
(ii) $E(aX + b) = aEX + b$
(iii) sind X und Y zusätzlich unabhängig, dann gilt : $E(XY) = EX \cdot EY$

Satz (6.3) Rechenregeln für Varianz
Seien X, Y unabhängige Zufallsvariablen
(i) $Var(X + Y) = VarX + VarY$,
insbesondere $Var(X + a) = VarX$
(ii) $Var(aX) = a^2 Var(X)$

§7 Die Ungleichung von Tschebischew:

Satz (7.1)
Sei X eine Zufallsgr öße und $c > 0$, dann gilt :

(1) $P(|X - \mu| \geq c) \leq \frac{VarX}{c^2}$

(2) $P(|X - \mu| > c) < \frac{VarX}{c^2}$

(3) $P(|X - \mu| < c) \geq 1 - \frac{VarX}{c^2}$

(4) $P(|X - \mu| \leq c) > 1 - \frac{VarX}{c^2}$

Satz (7.2) Arithmetische Mitte

$E\overline{X} = \mu$

$Var\overline{X} = \frac{\sigma^2}{n}$

$\sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}}$, ($\sqrt{n}$ - Gesetz)

Satz (7.3) Tschebischew Ungleichung für arithmetische Mitte
Es gilt :

$P(|\overline{X} - \mu| \geq c) \leq \frac{\sigma^2}{nc^2}$

$P(|\overline{X} - \mu| > c) < \frac{\sigma^2}{nc^2}$

$P(|\overline{X} - \mu| < c) \geq 1 - \frac{\sigma^2}{nc^2}$

$P(|\overline{X} - \mu| \leq c) > 1 - \frac{\sigma^2}{nc^2}$

Satz (9.3) Ungleichung von Tschebischew für Bernoulli Ketten
Sei X nach $B(n, p)$ verteilt.

$\frac{X}{n}$ beschreibt die relative Häufigkeit eines Treffers, also

$E\frac{X}{n} = \frac{1}{n}EX = p$

$Var\frac{X}{n} = \frac{1}{n^2}VarX = \frac{p \cdot q}{n}$, also

$P(\left|\frac{X}{n} - p\right| \geq \varepsilon) \leq \frac{p \cdot q}{n \cdot \varepsilon^2}$, bzw.

$P(\left|\frac{X}{n} - p\right| < \varepsilon) \geq 1 - \frac{p \cdot q}{n \cdot \varepsilon^2}$

$P(\left|\frac{X}{n} - p\right| > \varepsilon) < \frac{p \cdot q}{n \cdot \varepsilon^2}$

$P(\left|\frac{X}{n} - p\right| \leq \varepsilon) > 1 - \frac{p \cdot q}{n \cdot \varepsilon^2}$

Wegen $p \cdot q = p(1 - q) = p - p^2 \leq \frac{1}{4}$ kann man sogar eine von p unabhängige Abschätzung angeben :

$P(\left|\frac{X}{n} - p\right| \geq \varepsilon) \leq \frac{1}{4n \cdot \varepsilon^2}$

§9 Die Binomialverteilung:

Satz (9.1)
Ein Bernoulli Experiment habe die Trefferwahrscheinlichkeit p und die Nietenwahrscheinlichkeit $1-p=q$.
Dann gilt für die Bernoulli Kette der Länge n:

$$B(n,p,k)=\binom{n}{k}\cdot p^k q^{n-k}$$

Sei $X_1, X_2, ..., X_n$ Bernoulli Kette und $X = X_1 + X_2 + ... + X_n$, dann ist
$P(X=k)=B(n,p,k)$
Man sagt : X ist verteilt nach $B(n,p)$

Definition Wahrscheinlichkeitsverteilung

$$B(n,p): k \mapsto \binom{n}{k}\cdot p^k q^{n-k}$$

heißt *Binomialverteilung* $B(n,p)$ mit Parameter n und p.
Satz (9.2)
Ist X verteilt nach $B(n,p)$, dann gilt
$EX = n\cdot p,$
$VarX = n\cdot p\cdot q,$
$\sigma(X)=\sqrt{n\cdot p\cdot q}$

Definition
Die Verteilung

$$s \mapsto \frac{\binom{S}{s}\cdot\binom{N-S}{n-s}}{\binom{N}{n}}$$

heißt *hypergeometrische Verteilung* $H(N,S,n)$
Die hypergeometrische Verteilung ist aufwendig zu tabellieren.
Für $n << N$ ist $H(N,S,n)$ gut durch $B(n,\frac{S}{N})$ approximierbar!
Satz (9.5) Poisson Verteilung
Für die Wahrscheinlichkeit $B(n,p,k)$ gilt die Poisson Näherung

$$B(n,p,k)\approx\frac{\mu^k}{k!}e^{-\mu} \quad (\mu = n\cdot p),$$

die brauchbare Werte liefert für $\frac{\mu}{n} << 1$, d.h. "p klein" und $k << $ n.
Die Funktion

$$P(\mu): k \mapsto \frac{\mu^k}{k!}e^{-\mu} \quad ; k\in N_0$$

ist eine Wahrscheinlichkeitsverteilung.
$P(\mu)$ heißt Poisson Verteilung mit Parameter μ.
Für eine nach $P(\mu)$ verteilte Zufallsvariable X gilt:
$EX = \mu$
$VarX = \mu$

§10 Die Normalverteilung:

Satz (10.2)
Sei X eine nach $B(n, p)$ verteilte Zufallsgröße, so gilt

$$P(X \le k_2) \approx \Phi(\frac{k_2 - \mu + 0{,}5}{\sigma})$$

$$P(X = k_2) \approx \Phi(\frac{k_2 - \mu + 0{,}5}{\sigma}) - \Phi(\frac{k_2 - \mu - 0{,}5}{\sigma}),$$

$$P(k_1 \le X \le k_2) \approx \Phi(\frac{k_2 - \mu + 0{,}5}{\sigma}) - \Phi(\frac{k_1 - \mu - 0{,}5}{\sigma}),$$

für $k_1, k_2 = 1,2,...,n$

$\mu = n \cdot p$

$\sigma = \sqrt{n \cdot p \cdot q}$

Satz (10.3)
Für große n gilt näherungsweise

$$B(n, p, k) = \binom{n}{k} p^k (1-p)^{n-k} \approx \frac{1}{\sigma}\varphi\left(\frac{k-\mu}{\sigma}\right) = \frac{1}{\sqrt{n \cdot p \cdot (1-p)}} \cdot \Phi\left(\frac{k - n \cdot p}{\sqrt{n \cdot p(1-p)}}\right)$$

Satz (10.4)
Für große n gilt :

$$P(|X - \mu| \le c) \approx 2\Phi\left(\frac{c + 0{,}5}{\sigma}\right) - 1 \approx\approx 2\Phi\left(\frac{c}{\sigma}\right) - 1$$

$\mu = EX$

$\sigma = \sqrt{VarX}$

Aus den Übungsaufgaben:

$$P(|X - \mu| \le c) = P(\left|\frac{X}{n} - p\right| \le \frac{c}{n}) \approx 2\Phi(\frac{c + 0{,}5}{\sigma}) - 1 \quad \Rightarrow \quad P(\left|\frac{X}{n} - p\right| \le t) \approx 2\Phi(\frac{t \cdot n + 0{,}5}{\sigma}) - 1$$

Prüfen von Hypothesen:

I. Die Standardaufgabe, gesucht wird der (unbekannte) Mittelwert
1. Ermittlung des arithmetischen Mittels

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

2. Ermittlung der erwartungstreuen Schätzers für σ, wenn σ nicht angegeben ist

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \text{ dann } \sigma = \sqrt{S^2}$$

3. Entscheidung treffen

$$\left|\frac{\overline{X} - \xi_0}{\sigma / \sqrt{n}}\right| \le \lambda^{(n-1)}{}_{P\%} \qquad \Rightarrow \qquad \text{Annahme der Hypothese}$$

$$\left|\frac{\overline{X} - \xi_0}{\sigma / \sqrt{n}}\right| > \lambda^{(n-1)}{}_{P\%} \qquad \Rightarrow \qquad \text{Ablehnung der Hypothese}$$

=> Tabelle Student-T Verteilung

Bei Test $\xi = \xi_0$ gegen $\xi \neq \xi_0$

90% Sicherheitswahrscheinlichkeit	=	0,95
95% Sicherheitswahrscheinlichkeit	=	0,975
99% Sicherheitswahrscheinlichkeit	=	0,995

Bei Test $\xi \geq \xi_0$ gegen $\xi < \xi_0$

90% Sicherheitswahrscheinlichkeit	=	0,90
95% Sicherheitswahrscheinlichkeit	=	0,95
99% Sicherheitswahrscheinlichkeit	=	0,99

II. Gesucht wird ein Intervall, in dem sich der Mittelwert mit P% Wahrscheinlichkeit befindet:
1. Ermittlung des arithmetischen Mittels (siehe oben)
2. Ermittlung der erwartungstreuen Schätzers für σ, wenn σ nicht angegeben ist (siehe oben)
3. Ermittlung des Intervalls

$$-\lambda^{(n-1)}{}_{P\%} \le \left|\frac{\overline{X} - \xi}{\sigma / \sqrt{n}}\right| \le \lambda^{(n-1)}{}_{P\%}$$

4. nach ξ auflösen

III. Standardaufgabe, gesucht wird die Standardabweichung (Streuung)
1. Ermittlung des arithmetischen Mittels (siehe oben)
2. Ermittlung der erwartungstreuen Schätzers (siehe oben)
3. Das Wert für die χ^2 –Verteilung mit n-1 Freiheitsgraden ermitteln

=> Tabelle Chi-Quadrat Verteilung

Zweiseitiger Test:
Bei Hypothese $\sigma = \sigma_0$ gegen $\sigma \neq \sigma_0$

$\chi^{2\,(n-1)}_{\text{rechts, P\%}} < \frac{(n-1)S^2}{\sigma_0^2} < \chi^{2\,(n-1)}_{\text{links, P\%}}$ => Ablehnung der Hypothese

$\chi^{2\,(n-1)}_{\text{links, P\%}} \leq \frac{(n-1)S^2}{\sigma_0^2} \leq \chi^{2\,(n-1)}_{\text{rechts, P\%}}$ => Annahme der Hypothese

Beispiel:
Bei P% = 95% und 9 Freiheitsgraden
=> links: 0.025 / 9 FG = 2.70
=> rechts: 0.975 / 9 FG = 19.02

Einseitiger Test:
Bei Hypothese $\sigma \geq \sigma_0$ gegen $\sigma < \sigma_0$

$\frac{(n-1)S^2}{\sigma_0^2} < \chi^{2\,(n-1)}_{\text{einseitig rechts, P\%}}$ => Ablehnung der Hypothese

Beispiel:
Bei P% = 95% und 7 Freiheitsgraden
=> einseitig rechts: 0.05 / 7 FG = 2.17

Bei Hypothese $\sigma \leq \sigma_0$ gegen $\sigma > \sigma_0$

$\frac{(n-1)S^2}{\sigma_0^2} > \chi^{2\,(n-1)}_{\text{einseitig links, P\%}}$ => Ablehnung der Hypothese

Beispiel:
Bei P% = 95% und 7 Freiheitsgraden
=> einseitig links: 0.95 / 7 FG = 14.07

IV. Gesucht wird ein Intervall, in dem sich die gesuchte Standardabweichung mit P% Wahrscheinlichkeit befindet:
1. Ermittlung des arithmetischen Mittels (siehe oben)
2. Ermittlung der erwartungstreuen Schätzers (siehe oben)
3. Ermittlung des Intervalls

$\chi^{2\,(n-1)}_{\text{links, P\%}} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^{2\,(n-1)}_{\text{rechts, P\%}}$

4. Nach σ auflösen
Beispiel:
Bei P% = 90% und 8 Freiheitsgraden
=> links: 0.05 / 9 FG = 2.73
=> rechts: 0.95 / 9 FG = 15.51

V. Vergleich auf Gleichheit von zwei normalverteilten Grundgesamtheiten mit der selben unbek. Streuung σ
Vorgegeben:
H_0: $\xi = \eta$ bei einer Sicherheitswahrscheinlichkeit von P%
Stichproben X_n und Y_m

1. Ermittle arithmetische Mittel $\overline{x}$ und $\overline{y}$
2. berechne:

$$t = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{1}{n} + \frac{1}{m}} \sqrt{\frac{1}{n+m-2} \sum_{i=1}^{n} (x_i - \overline{x})^2 + \sum_{i=1}^{m} (y_i - \overline{y})^2}}$$

3. Entscheidung:

$|t| > \gamma_{P\%}^{(n-m+2)}$ => Hypothese wird abgelehnt

$|t| < \gamma_{P\%}^{(n-m+2)}$ => Hypothese wird angenommen

=> Tabelle Student-T Verteilung

Soll ξ - η = C ein konstanter Unterschied festgestellt werden, wird der Zähler ersetzt durch
$\overline{x}$ - $\overline{y} - C$

Aufgabentypen:

Gegeben ist - Wahrscheinlichkeit p für ein Experiment
- Sicherheitswahrscheinlichkeit P%

Frage: Wie viele Experimente, damit p mit Sicherheit P% eintrifft ?

Beispiel: - Wahrscheinlichkeit 3%
- Sicherheitswahrscheinlichkeit 95%

Ansatz:

$$P(3\%) = 1 - P(97\%) = 1 - 0{,}97^n \geq 0{,}95 \Rightarrow 0{,}97^n \leq 0{,}05 \Rightarrow n \cdot \ln 0{,}97 \leq \ln 0{,}05 \Rightarrow n \geq \frac{\ln 0{,}05}{\ln 0{,}97} = 98{,}35 \Rightarrow n \geq 99$$

Gegeben ist - Wahrscheinlichkeit p für ein Experiment (ggf. umkehren)
- Anzahl der Versuche (<10)
oder
- Anzahl der Versuche (>10)
- Anzahl der Treffer

Frage: Wie groß ist die Wahrscheinlichkeit für dieses Experiment ?

Versuche <10:

$$B(n,p): k \mapsto \binom{n}{k} \cdot p^k q^{n-k}$$

Versuche >10:

$$P(X \leq k_2) \approx \Phi\left(\frac{k_2 - \mu + 0{,}5}{\sigma}\right)$$

$$\mu = n \cdot p$$

$$\sigma = \sqrt{n \cdot p \cdot q}$$

Gegeben: - Zufallsexperiment X

Frage: Wie lautet der Erwartungswert (Durchschnitt), die Varianz oder die Standardabweichung von X ?

$$\mathrm{EX} = \mu$$

$$E(X) = EX = x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) + \ldots + x_n \cdot P(X = x_n)$$

VarX:

$$VarX = \sum_{i=1}^{k} (x_1 - \mu)^2 \cdot P(X = x_i) = E((X - \mu)^2)$$

oder

$$VarX = E(X^2) - (EX)^2$$

Standardabweichung:

Die Größe $\sqrt{VarX}$ heißt Standardabweichung, und wird auch mit σ oder σ(X) bezeichnet.

Kombinatorik:

Gegeben: - Eine Menge Ω
- Eine Vorschrift A

Frage: Wie groß ist die Wahrscheinlichkeit, daß A eintrifft ?

1. Ermittlung von $|\Omega|$
2. Ermittlung von $|A|$
3. Wahrscheinlichkeit = $\frac{|A|}{|\Omega|}$