# Wahrscheinlichkeitstheorie und Statistik

ETH HS18
Juni 2019 Noah Hütter
40[redacted]-0604-00



## Grundbegriffe

**Grundraum $\Omega$:** Kollektion aller elementarer Ereignissen

Würfel: $\boxed{\Omega = \{1,2,3,4,5,6\}}$

**Elementarereignis $\omega$:** Eine Realisierung des Experiments

$\boxed{\omega \in \Omega}$

**W-Raum** $(\Omega, \mathcal{F}, P)$: $\mathcal{F} \subseteq \mathcal{P}(\Omega)$

$\mathcal{P}(\Omega)$: Potenzmenge: Menge aller Teilmengen

**P:** W-Mass $\boxed{P(\Omega) = 1}$

$A, B \in \mathcal{F}$: $\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$

$\boxed{P(\emptyset) = 0}$ $\boxed{P(A^c) = 1 - P(A)}$ $\boxed{A \subset B \Rightarrow P(A) \leq P(B)}$

**Laplace Modell:** Alle Ereignisse gleich wahrscheinlich

$\mathcal{F} = \mathcal{P}(\Omega)$ $\boxed{P(\omega) = \frac{1}{|\Omega|}}$ $\boxed{P(A) = \frac{|A|}{|\Omega|}}$

**Permutation:** Anz. Kombinationen mit n Elementen aus k. Reihenfolge beachten
(= n untersch. Objekte auf k Plätze verteilen)

$$\boxed{P(n,k) = \frac{n!}{(n-k)!}}$$

**Potenz:** k Stellen mit je n Möglichkeiten
Reihenfolge beachten

$\boxed{\# = n^k}$

**Zufallsvariable:** X ist eine Abbildung $X: \Omega \to \mathbb{R}$ s.d. für
$a \in \mathbb{R}$: $\{\omega \in \Omega, X(\omega) \leq a\} \in \mathcal{F}$

**Kombination** (Binomialkoeffizient): Anz. Komb. mit n Elementen aus k. Reihenfolge egal
(= n identische Obj. auf k Plätze verteilen)

$\boxed{C(n,k) = \binom{n}{k} = \frac{n!}{(n-k)!\,k!}}$ $\boxed{\binom{n}{k} = \binom{n}{n-k}}$ $\boxed{\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}}$

$\boxed{(x+y)^n = \sum_{p=0}^{n} \binom{n}{p} x^{n-p} y^p}$

n, k

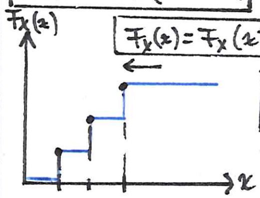
1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1

## Diskrete Verteilungen

### Verteilfunktion

**Zufallsvariable X:** Wkeit, dass X den Wert kleiner als x annimmt ist die Verteilfunktion

$\boxed{F_X(x) = P(X \leq x)}$ $F_X(x)$ ist monoton$^+$ rechtsstetig




$\boxed{F_X(x) = F_X(x^+)}$ $\boxed{F_X(x^-) = P(X < x)}$

$\boxed{\lim_{x \to -\infty} F_X(x) = 0}$ $\boxed{\lim_{x \to +\infty} F_X(x) = 1}$

$\boxed{F_X(x) = \lim_{h \downarrow 0} F_X(x+h) = F_X(x^+)}$

Sprunghöhe bei x: $\boxed{P(X=x) = F_X(x) - F_X(x^-)}$

$P(a < X \leq b) = F_X(b) - F_X(a)$
$P(a < X < b) = F_X(b^-) - F_X(a)$
$P(a \leq X < b) = F_X(b^-) - F_X(a^-)$

### Dichtefunktion

⚠ nur wenn $F_X(x)$ stetig ⚠

$\boxed{f_X(x) = \frac{d}{dx} F_X(x)}$ $\boxed{\int_{-\infty}^{\infty} f_X(x)dx = 1}$ $\boxed{F_X(x) = \int_{-\infty}^{x} f_X(s)ds}$

$P(X \leq a) = \int_{-\infty}^{a} f(x)dx$

### Funktion von r.V.

Gegeben: $f_Y(y)$, $U = g(Y)$ Gesucht: $F_U(a) = ?$

$f_Y \xrightarrow{\int} F_Y \xrightarrow{\text{Umformen}} F_{g(Y)=U} \xrightarrow{\frac{d}{dx}} f_U$

⚠ Fallunterscheidung & Wertebereich ⚠

1 Wertebereich $Y \in [0, \infty) \to U \in [?, ?]$

### Bedingte WSK

$P(A|B) =$ "Wsk dass A eintrifft wenn B eintrat"

$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}}$ $P(A \cap B) = P(A, B)$ – Beide Ereignisse treffen ein

$\boxed{P(A) = \sum_i^n P(A|B_i) \cdot P(B_i)}$ $\boxed{P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{\underbrace{\sum_j^n P(A|B_j) \cdot P(B_j)}_{= P(A)}}}$

Totale WSK Bayes

$P(B_i)$: „a priori Wsk der Ursache $B_i$"
$P(B_i|A)$: „a posteriori Wsk der Ursache $B_i$"

$\boxed{P(A^c|B_i) = 1 - P(A|B_i)}$

### Unabhängigkeit

Zwei Ereignisse heissen unabhängig, falls

$\boxed{A, B \text{ unabh.} \iff P(A \cap B) = P(A) \cdot P(B)}$

Bei mehreren Ereignissen A, B, C

$A, B, C$ unabh. $\iff$
$P(A \cap B) = P(A) \cdot P(B)$
$P(A \cap C) = P(A) \cdot P(C)$
$P(B \cap C) = P(B) \cdot P(C)$
$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$

Bedingte WSK: $\boxed{A, B \text{ unabh.} \iff P(A|B) = P(A),\ P(B|A) = P(B)}$

Für Zufallsvariablen X, Y gilt

$X, Y$ unabh. $\iff P(X=x, Y=y) = P(X=x) \cdot P(Y=y)$
$\iff f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$
$\iff p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$
$\Rightarrow E[XY] = E[X] \cdot E[Y]$
$\Rightarrow Var(X+Y) = Var(X) + Var(Y)$
$\Rightarrow Cov(X,Y) = 0$

# Diskrete Verteilungen

## Bernoulli Bin(1, p)

Einfaches 0-1 Experiment mit Erfolgsparameter p

$$\boxed{p_X(k) = \begin{cases} p, & k=1 \\ 1-p, & k=0 \end{cases}} \quad \boxed{m_X = p} \quad \boxed{\sigma_X^2 = p(1-p)}$$

$W_X = \{0,1\}$

$$\varphi_X(t) = E[e^{itX}] = (1-p) + p\cdot e^{it}$$

$p \in [0,1]$ Erfolgsparameter

## Binomial Bin(n, p)

Anz. Erfolge bei n unabh. 0-1-Experimenten

$$\boxed{p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}} \quad \boxed{m_X = np} \quad \boxed{\sigma_X^2 = np(1-p)}$$

$W_X = \{0, .., n\}$

$$\varphi(t) = E[e^{itX}] = ((1-p) + pe^{it})^n$$

$n \in \mathbb{N}$ Anz. Wiederholungen

## Geometrisch Geom(p)

Erster Erfolg bei $\infty$ unabh. 0-1 Experimenten

$$\boxed{p_X(k) = p\cdot(1-p)^{k-1}} \quad \boxed{m_X = \frac{1}{p}} \quad \boxed{\sigma_X^2 = \frac{1-p}{p^2}}$$

$W_X = \{1, 2, ..\}$ $\quad X = \inf\{k \in [1, .., |X_k = 1]\}$ ← kl. idx k s.d. $X_k = 1$

$$\varphi(t) = E[e^{itX}] = \frac{p\cdot e^{it}}{1-(1-p)e^{it}}$$

## Poisson Pois(λ)

Bin(n,p) mit $n \to \infty$, p sehr klein: $np \approx \lambda$

$$\boxed{p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}} \quad \boxed{m_X = \lambda} \quad \boxed{\sigma_X^2 = \lambda} \quad W_X = \{0, 1, ..\}$$

$$\varphi(t) = E[e^{itX}] = e^{\lambda(e^{it}-1)}$$

$\lambda > 0$: Poisson-Parameter

↑ da $\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x$

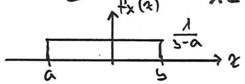
wenn $Y \sim Pois(\lambda)$ $Z \sim Pois(4\lambda)$ ist $Y+Z \sim Pois(5\lambda)$

# Stetige Verteilungen

## Gleichverteilt Uniform(a,b)

$$\boxed{f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & \text{sonst} \end{cases}} \quad \boxed{F_X(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & x \in [a,b] \\ 1, & x > b \end{cases}}$$

$$m_X = \frac{a+b}{2} \quad \sigma_X^2 = \frac{1}{12}(b-a)^2 \quad \varphi_X(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}$$

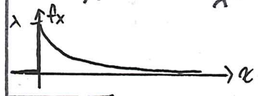



a, b: linke, rechte Grenze

## Exponential Exponential(λ)

$$\boxed{f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{sonst} \end{cases}} \quad \boxed{F_X(x) = \begin{cases} 1-e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}}$$

$$m_X = \frac{1}{\lambda} \quad \sigma_X^2 = \frac{1}{\lambda^2} \quad \varphi_X(t) = \frac{\lambda}{\lambda - it}$$



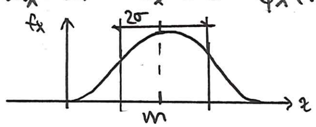

$\lambda > 0$: Rate

## Normal $\mathcal{N}(m, \sigma)$

$$\boxed{f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}} \quad F_X(x) = \phi\left(\frac{x-m}{\sigma}\right)$$

↑ tabelliert

$$m_X = m \quad \sigma_X^2 = \sigma \quad \varphi_X(t) = e^{imt - \frac{1}{2}\sigma^2 t^2}$$




$m \in \mathbb{R}$ Zentrierungsparameter

$\sigma > 0$ Breitenparameter

# Erwartungswert & Varianz

## Erwartungswert $m_X = E[X]$

$$\boxed{E[X] = \sum_x x\cdot p_X(x)} \text{ diskret} \quad \boxed{E[X] = \int_{-\infty}^{\infty} x\cdot f_X(x)dx} \text{ stetig}$$

$$\boxed{E[Y] = E[g(X)] = \sum_x g(x)\cdot p_X(x) = \int_{-\infty}^{\infty} g(x)\cdot f_X(x)dx} \text{ Funktion von r.v.}$$

Linearität: $\boxed{E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y]}$

$$\boxed{P(X \geq a) \leq \frac{1}{a} E[X]} \quad \boxed{P(|X - m_X| \geq a) \leq \frac{1}{a^2} Var(X)}$$

Markov — Tschebycheff

## Varianz $Var(X) = \sigma_X^2$

$$\boxed{\sigma_X^2 = E[(X-m_X)^2] = E[X^2] - m_X^2}$$

$Var(\alpha X) = \alpha^2 \cdot Var(X)$ $\quad$ Std. Abw. $\sigma_X = \sqrt{Var(X)}$

# Gemeinsame Dichte & Verteilung

## Stetige r.v. $f_X, F_X$

$$\boxed{P(X \in [a_1, b_1], Y \in [a_2, b_2]) = \int_{a_2}^{b_2}\int_{a_1}^{b_1} f_{X,Y}(x,y)dx\,dy}$$

↑ gem. Vert. von X & Y

Randdichten $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$ $\quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$

Erwartungswert: $E[g(X,Y)] = \iint_{-\infty}^{\infty} g(x,y)\cdot f_{X,Y}(x,y)dx\,dy$

Normierung: $\boxed{\iint_{-\infty}^{\infty} f_{X,Y}(x,y)dx\,dy = 1}$

## Diskrete r.v. $p_{X,Y}$

$$\boxed{p_{X,Y}(x,y) = P(X=x, Y=y)}$$

Randvert. $p_X(x) = \sum_{\forall y} p_{X,Y}(x,y)$ $\quad p_Y(y) = \sum_{\forall x} p_{X,Y}(x,y)$

Erwartungswert: $E[g(X,Y)] = \sum_x \sum_y g(x,y)\cdot p_{X,Y}(x,y)$

Normierung: $\boxed{\sum_x \sum_y p_{X,Y}(x,y) = 1}$

| Y \ X | $x_1$ | $x_2$ | |
|---|---|---|---|
| $y_1$ | a | b | $\Sigma = P(Y=y_1)$ |
| $y_2$ | c | d | $\Sigma = P(Y=y_2)$ |
| | $\Sigma = P(X=x_1)$ | $\Sigma = P(X=x_2)$ | |

$P(X=x_2, Y=y_1) = P(X=x_2) - d$
$P(X=x_1, Y=y_2) = P(Y=y_2) - d$ $\Big\}\ \epsilon\, 0..1$

$a+b+c+d = 1$

# Lineare Prognose

## Kovarianz

$$\boxed{Cov(X,Y) = E[(X-m_x)(Y-m_y)] = E[XY] - m_x m_y}$$

$Cov(X,Y) = Cov(Y,X)$ $\quad$ $Cov(X,X) = Var(X)$

$Var(X+Y) = Var(X) + Var(Y) + 2cov(X,Y)$

$X,Y$ unabhängig $\Rightarrow$ $Cov(X,Y)=0$

$\cancel{\Leftarrow}$ $\quad$ $Cov(A+B,C) = Cov(A,C) + Cov(B,C)$

## Korrelationskoeff.

$$\boxed{\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}}} \quad \in [-1,1]$$

## Beste lin. Prognose

Szenario: $U = V + W$. $W$ nicht bekannt (z.B. Rauschen)

$V$ messen und $U$ schätzen $\to \hat{U}$

$$\boxed{\hat{U} = \frac{Cov(U,V)}{Var(V)}(V - m_V) + m_U}$$

$\hat{U}$: beste lin. Prognose von $U$. $\hat{U}$ ist r.v. und abhängig von $V$

1. $m_V$ 2. $m_U$ 3. $Var(V)$ 4. $Cov(U,V)$ 5. $\hat{U} = ..$

# Statistik

## Chi-Quadrat Anpassungstest

Geg: hypothetisch diskrete Verteilung + Tabelle

| Kategorie $i$ | 1 | 2 | .. | n |
|---|---|---|---|---|
| Eintreten $N_i$ | .. | .. | .. | .. |

Ges.: Wollen hypothetisch Verteilung verworfen oder akzeptieren.

1. Nullhypothese formulieren
   $H_0 = \ldots$ entspricht einer ... Vert. mit Parameter...
2. Definiere Niveau des Tests $\alpha = ..$
3. Berechne $p_i$ jeder Kategorie unter Annahme, dass $H_0$ stimmt
4. $N = \sum_i N_i$, $\quad t = \sum_{i=1}^{n} \frac{(N_i - N\cdot p_i)^2}{N \cdot p_i}$
5. Freiheitsgrad $\nu = \#\text{Kategorien} - 1$
6. Lese aus Tabelle für $\chi^2_{(1-\alpha);\nu}$ den Wert $a$
7. $\begin{cases} t > a & H_0 \text{ verwerfen} \\ t \le a & H_0 \text{ akzeptieren} \end{cases}$

- Nur aussagekräftig wenn $H_0$ verworfen werden kann
- Kategorie zusammen fassen (z.B. $x \geq 7$)
- Mehr Daten $\Rightarrow$ aussagekräftiger Test

## Chi-Quadrat Unabhängigkeitstest

Ges.: sog. Kontingenztafel

| | | Merkmal $j$: 1 | 2 | -- | m | |
|---|---|---|---|---|---|---|
| Merkmal $i$ | 1 | $N_{11}$ | $N_{12}$ | | : | $N_{1*}$ |
| | 2 | $N_{21}$ | $N_{22}$ | $\ddots$ | | $N_{2*}$ |
| | : | | .. | | | : |
| | n | | | | | |
| | | $N_{*1}$ | $N_{*2}$ | .. | | $N$ |

$N_{ij}$ bezeichnet wieviele Vorkommnisse in kat. $(i,j)$

Ges.: Merkmale $i$ & $j$ unabhängig?

1. $H_0$: Merkmale ... und ... sind unabhängig
2. Niveau $\alpha = ..$
3. $p_{ij} = \frac{N_{ij}}{N}$ $\quad$ $N = \sum_i \sum_j N_{ij}$
4. $t = N \cdot \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{(p_{ij} - p_{i*}\cdot p_{*j})^2}{p_{i*}\cdot p_{*j}}$
5. Freiheitsgrad $\nu = (n-1)(m-1)$
6. $a$ aus Tabelle für $\chi^2_{(1-\alpha);\nu}$ lesen
7. $\begin{cases} t > a & H_0 \text{ verwerfen} \\ t \le a & H_0 \text{ akzeptieren} \end{cases}$

- $N_{i*}$ = Summe der i-ten Reihe

## Null- vs. Alternativhypothese

Geg: Hyp. stetige Verteilung A und stetige alternative Verteilung B und Menge an Stichproben $X_1, X_2, .., X_N$

Ges.: A verworfen oder akzeptieren. Wenn verworfen dann B annehmen

1. $H_0 := \ldots$ Entspricht einer A Verteilung mit Parameter...
   $H_1 := \ldots$ Entspricht einer B Verteilung mit Parameter...
   Alternativ
2. Niveau $\alpha = ..$
3. $\bar{X} = \frac{1}{N}\cdot(X_1 + \ldots + X_N)$
4. $E[\bar{X}_{H_0}] = ..$ $\quad$ $Var(\bar{X}_{H_0}) = ..$ $\quad$ $E[\bar{X}_{H_1}] = ..$ $\quad$ $Var(\bar{X}_{H_1}) = ..$
5. $m_{H_0} > m_{H_1}$ : Wähle $a$ sodass $\alpha = P(\bar{X}_{H_0} \leq a)$
   $m_{H_0} < m_{H_1}$ : Wähle $a$ sodass $\alpha = P(\bar{X}_{H_0} \geq a)$
6. $m_{H_0} > m_{H_1}$ : Verwerfe $H_0$ wenn $\bar{X} < a$
   $m_{H_0} < m_{H_1}$ : Verwerfe $H_0$ wenn $\bar{X} > a$

Macht: $\begin{cases} P(\bar{X}_{H_1} \leq a) & m_{H_0} > m_{H_1} \\ P(\bar{X}_{H_1} \geq a) & \text{else} \end{cases}$

Fehler:
1. Art: Obwohl $H_0$ stimmt, ist $a$ gross $\to$ man wählt $H_1$ ↯
2. Art: Obwohl $H_1$ stimmt, ist $a$ klein $\to$ man wählt $H_0$ ↯
$P(\text{Fehler 2. Art}) = 1 - \text{Macht}$

## Maximum-Likelihood Schätzer

Geg: Verteilung mit unbekannten Parametern $\lambda$
Messwerte $X_1, .., X_n \overset{iid}{\sim}$ Verteilung

Ges: Schätzung von $\lambda$

1. Idee: Maximiere gemeinsame Dichte/Verteilung
2. Definiere & Berechne Likelihood-Funktion
   $L(\lambda | x_1, .., x_n) = f_{(X_1,..,X_n)}(x_1,..,x_n|\lambda) = \prod_{i=1}^{n} f_{X_i}(x_i|\lambda)$
   $= P_{(X_1,..,X_n)}(x_1,..,x_n|\lambda) = \prod_{i=1}^{n} P_{X_i}(x_i|\lambda)$
3. Maximiere $\log(L) = \ell$
   $\ell(\lambda|x_1,..,x_n) = \log L(\lambda|x_1,..,x_n) \to \frac{\partial \ell}{\partial \lambda} \overset{!}{=} 0$
4. Löse nach $\lambda$ auf $\to \tilde{\lambda} = \ldots$ und setze numerische Werte ein

- iid: independent and identically distributed
- Bsp.: Poisson $L(\lambda|x_i) = e^{-n\lambda}\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!} \rightsquigarrow \tilde{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$\log(x\cdot y) = \log(x) + \log(y)$

$\log(x^a) = a \cdot \log(x)$

$\frac{d}{dx}\ln x = \frac{1}{x}$ $\quad$ $\frac{d}{dx}\log_b x = \frac{1}{x \cdot \ln(b)}$