Inhalt

Wahrscheinlichkeitstheorie Grundlagen 2
Verteilungen 6
Transformationen von Verteilungen 10
Erwartungswert und Varianz (Varianz = mittlere Quadratische Abweichung) 12
Gesetze der Großen Zahlen 17
Schätzertheorie 23
Intervallschätzung 28
Hypothesentests 29
Stochastik 35
  Markovketten 37
  Klasseneigenschaften 38
  Markov Ketten in stetiger Zeit 40

Vorwort: Die Bilder und Definitionen sind aus dem Skriptum vom Herrn Prof. Grill, Bilder aber auch Teils aus YouTube Videos von StatQuest etc.

Größtenteils nur Screenshots von Definitionen, manchmal auch mit eigenen Intuitionen oder Beispielen hinzugefügt.

Der Stochastik Part ist relativ mager, da ich den selbst nicht gut verstanden habe.

Wahrscheinlichkeitstheorie Grundlagen

Axiome von Kolmogorow („Logische Gesetzte der Wahrscheinlichkeiten“)

$0 \leq \mathbb{P}(A) \leq 1.$

$\mathbb{P}(\emptyset) = 0.$

$\mathbb{P}(\Omega) = 1.$

(Additivität) Wenn A und B disjunkte Ereignisse ($A \cap B = \emptyset$) sind, dann gilt

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

(Abzählbare Additivität, Sigmaadditivität) Wenn $A_n, n \in \mathbb{N}$ disjunkte Ereignisse sind (d.h., $A_i \cap A_j = \emptyset, i \neq j$), dann gilt

$$\mathbb{P}(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n).$$

Definition 2.2: Wahrscheinlichkeitsraum

Ein Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$ besteht aus einer Menge Ω, der *Grundmenge* und einem Wahrscheinlichkeitsmaß $\mathbb{P}$.

Die Elemente $\omega \in \Omega$ heißen *Elementarereignisse*, Teilmengen $A \subseteq \Omega$ *Ereignisse*. Das Wahrscheinlichkeitsmaß $\mathbb{P}$ ist auf der Menge dieser Ereignisse definiert (und erfüllt die Axiome von Kolmogorov).

Beispiele für **Wsl. Raum**:

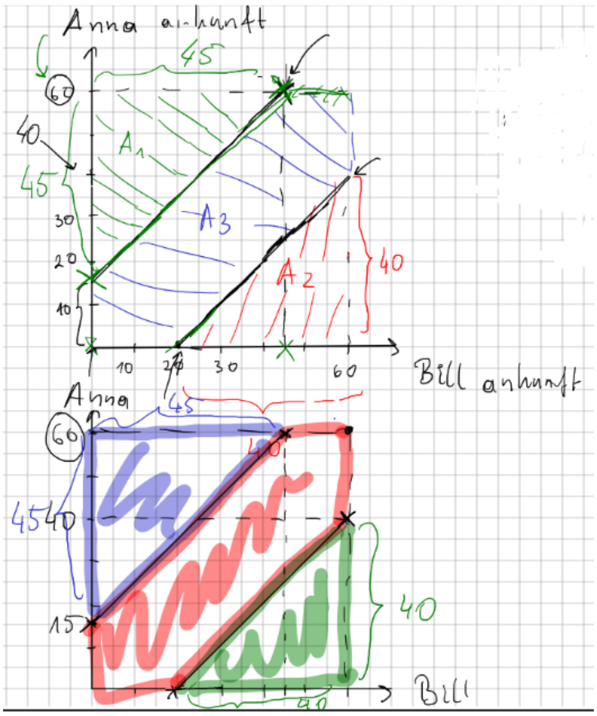
- Wsl. das nach n Münzenwürfen Kopf kommt: P({n}) = 2^-n und $\Omega = \mathbb{N}$
- Wsl. die mit dem Intervall steigt: P([a,b]) = b-a und $\Omega = [a, b]$

Definition 2.3: Laplacescher Wahrscheinlichkeitsraum

Wenn Ω endlich ist und

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}$$

(d.h. alle Elementarereignisse haben dieselbe Wahrscheinlichkeit), dann heißt der Wahrscheinlichkeitsraum ein Laplacescher Wahrscheinlichkeitsraum.



Prinzip nicht wirklich anwendbar bei **unendlichen** Wahrscheinlichkeitsräumen. Jedoch kann man das Prinzip **geometrisch** auch denken:

Links ist das Standard Beispiel. X-Achse ist Anna, y-Achse Herbert, beide kommen iwann zwischen der [0, 60] Minute an und trinken 10 min ihren Kaffee, was ist die Wsl. das sie sich treffen (rot markierte fläche)

Satz 2.1

1. $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$.
2. Wenn $A \subseteq B$, dann gilt $\mathbb{P}(A) \leq \mathbb{P}(B)$.
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
4. Für Ereignisse A_n mit $A_n \subseteq A_{n+1}$ gilt
$$\mathbb{P}(\bigcup_n A_n) = \lim_n \mathbb{P}(A_n).$$
5. Für Ereignisse A_n mit $A_n \supseteq A_{n+1}$ gilt
$$\mathbb{P}(\bigcap_n A_n) = \lim_n \mathbb{P}(A_n).$$
6. Für beliebige Ereignisse $A_n, n \in \mathbb{N}$ gilt
$$\mathbb{P}(\bigcup_n A_n) \leq \sum_n \mathbb{P}(A_n).$$

Satz 2.2: Additionstheorem

$A_1, \dots A_n$ seien beliebige Ereignisse. Dann gilt

$$\mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n (-1)^{i-1} S_i$$

mit

$$S_i = \sum_{1 \leq j_1 < j_2 < \dots < j_i \leq n} \mathbb{P}(A_{j_1} \cap \dots \cap A_{j_i}).$$

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C).$$

Definition 2.4

A und B seien zwei Ereignisse mit $\mathbb{P}(B) > 0$. Dann heißt

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

die bedingte Wahrscheinlichkeit von A unter (der Bedingung) B.

Satz 2.3: Multiplikationssatz

$A_1, \dots A_n$ seien Ereignisse mit $\mathbb{P}(A_1 \cap \dots \cap A_{n-1}) > 0$. Dann gilt

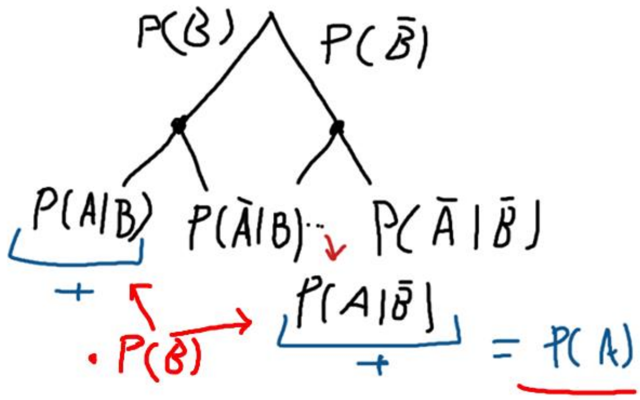
$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \dots \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1}).$$

Satz 2.4: Satz von der vollständigen Wahrscheinlichkeit

B_i seien disjunkte Ereignisse mit $\mathbb{P}(B_i) > 0$ und $\bigcup_i B_i = \Omega$ und A ein beliebiges Ereignis. Dann gilt

$$\mathbb{P}(A) = \sum_i \mathbb{P}(B_i)\mathbb{P}(A|B_i).$$

Wenn man es sich mit einem Baum überlegt: P(A) ist einfach die Summe aller Wege, die zu **Blättern führt,** wo **A** vorkommt.



Dieser Baum ist auch immer **umkehrbar** mithilfe des **Satz von Bayes** (Für Beispiele siehe Blutgruppen Bsp.)

Satz 2.5: Satz von Bayes

Unter denselben Voraussetzungen wie im vorigen Satz gilt (wenn der Nenner positiv ist)

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(B_j)\mathbb{P}(A|B_j)}{\sum_i \mathbb{P}(B_i)\mathbb{P}(A|B_i)}.$$

Verteilungen

Definition 2.6

Eine Zufallsvariable X ist eine Abbildung von Ω nach $\mathbb{R}^d$.

Beispiel: Zufallsvariable **S**, welche die Augensumme zweier gewürfelter Würfel definiert.

$S: \Omega \to \mathbb{R}; \quad (n_1, n_2) \mapsto n_1 + n_2,$

Definition 2.7

Die Verteilung einer Zufallsvariable ist das Wahrscheinlichkeitsmaß

$$\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega : X(\omega) \in A\})(A \subseteq \mathbb{R}^d).$$

Definition 2.8

Wenn der Wertebereich von X endlich oder höchstens abzählbar ist, dann nennen wir X diskret.

In diesem Fall kann die Verteilung von X durch die Wahrscheinlichkeitsfunktion

$$p_X(x) = \mathbb{P}(X = x) = \mathbb{P}_X(\{x\})$$

angegeben werden.

Definition 2.9

Die hypergeometrische Verteilung $H(N, A, n)$ $(n, A, N \in \mathbb{N},\ 0 \leq n, A \leq N)$ hat die Wahrscheinlichkeitsfunktion

$$p(x) = \frac{\binom{A}{x}\binom{N-A}{n-x}}{\binom{N}{n}}.$$

Ziehe mit Zurücklegen: Klassikerbeispiel **Lotto 45**: **Wahrscheinlichkeit** für einen **4rer**:

$$P(X = 4) = p_x(4) = H(45,6,6) = \frac{\binom{6}{4} \cdot \binom{39}{2}}{\binom{45}{6}}$$

Falls nicht zurückgelegt wird => **Binomialverteilung**

Definition 2.10

Die Binomialverteilung $B(n, p)$:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

Definition 2.11

X heißt alternativverteilt $(X \sim A(p))$ wenn

$$\mathbb{P}(X = 1) = p, \mathbb{P}(X = 0) = 1 - p.$$

Definition 2.12

Für ein Ereignis A heißt die Funktion

$$I_A : \Omega \to \{0, 1\}$$

mit

$$I_A(\omega) = \begin{cases} 1 & \text{wenn } \omega \in A, \\ 0 & \text{wenn } \omega \notin A \end{cases}$$

Indikator von A.

Definition 2.13

Die Verteilungsfunktion von X ist gegeben durch

$$F_X(x) = \mathbb{P}_X((-\infty, x]) = \mathbb{P}(X \leq x)$$

(wenn X d-dimensional ist, ist auch x d-dimensional, und die Ungleichung ist komponentenweise zu verstehen, also $X \leq x$ wenn $X_i \leq x_i$ für alle $i = 1, \ldots, d$ und $(-\infty, x] = (-\infty, x_1] \times \ldots \times (-\infty, x_d]$). Damit ist

$$F_X(x) = F_{(X_1,\ldots,X_d)}(x_1, \ldots, x_d) = \mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d).$$

Für $d = 1$ kann man die Verteilungsfunktionen einfach charakterisieren:

Satz 2.6

$F : \mathbb{R} \to \mathbb{R}$ ist genau dann eine Verteilungsfunktion, wenn

1. $0 \leq F(x) \leq 1$ für alle x,
2. F ist monoton nichtfallend,
3. F ist rechtsstetig,
4. $\lim_{x \to -\infty} F(x) = 0$,
5. $\lim_{x \to \infty} F(x) = 1$.

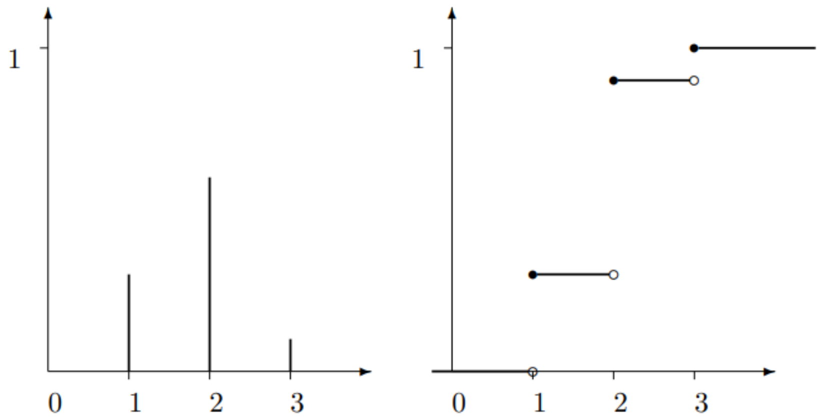


Abbildung 2.3: Wahrscheinlichkeits- und Verteilungsfunktion der hypergeometrischen Verteilung $H_{5,3,3}$

Satz 2.7

$F : \mathbb{R}^2 \to \mathbb{R}$ ist genau dann eine Verteilungsfunktion, wenn

1. $0 \leq F(x_1, x_2) \leq 1$ für alle x_1, x_2,
2. F ist monoton nichtfallend in jeder Argumentvariable,
3. F ist rechtsstetig,
4. $\lim_{x_1 \to -\infty} F(x_1, x_2) = \lim_{x_2 \to -\infty} F(x_1, x_2) = 0$,
5. $\lim_{x_1, x_2 \to \infty} F(x_1, x_2) = 1$,
6. Für $a_1 < b_1, a_2 < b_2$ gilt

$$F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \geq 0.$$

Definition 2.14

Wenn F_X in der Form

$$F_X(x) = \int_{-\infty}^{x} f_X(u) du$$

bzw.

$$F_X(x) = \int_{-\infty}^{x_d} \dots \int_{-\infty}^{x_1} f_X(u_1, \dots, u_d) du_1 \dots du_d$$

(falls $X = (X_1, \dots, X_d)$ mehrdimensional ist) darstellbar ist, dann heißt f_X die Dichte der Verteilung von X, und wir nennen X stetig (verteilt).

Für eine **Dichtefunktion** (**kontinuierlich**) muss gelten: Für eine **Dichtefunktion** (**kontinuierlich**) muss gelten:

$$f(x) \geq 0 \text{ für alle } x \in \mathbb{R}^d \text{ und } \int f(x)dx = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1 \dots, x_d) dx_1 \dots dx_d = 1.$$

Für eine **Wahrscheinlichkeitsfunktion** (**diskret**) muss gelten: $p(x) \geq 0$ für alle $x \in \mathbb{R}^d$ und $\sum p(x) = 1.$

Eine **Verteilung** wird genau dann gemischt genannt, wenn F(x) **nicht verschwindende Ableitungen** und **Sprünge** hat. Daraus folgt das die Verteilung eine **Wahrscheinlichkeitsfunktion** und eine **Dichtefunktion** hat.

Die Verteilungsfunktion muss daher eine Kombination aus den beiden Teilen sein.

$$\mathbb{P}(X \in A) = \sum_{x \in A} p_X(x) + \int_A f_X(x) dx.$$

Eine **Randdichte** (stetig) oder **Randverteilung** (diskret) wenn eine Verteilung (z.B bei einer 2-D Verteilung mit X und Y mit Tabelle) aus der **gemeinsamen** Verteilung gewonnen wird (z.B. X aus der Verteilung von X und Y)

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

$$p_X(x) = \sum_y p_{X,Y}(x,y)$$

Name	Symbol	$f(x)$	$\mathbb{E}(X)$	$\mathbb{V}(X)$
Gleichverteilung (stetig) $a < b$	$U(a,b)$	$\frac{1}{b-a}(a \leq x \leq b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential $\lambda > 0$	$E(\lambda)$	$\lambda e^{-\lambda x}(x \geq 0)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma $\alpha, \lambda > 0$	$\Gamma(\alpha,\lambda)$	$\frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)}e^{-\lambda x}(x \geq 0)$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
Cauchy $a > 0$	$C(a)$	$\frac{a}{\pi(x^2+a^2)}$	N.A.	N.A.
Normal $\mu \in \mathbb{R},\ \sigma^2 > 0$	$N(\mu,\sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(x-\mu)^2}{2\sigma^2})$	μ	σ^2
Beta 1. Art $\alpha, beta > 0$	$B_1(\alpha,\beta)$	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}(0 \leq x \leq 1)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$
Beta 2. Art $\alpha, beta > 0$	$B_1(\alpha,\beta)$	$\frac{x^{\alpha-1}(1+x)^{-\alpha-\beta}}{B(\alpha,\beta)}(0 \leq x)$	$\frac{\alpha}{\beta-1}\ (\beta > 1)$	$\frac{\alpha(\alpha+\beta-1)}{(\beta-2)(\beta-1)^2}\ (\beta > 2)$
Chiquadrat $n \in \mathbb{N}$	χ^2_n	$= \Gamma(n/2, 1/2)$		
Erlang $n \in \mathbb{N},\ \lambda > 0$	$Er(n,\lambda)$	$= \Gamma(n,\lambda)$		
t-Verteilung $n \in \mathbb{N}$	t_n	$\frac{2}{\sqrt{n}B(n/2,1/2)(1+x^2/n)^{(n+1)/2}}$	$0\ (n > 1)$	$\frac{n}{n-2}\ (n > 2)$
F-Verteilung $m, n \in \mathbb{N}$	$F_{n.m}$	$\frac{x^{n/2-1}n^{n/2}(1+nx/m)^{-(m+n)/2}}{m^{n/2}B(n/2,m/2)}(0 \leq x)$	$\frac{m}{m-2}\ (m > 2)$	$\frac{m^2(m+n-2)}{n(m-4)(m-2)^2}\ (m > 4)$

Name	Symbol	$p(x)$	$\mathbb{E}(X)$	$\mathbb{V}(X)$
Binomial $n \in \mathbb{N}, 0 \leq p \leq 1$	$B(n,p)$	$\binom{n}{x}p^x(1-p)^{n-x}(0 \leq x \leq n)$	np	$np(1-p)$
Gleichverteilung (diskret) $a \leq b, a, b \in \mathbb{Z}$	$D(a,b)$	$\frac{1}{b-a+1}(a \leq x \leq b)$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$
Geometrisch $0 < p < 1$	$G(p)$	$p(1-p)^x, (x \geq 0)$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
Geometrisch (alternativ) $0 < p < 1$	$G^*(p)$	$p(1-p)^{x-1}, (x \geq 1)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negativ Binomial $n > 0, 0 < p < 1$	$NB(n,p)$	$\binom{n+x-1}{n}p^n(1-p)^x, (x \geq 0)$	$\frac{n(1-p)}{p}$	$\frac{n(1-p)}{p^2}$
Negativ Binomial (alternativ) $n \in \mathbb{N}, 0 < p < 1$	$NB^*(n,p)$	$\binom{x-1}{n-1}p^{n-x}(1-p)^x, (x \geq n)$	$\frac{n}{p}$	$\frac{n(1-p)}{p^2}$
Poisson $\lambda > 0$	$P(\lambda)$	$\frac{\lambda^x e^{-\lambda}}{x!}$	λ	λ
Hypergeometrisch $N, A, n \in \mathbb{N},\ n, A \leq N$	$H(N,A,n)$	$\frac{\binom{A}{x}\binom{N-A}{n-x}}{\binom{N}{n}}(0 \leq x \leq n)$	$\frac{nA}{N}$	$\frac{nA(N-A)(N-n)}{N^2(N-1)}$

Bsp: **Exponentialverteilung** => Wsl das zeitlicher Abstand zw. 2 Störungen zwei Stunden ist. (Stunde, Minute etc. kontinuierlich!

Poissionverteilung => Wsl das nach zwei Wochen keine Störung ist (Woche = Diskret)

Binomialverteilung => 10x Würfeln, Wsl für fünf 6er

Geometrisch => Münzwurf bis Kopf gezogen wird, damit leitet man die Exponentialverteilung her

$$P(X > nx) = \lim_{n\to\infty}(1 - \lambda/n)^{nx} = e^{-\lambda x}$$

Hypergeometrisch => Lotto

Gamma => wie Exponentialverteilung nur bis zur k-ten Störung z.B

Definition 2.17

Die Zufallsvariablen $(X_1, \dots, X_n)$ heißen unabhängig, wenn für alle $x_1, \dots x_n$

$$F_{X_1,\dots,X_n}(x_1, \dots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i).$$

Die unendliche Folge $(X_n, n \in \mathbb{N})$ heißt unabhängig, wenn jede endliche Teilfolge unabhängig ist.

Bei stetigen Dichten brauchen wir einen anderen Ansatz, da dort die Punktwahrscheinlichkeiten 0 sind (der Nenner ist damit dann auch 0)

Definition 2.18

X, Y seien stetig verteilt mit Dichte $f_{X,Y}$. Die bedingte Dichte von X unter $Y = y$ ist

$$f_X(x|Y = y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

Transformationen von Verteilungen

Satz 2.9

Wenn X eine diskrete Verteilung mit Wahrscheinlichkeitsfunktion p_X hat und $Y = g(X)$ gilt, dann ist

$$p_Y(y) = \mathbb{P}(Y = y) = \mathbb{P}(g(X) = y) = \sum_{x:g(x)=y} p_X(x).$$

z.B: Y = (X-2)^2 wobei X werte von 0-6 annehmen kann, jetzt kann man all diese Werte durchprobieren, dadurch wird klar das Y andere Werte als X annehmen wird.

Satz 2.10: Transformationssatz für Dichten

$X = (X_1, \dots, X_n)$ sei stetig verteilt mit der Dichte f_X. $g : \mathbb{R}^n \to \mathbb{R}^n$ sei stetig differenzierbar und eindeutig umkehrbar. $Y = g(X)$ (d.h. $Y_i = g_i(X_1, \dots, X_n)$) ist dann ebenfalls stetig verteilt mit der Dichte

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))|\frac{\partial g^{-1}}{\partial y}(y)| = f_X(g^{-1}(y))\frac{1}{|\frac{\partial g}{\partial x}(g^{-1}(y))|} & \text{wenn } y \in g(\mathbb{R}^n), \\ 0 & \text{sonst.} \end{cases}$$

Dabei ist

$$\frac{\partial g}{\partial x} = \det((\frac{\partial g_i}{\partial x_j})_{n \times n})$$

die Funktionaldeterminante.

Ich habe die Herleitung selbst nicht genau verstanden, im Skriptum werden da paar magische Umformungen gemacht, paar Schritte machen Sinn paar nicht.

Satz 2.11

X und Y seien unabhängig mit Dichte f_X und f_Y. Dann ist die Dichte von $X + Y$

$$f_{X+Y}(z) = f_X * f_Y(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx.$$

Satz 2.12

X und Y seien unabhängig mit Dichte f_X und f_Y. Dann hat $Z = X/Y$ die Dichte

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(y) f_X(zy)|y| dy.$$

Satz 2.13

X und Y seien unabhängig mit Wahrscheinlichkeitsfunktion p_X und p_Y. Dann ist die Wahrscheinlichkeitsfunktion von $X + Y$ die (diskrete) Faltung von p_X und p_Y:

$$p_{X+Y}(z) = p_X * p_Y(z) = \sum_x p_X(x) p_Y(z-x).$$

Definition 2.20: Quantil

x heißt p-Quantil $(0 > p > 1)$ der Verteilung mit der Verteilungsfunktion F, wenn

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

Definition 2.21

Die verallgemeinerte Inverse der Verteilungsfunktion F ist gegeben durch

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

Satz 2.14

Wenn U auf $[0, 1]$ gleichverteilt ist, dann hat

$$X = F^{-1}(U)$$

die Verteilungsfunktion F.

Erwartungswert und Varianz (Varianz = mittlere Quadratische Abweichung)

Definition 2.22

Der Erwartungswert einer Zufallsvariable X ist

$$\mathbb{E}(X) = \sum_x x p_X(x)$$

für diskrete Zufallsvariable, und

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

für stetige Zufallsvariable. Falls X gemischt verteilt ist, gilt

$$\mathbb{E}(X) = \sum_x x p_X(x) + \int_{-\infty}^{\infty} x f_X(x) dx.$$

Satz 2.15: Eigenschaften des Erwartungswerts

1. Linearität: $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$,
2. Additivität: $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$,
3. Monotonie: wenn $X \leq Y$, dann ist auch $\mathbb{E}(X) \leq \mathbb{E}(Y)$,
4. wenn X und Y unabhängig sind, dann gilt $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Satz 2.16: Satz vom unachtsamen Statistiker

X sei eine Zufallsvariable, $g : \mathbb{R} \to \mathbb{R}$ eine Funktion und$Y = g(X)$. Dann ist

1. Wenn X diskret verteilt ist mit Wahrscheinlichkeitsfunktion p_X,

$$\mathbb{E}(Y) = \sum_x g(x) p_X(x).$$

2. Wenn X stetig verteilt ist mit Dichte f_X,

$$\mathbb{E}(Y) = \int g(x) f_X(x) dx.$$

Diese Formel gilt sinngemäß auch für mehrdimensionales $X \in \mathbb{R}^k$ und $g : \mathbb{R}^k \to \mathbb{R}$:

$$\mathbb{E}(Y) = \int g(x) f_X(x) dx = \int \dots \int g(x_1, \dots, x_k) f_{X_1,\dots,X_k}(x_1, \dots, x_k) dx_1 \dots dx_k.$$

3. Wenn X gemischt verteilt ist mit Wahrscheinlichkeitsfunktion p_X und Dichte f_X,

$$\mathbb{E}(Y) = \sum_x g(x) p_X(x) + \int g(x) f_X(x) dx.$$

Vorteil: Wir brauchen die Verteilung von **Y** nicht um ihren **Erwartungswert** auszurechnen!

Definition 2.23

Die Varianz von X ist
$$\mathbb{V}(X) = \mathbb{E}((X - \mathbb{E}(X))^2).$$

Satz 2.17: Steinerscher Verschiebungssatz

Für beliebiges reelles a gilt
$$\mathbb{E}((X - a)^2) = \mathbb{V}(X) + (\mathbb{E}(X) - a)^2.$$

Satz 2.18: Eigenschaften der Varianz

1. $\mathbb{V}(X) \geq 0$,
2. $\mathbb{V}(X) = 0$ genau dann, wenn $\mathbb{P}(X = \mathbb{E}(X)) = 1$,
3. $\mathbb{V}(aX + b) = a^2\mathbb{V}(X)$,
4. wenn X und Y unabhängig sind, dann ist $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$.

1. Muss logisch sein, ein Datensatz kann keine negative Abweichung von Mittelwert haben.
2. Es existiert keine Abweichung, wenn alle Elemente in einem Punkt sind mit P = 1
3. Das „+b“ ist für eine Abweichung ja irrelevant, die Abweichung bleibt ja gleich. Jedoch ist die Skalierung wichtig, da sich jeder Abstand um diese Skalierung auch verändert. (Daraus folgt die Quadrierung aus der Formel der Varianz)

Weiters gilt: V(X) = E(X^2) – E(X^2)

Kleinere Varianz, kleinerer IQR (Interquantilenabstand = x0.75 – x0.25)

Definition 2.24

Wenn $\mathbb{E}(X) = 0$, dann heißt X zentriert.
Wenn $\mathbb{E}(X^2) = 1$, dann heißt X normiert.
Wenn $\mathbb{E}(X) = 0$ und $\mathbb{V}(X) = 1$, dann heißt X standardisiert.
Wenn $\mu = \mathbb{E}(X)$ endlich ist, dann ist $X^o = X - \mu$ zentriert (die Zentrierung von X).
Wenn $\mathbb{E}(X^2)$ endlich und positiv ist, dann ist $X^* = \frac{X}{\sqrt{\mathbb{E}(X^2)}}$ normiert (die Normierung von X).
Wenn $\sigma^2 = \mathbb{V}(X)$ endlich und positiv ist, dann ist $X^{o*} = \frac{X-\mu}{\sigma}$ standardisiert (die Standardisierung von X).

Definition 2.25

X und Y seien Zufallsvariable mit endlicher Varianz. Dann heißt
$$\mathbf{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$
die Kovarianz von X und Y.

Wahrscheinlichkeit und Stochastik WS2021
Beschreibt den linearen Zusammenhang zwischen zwei Variablen.

Große Cov() => X steigt gleich mit Y

Kleine Cov() => Wenn X fällt steigt Y (inv. Prop.)

Definition 2.26

X und Y seien zwei Zufallsvariable mit positiver endlicher Varianz. Dann heißt

$$\rho(X,Y) = \frac{\mathbf{Cov}(X,Y)}{\sqrt{\mathbb{V}(X)\mathbb{V}(Y)}}$$

der Korrelationskoeffizient von X und Y.

Satz 2.19

Es gilt

$$-1 \leq \rho(X,Y) \leq 1.$$

$\rho(X,Y) = 1$ genau dann, wenn es Konstante $a > 0$ und b gibt mit $Y = aX + b$.
$\rho(X,Y) = -1$ genau dann, wenn es Konstante $a < 0$ und b gibt mit $Y = aX + b$.
$\rho(X,Y) = 0$ genau dann, wenn X und Y unkorreliert sind.

Satz 2.20: Cauchy-Schwarz Ungleichung

Wenn $\mathbb{E}(X^2)$ und $\mathbb{E}(Y^2)$ endlich sind, dann gilt

$$\mathbb{E}(XY) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}.$$

Gleichheit gilt genau dann, wenn es ein $a > 0$ gibt, sodass $Y = aX$.

Satz 2.21: Ungleichung von Markov

X sei eine nichtnegative Zufallsvariable, $\lambda > 0$. Dann ist

$$\mathbb{P}(X \geq \lambda) \leq \frac{1}{\lambda}\mathbb{E}(X).$$

für X = Augenzahl bei Würfelwurf, sieht man das es z.B. für jedes Augenzahl funkt.

Intuition: Alle Wsl. die kleiner als E(X) sind, müssen das erfüllen weil der Quotient größer 1 ist und eine WSL nicht größer 1 sein kann.

Satz 2.22: Ungleichung von Chebychev

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda) \leq \frac{\mathbb{V}(X)}{\lambda^2}.$$

Folgt durch die Ungleichung von Markov. (Betrag quadrieren, dann Markov anwenden)

Satz 2.23: Ungleichung von Kolmogorov

$X_1, \ldots, X_n$ seien unabhängig mit Erwartungswert 0, $S_0 = 0$, $S_n = X_1 + \ldots + X_n$. Dann ist

$$\mathbb{P}(\max_{k \leq n} |S_k| \geq \lambda) \leq \frac{\mathbb{V}(S_n)}{\lambda^2}.$$

Mit dieser Ungleichung kann **global** für alle Zufallsvariablen Sn eine Abschätzung gemacht werden.

Die **Ungleichungen** sind deswegen **nützlich**, um Wahrscheinlichkeiten **abschätzen** zu können.

Definition 2.27

$$M_n(X) = \mathbb{E}(X^n)$$

heißt das n-te Moment von X,

$$m_n(X) = \mathbb{E}((X - \mathbb{E}(X))^n)$$

das n-te zentrale Moment von X.

Definition 2.28

X sei eine diskrete Zufallsvariable mit nichtnegativen ganzzahligen Werten. Dann heißt die Funktion

$$p_X^*(z) = \mathbb{E}(z^X) = \sum_{n=0}^{\infty} z^n p_X(n)$$

(für reelles oder auch komplexes z) die *wahrscheinlichkeitserzeugende Funktion* von X.

s

Definition 2.29

Die Funktion

$$M_X(t) = \mathbb{E}(e^{Xt})$$

heißt die *momentenerzeugende Funktion* von X.

Definition 2.30

Die charakteristische Funktion ϕ_X der Zufallsvariable X ist gegeben durch

$$\phi_X(t) = \mathbb{E}(e^{iXt}) = \mathbb{E}(\cos(Xt)) + i\mathbb{E}(\sin(Xt)), t \in \mathbb{R}.$$

(Charakteristische Funktion als Abhilfe, wenn die momenterzeugende Funktion nicht existiert)

Satz 2.24

F[r die momentenerzeugende Funktion und die charakteristische Funktion gilt:

1. Wenn X und Y unabhängig sind, dann gilt
$$M_{X+Y}(t) = M_X(t)M_Y(t),$$
$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t).$$

2. Für $Y = aX + b$, $a, b \in \mathbb{R}$:
$$M_Y(t) = e^{bt}M_X(at),$$
$$\phi_Y(t) = e^{ibt}\phi_X(at).$$

3. Wenn $m_k = \mathbb{E}(X^k)$ endlich ist, dann ist ϕ_X k-mal differenzierbar, und
$$\phi_X^{(k)}(0) = i^k m_k.$$
Wenn wir zusätzlich annehmen, dass $M_X(t)$ für ein $t \neq 0$ endlich ist, dann gilt auch
$$M_X^{(k)}(0) = m_k.$$
(unter Umständen, wenn die Momentenerzeugende nur auf einer Seite der Null endlich ist, ist hier die einseitige Ableitung zu verwenden. Wenn das nicht der Fall ist, wenn es also negative und positive Werte von t gibt, für die $M_X(t)$ endlich ist, dann sind auch alle Momente m_k endlich).

Wahrscheinlichkeit und Stochastik WS2021

Gesetze der Großen Zahlen

Satz 2.25: schwaches Gesetz der großen Zahlen

$(X_n, n \in \mathbb{N})$ sei eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit Erwartungswert μ und endlicher Varianz, $S_n = X_1 + \ldots + X_n$. Dann gilt

$$\mathbb{P}(|\frac{S_n}{n} - \mu| \geq \epsilon) \to 0$$

für jedes $\epsilon > 0$.

Beweis mithilfe von Chebychev-Ungleichung => V(X) / (ne)^2

Definition 2.31

Wir betrachten Zufallsvariable $X_n, n \in \mathbb{N}$ und X. Die Folge X_n konvergiert gegen X

1. in Wahrscheinlichkeit, wenn für alle $\epsilon > 0$
$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0.$$
2. mit Wahrscheinlichkeit 1 (oder "fast sicher"), wenn
$$\mathbb{P}(X_n \text{ konvergiert gegen } X) = 1.$$
3. in Verteilung, wenn für alle Stetigkeitspunkte x von F_X
$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x).$$

2) ist die stärkste

3) die schwächste

Implikationskette: 2) => 1) => 3)

Satz 2.26: Starkes Gesetz der großen Zahlen

$(X_n, n \in \mathbb{N})$ sei eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit endlichem Erwartungswert $\mu = \mathbb{E}(X_n)$, $S_n = X_1 + \ldots + X_n$. Dann konvergiert $\frac{S_n}{n}$ mit Wahrscheinlichkeit 1 gegen μ.

Gegeben sei eine Folge $(X_n)_{n\in\mathbb{N}}$ von Zufallsvariablen. Diese heißt unabhängig und identisch verteilt, wenn die folgenden beiden Kriterien erfüllt sind:

- Die Familie der Zufallsvariablen X_n der Folge sind stochastisch unabhängige Zufallsvariablen.
- Die Zufallsvariablen besitzen alle dieselbe Verteilung. Das bedeutet, es existiert eine Wahrscheinlichkeitsverteilung P, so dass $X_n \sim P$ für alle $n \in \mathbb{N}$.

$$\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n}(X_i - \operatorname{E}(X_i))$$

fast sicher gegen 0 konvergiert. Das bedeutet, dass

$$P(\lim_{n\to\infty} \overline{X}_n = 0) = 1$$

Satz 2.27: Zentraler Grenzwertsatz

$(X_n, n \in \mathbb{N})$ sei eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit $\mathbb{E}(X) = \mu$, $\mathbb{V}(X) = \sigma^2$. Dann ist S_n näherungsweise normalverteilt mit Mittel $n\mu$ und Varianz $n\sigma^2$, d.h., es gilt

$$\lim_{n\to\infty} \mathbb{P}(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du.$$

Satz 2.28: deMoivre-Laplace

X_n sei binomialverteilt: $X_n \sim B(n,p)$. Dann gilt

$$\mathbb{P}(X_n = k) = \frac{1}{\sqrt{2n\pi p(1-p)}} e^{-\frac{(k-np)^2}{2np(1-p)}} + o(\frac{1}{\sqrt{n}})$$

gleichmäßig in k für $n \to \infty$.

Stärker als **Zentraler Grenzwertsatz** da Konvergenz von Dichtefunktion gezeigt ist

1. Was ist ein Wahrscheinlichkeitsraum?

Ein **Wahrscheinlichkeitsraum** (Ω, P) besteht aus einer **Menge Ω**, der Grundmenge und einem **Wahrscheinlichkeitsmaß P**. Die Elemente ω ∈ Ω heißen Elementarereignisse, Teilmengen A ⊆ Ω Ereignisse. Das **Wahrscheinlichkeitsmaß P ist auf der Menge dieser Ereignisse definiert** (und erfüllt die Axiome von Kolmogorov).

2. Was ist ein Ereignis?

Die einzelnen Elemente von **Ω** nennt man Elementarereignisse, jede Teilmenge von **Ω** ist ein Ereigniss.

3. Was besagen die Axiome von Kolmogorov?

Grundlegende Regeln die für das **Wahrscheinlichkeitsmaß P.**

1. $0 \leq \mathbb{P}(A) \leq 1$.
2. $\mathbb{P}(\emptyset) = 0$.
3. $\mathbb{P}(\Omega) = 1$.
4. (Additivität) Wenn A und B disjunkte Ereignisse ($A \cap B = \emptyset$) sind, dann gilt

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

4a. (Abzählbare Additivität, Sigmaadditivität) Wenn $A_n, n \in \mathbb{N}$ disjunkte Ereignisse sind (d.h., $A_i \cap A_j = \emptyset, i \neq j$), dann gilt

$$\mathbb{P}(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n).$$

4. Welche Eigenschaften von Wahrscheinlichkeitsmaßen kennen Sie?

Siehe 3)

5. Wie kann man die Wahrscheinlichkeit einer Vereinigung berechnen?

Falls A und B **disjunkt**: $P(A \cup B) = P(A) + P(B)$

Falls **nicht:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

6. Wie ist die bedingte Wahrscheinlichkeit definiert?

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

7. Wie kann man die Wahrscheinlichkeit eines Durchschnitts berechnen?

Falls **unabhängig:** $P(A \cap B) = P(B) * P(A)$

Falls **nicht**: $P(A \cap B) = P(B) * P(A|B)$

8. Wann heißen zwei Ereignisse unabhängig?

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

9. Wann heißen mehr als zwei Ereignisse unabhängig?

Wenn **gilt** $P(A \cap B \cap C \cap ..) = P(A) * P(B) * P(C)$ …

10. Wann heißen mehr als zwei Ereignisse paarweise unabhängig?

Für Ereignisse A, B, C

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

11. Was besagt der Satz der vollständigen Wahrscheinlichkeit?

$$P(A) = \sum_i P(B_i) * P(A|B_i) = \sum_i P(A \cap B_i)$$

Sozusagen alle Blätter wo „A“ vorkommt summieren.

12. Was besagt der Satz von Bayes?

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(B_j)\mathbb{P}(A|B_j)}{\sum_i \mathbb{P}(B_i)\mathbb{P}(A|B_i)}.$$

Der Nenner entspricht P(A) lt. dem **Satz der vollständigen Wahrscheinlichkeit**, das heißt die Wsl. von Bj unter Bedingung von A ist, P(Bj) * P(A|Bj) / P(A). (Finde aber die Formel schöner wenn oben stehen würde P(A)*P(Bj|A), sollte aber äquivalent sein)

13. Was ist eine Zufallsvariable?

Ist eine **Abbildung** von Ω nach R^d (d = Dimensionen).

Beispiel: Zufallsvariable **S**, welche die Augensumme zweier gewürfelter Würfel definiert.

$$S: \Omega \to \mathbb{R}; \quad (n_1, n_2) \mapsto n_1 + n_2,$$

14. Welche Typen von Zufallsvariablen gibt es?

Stetige und **diskrete** Zufallsvariablen.

15. Wie kann man die Verteilung einer Zufallsvariable angeben? (Verstehe die frage tbh nd)

$$F_X(x) = \mathbb{P}_X((-\infty, x]) = \mathbb{P}(X \leq x)$$

16. Welche Eigenschaften hat eine Verteilungsfunktion?

0 <= F(X) <=1

F ist **monoton** nicht fallend

F ist **rechtsstetig**

Limes gegen -**unendlich** => 0

Limes gegen +**unendlich** => 1

17. Welche Eigenschaften hat eine Wahrscheinlichkeitsfunktion?

Summe über die Wsl. Funktion muss 1 ergeben.

18. Welche Eigenschaften hat eine Dichtefunktion?

Integral über die Dichte muss 1 sein

19. Wann heißen zwei/mehrere Zufallsvariable unabhängig?

$$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n} F_{X_i}(x_i).$$

20. Was versteht man unter einer "Randverteilung"?

Wenn **eine** Verteilung aus einer **gemeinsamen** Verteilung gewonnen wird. (der diskrete Fall)

z.B. einfach alle **Zeilen summieren** bei einer Tabelle.

21. Wie kann man die Dichte einer transformierten Zufallsvariable bestimmen?

Mithilfe des Transformationssatz für Dichten.

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))|\frac{\partial g^{-1}}{\partial y}(y)| = f_X(g^{-1}(y))\frac{1}{|\frac{\partial g}{\partial x}(g^{-1}(y))|} & \text{wenn } y \in g(\mathbb{R}^n), \\ 0 & \text{sonst.} \end{cases}$$

Dabei ist

$$\frac{\partial g}{\partial x} = \det((\frac{\partial g_i}{\partial x_j})_{n\times n})$$

die Funktionaldeterminante.

22. Wie berechnet man die bedingte Dichte?

Das Problem tritt auf, da bei der bedingten Wsl. der Nenner 0 sein würde für Punktwsl.

$$f_X(x|Y=y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

23. Wie ist der Erwartungswert einer Zufallsvariable definiert?

Definition 2.22

Der Erwartungswert einer Zufallsvariable X ist

$$\mathbb{E}(X) = \sum_x x p_X(x)$$

für diskrete Zufallsvariable, und

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

für stetige Zufallsvariable. Falls X gemischt verteilt ist, gilt

$$\mathbb{E}(X) = \sum_x x p_X(x) + \int_{-\infty}^{\infty} x f_X(x) dx.$$

24. Wie ist die Varianz einer Zufallsvariable definiert?

Durch die **Mittlere Quadratische Abweichung.** V(X) = E((X-E(X))^2)

25. Wie ist die Kovarianz von zwei Zufallsvariablen definiert?

$\mathbf{Cov}(X,Y) = \mathbb{E}((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ Zeigt an wie sehr 2 Zufallsvariablen korrelieren.

26. Welche Eigenschaften des Erwartungswerts kennen Sie?

Linearität, **Additivität**, **Monotonie** (X<Y ⇔E(X)<E(Y),

Falls X und Y **unabhängig**, E(XY) = E(X)E(Y)

27. Welche Eigenschaften der Varianz kennen Sie?

V(X) >= 0

V(X) = 0 ⇔P(X=E(X)) = 1

V(aX+b) = a^2*V(X)

Falls **unabhängig:** V(X+Y) = V(X) + V(Y) (+ 2*Cov(X,Y)) falls nicht unabhängig

28. Was besagt die Ungleichung von Markov?

$$\mathbb{P}(X \geq \lambda) \leq \frac{1}{\lambda}\mathbb{E}(X).$$

Intuition: Alle Wsl. die kleiner als E(X) sind, müssen das erfüllen weil der Quotient größer 1 ist und eine WSL nicht größer 1 sein kann.

Schätzt die Wahrscheinlichkeit nach oben ab.

Mit **Markov** kann **Chebychev** und **Kolmogorov** gezeigt werden.

29. Was besagt die Ungleichung von Chebychev?

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda) \leq \frac{\mathbb{V}(X)}{\lambda^2}.$$

Beweis: In P beide Seiten quadrieren. => Markov anwenden

30. Was besagt die Ungleichung von Kolmogorov?

$$\mathbb{P}(\max_{k \leq n} |S_k| \geq \lambda) \leq \frac{\mathbb{V}(S_n)}{\lambda^2}.$$

Ebenfalls eine **Abschätzung** nach **oben** für jede Zufallsvariable, die man sich durch die Summe aller Zufallsvariablen berechnen kann (wobei **E(X_i) = 0 gelten muss**)

31. Wie sind die Momente einer Zufallsvariable definiert?

M_n = E(X^n)

32. Was ist die momentenerzeugende Funktion einer Zufallsvariable?

M_x(t) = E(e^Xt)

Name kommt durch die **Potenzreihenentwicklung** der Exp. Funktion

33. Was ist die wahrscheinlichkeitserzeugende Funktion einer diskreten Zufallsvariable?

$$p_X^*(z) = \mathbb{E}(z^X) = \sum_{n=0}^{\infty} z^n p_X(n)$$

34. Was ist die charakteristische Funktion einer Zufallsvariable?

Die Momenterzeugende nur wird **t=i*t** gesetzt und dann **Eulersche Identität** angewandt

35. Was besagt das schwache Gesetz der großen Zahlen?

$$\mathbb{P}(|\frac{S_n}{n} - \mu| \geq \epsilon) \to 0$$

Wobei Sn Summe von ident. Verteilten und unabhängigen Zufallsvariablen ist.

Diese **Summe zentriert** wird für jedes **beliebig kleine epsilon** eine Wsl. von 0 geben.

Das heißt sie konvergiert **in Wahrscheinlichkeit!**

Beweis: mit Chebychev => lim n=>unendlich V(Sn) / e^2*n^2 = n* V(X) /e^2*n^2 = V(X) /e^2*n

36. Was besagt das starke Gesetz der großen Zahlen?

Ähnlich wie das schwache Gesetz. **Sn / n konvergiert** gegen **µ** mit **Wsl. 1**

37. Was besagt der zentrale Grenzwertsatz?

Satz 2.27: Zentraler Grenzwertsatz

$(X_n, n \in \mathbb{N})$ sei eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit $\mathbb{E}(X) = \mu$, $\mathbb{V}(X) = \sigma^2$. Dann ist S_n näherungsweise normalverteilt mit Mittel $n\mu$ und Varianz $n\sigma^2$, d.h., es gilt

$$\lim_{n\to\infty} \mathbb{P}(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du.$$

Heißt wir können viele Verteilungen mit großem N mit der Normalverteilung approximieren.

Schätzertheorie

Definition 5.1

Ein statistisches Modell ist eine Menge $\mathcal{P}$ von Verteilungen. Wenn diese Verteilungen durch endlich viele reelle Zahlen (die Parameter) beschrieben werden können, also

$$\mathcal{P} = \{P_\theta, \theta \in \Theta\}$$

mit $\Theta \subseteq \mathbb{R}^d$, dann sprechen wir von einem parametrischen Modell, sonst von einem nichtparametrischen Modell.

Eine Stichprobe ist eine Folge $(X_1, \ldots, X_n)$ von unabhängigen Zufallsvariablen mit einer (unbekannten) Verteilung aus $\mathcal{P}$.

Definition 5.2

Eine Statistik T ist eine Zufallsvariable, die aus der Stichprobe berechnet werden kann:

$$T = T(X_1, \ldots, X_n)$$

Insbesondere dürfen in dieser Funktion die unbekannten Parameter nicht vorkommen.

Beispiele für eine Statistik: **Stichprobenmittelwert**, **Formel für die Varianz..**

Definition 5.3

Ein Schätzer ist eine Folge $(\hat{\theta}_n)$ von Statistiken $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$.

Definition 5.4

Ein Schätzer $\hat{\theta}_n$ heißt

- schwach konsistent, wenn $\hat{\theta}_n \to \theta$ in Wahrscheinlichkeit konvergiert (also $\mathbb{P}_\theta(|\hat{\theta}_n - \theta| > \epsilon) \to 0$ für alle $\epsilon > 0$).
- stark konsistent, wenn $\hat{\theta}_n \to \theta$ mit Wahrscheinlichkeit 1 konvergiert (also $\mathbb{P}_\theta(\hat{\theta}_n \to \theta) = 1$.
- erwartungstreu, wenn für alle $\theta \in \Theta$

$$\mathbb{E}_\theta(\hat{\theta}_n) = \theta.$$

- effizient, wenn er erwartungstreu ist und die kleinste Varianz unter allen erwartungstreuen Schätzern hat, also wenn für jeden weiteren erwartungstreuen Schätzer $\tilde{\theta}_n$

$$\mathbb{V}_\theta(\hat{\theta}_n) \leq \mathbb{V}_\theta(\tilde{\theta}_n) \text{ für alle } \theta. \tag{5.1}$$

Beispiele: Stichprobenmittelwert, Stichprobenvarianz (nicht erwartungstreu, deswegen durch **n-1** rechnen und nicht durch **n)**

Gutes Beispiel wie gut man sich mit den Regeln von Erwartungswerten und Varianzen auskennt: Stichprobenvarianz Beweis komplett verstehen!

Definition 5.5: Momentenmethode

θ sei ein einzelner reeller Parameter (also der Parameterraum Θ eindimensional). Dann können wir den Erwartungswert von X als Funktion von θ schreiben:

$$\mathbb{E}_\theta(X) = m(\theta).$$

Wenn die Funktion m stetig umkehrbar ist, dann ist

$$\hat{\theta}_n = m^{-1}(\bar{X}_n)$$

ein (stark) konsistenter Schätzer, der Momentenschätzer.

Wenn es $d > 1$ Parameter gibt, dann verwendet man zusätzlich höhere Momente:

$$m_i(\theta) = m_i(\theta_1, \ldots, \theta_d) = \mathbb{E}_\theta(X^i).$$

Wir ersetzen die theoretischen Momente durch ihre Schätzer, das ergibt die d Gleichungen

$$m_i(\theta_1, \ldots, \theta_d) = \frac{1}{n}\sum_{j=1}^{n} X_j^i, \; i = 1, \ldots, d$$

in den d Variablen $\theta_1, \ldots, \theta_d$. Auflösen dieses Gleichungssytems nach θ ergibt den Momentenschätzer für $\theta = (\theta_1, \ldots, \theta_d)$. Wenn das Ergebnis stetig von den rechten Seiten des Gleichungssystems abhängt, erhalten wir dadurch wieder einen konsistenten Schätzer.

$$\mathbb{E}_\theta(X) = \frac{\theta}{\theta+1} \Rightarrow \bar{X}_n = \frac{\hat{\theta}_n}{\hat{\theta}_n+1}$$ $\Rightarrow$ umformen nach $\hat{\theta}_n$

$$\bar{X}_n \cdot \hat{\theta}_n + \bar{X}_n = \hat{\theta}_n$$

$$\bar{X}_n = \hat{\theta}_n - \bar{X}_n \cdot \hat{\theta}_n$$

$$\bar{X}_n = \hat{\theta}_n (1-\bar{X}_n) \Rightarrow \hat{\theta}_n = \frac{\bar{X}_n}{1-\bar{X}_n}$$ Momentschätzer

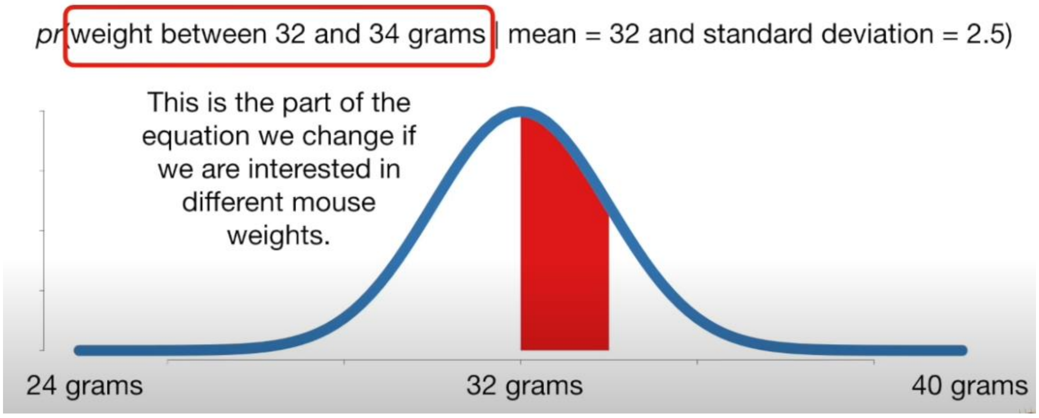
Beispiel für einen Momentenschätzer.

Wahrscheinlichkeit und Stochastik WS2021

Für den **Maximum Likelihood Schätzer**, muss erstmal klar gemacht werden was der Unterschied ist zwischen **Probability** und **Likelihood.**

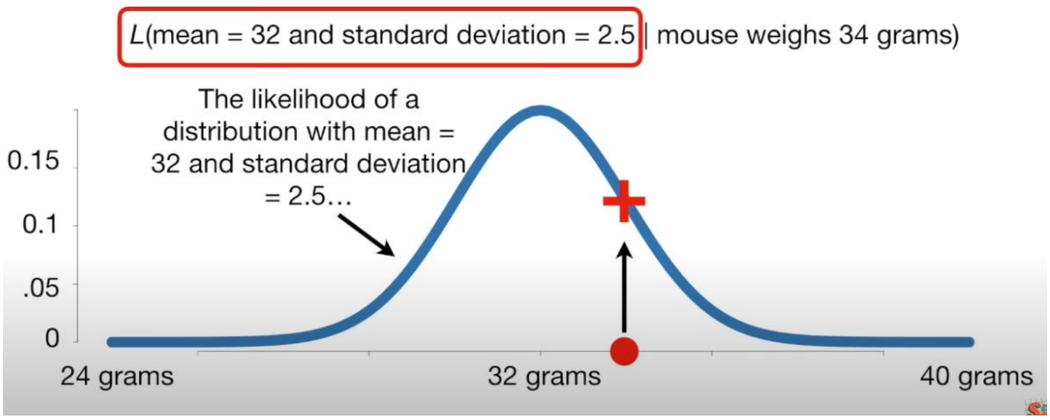
Bsp**: Normalverteilte Gewichte von Menschen**

Probability:



Der Rote part ändert sich, mean und deviation bleibt gleich.

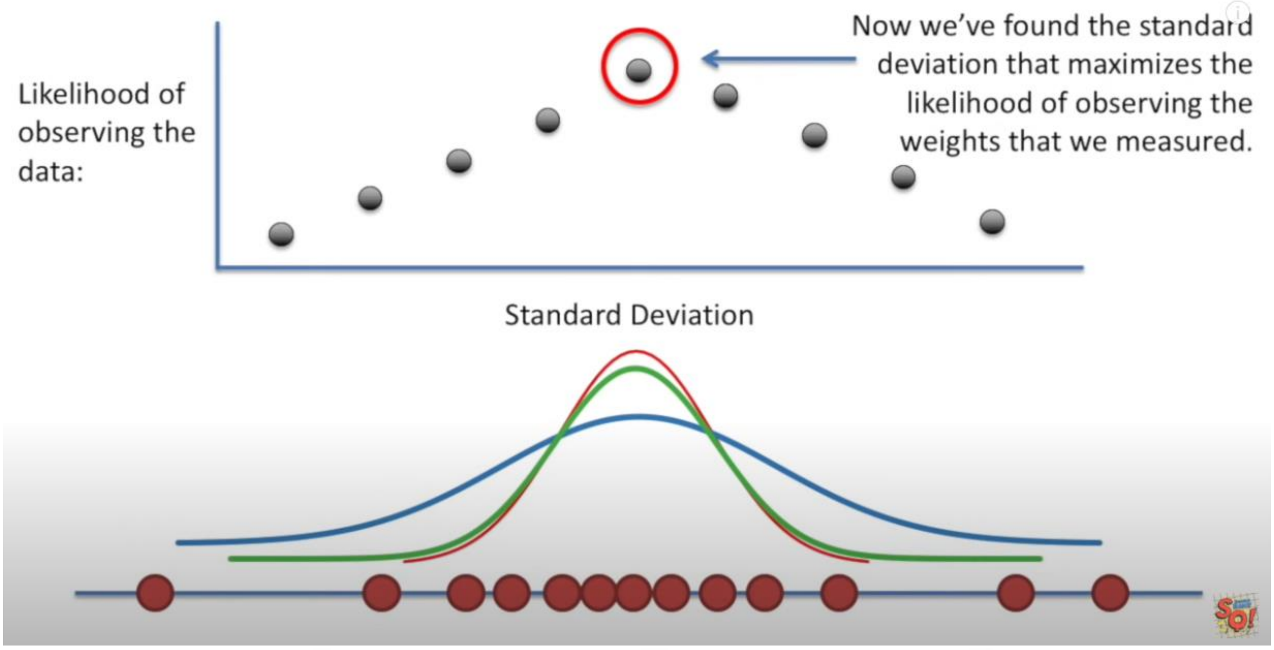
Likelihood:

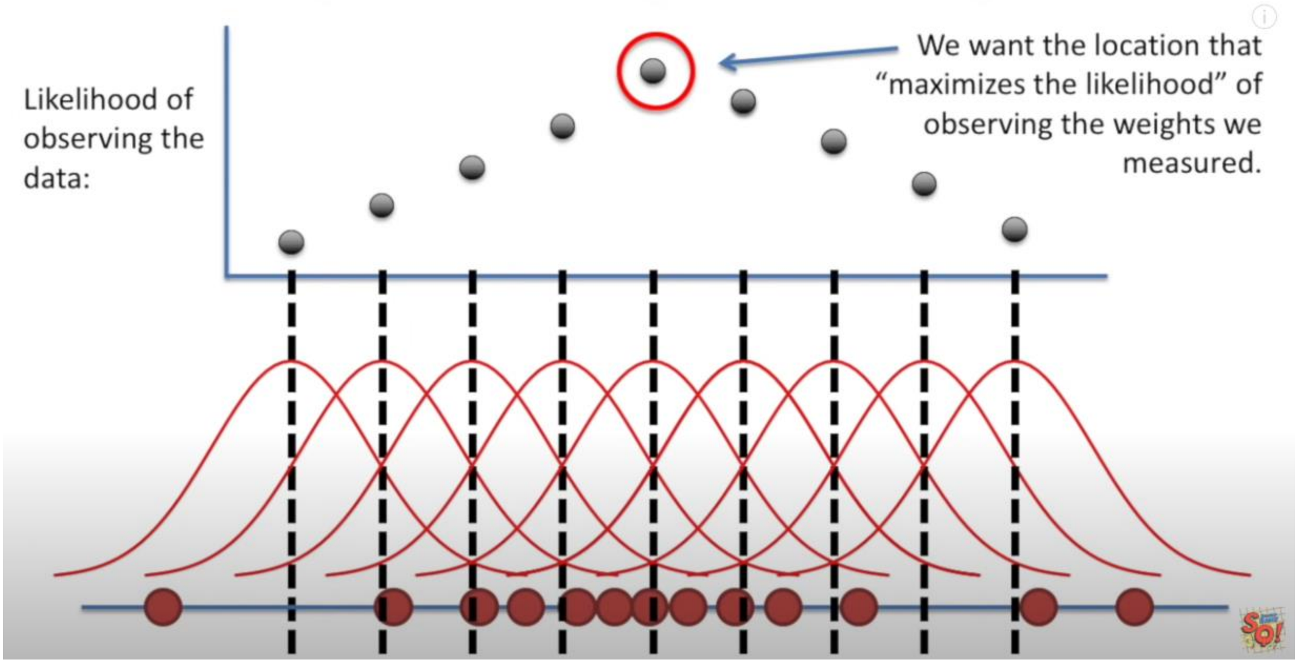


Man maximiert die **Punktwahrscheinlichkeiten** für eine Stichprobe.

(https://www.youtube.com/watch?v=pYxNSUDSFH4 https://www.youtube.com/watch?v=XepXtl9YKwc)

Man muss es sich so vorstellen, dass die Verteilung so verändert wird, so dass sie der **Stichprobe** anpasst.





Definition 5.6: Likelihoodfunktion

Die Likelihoodfunktion ist

$$L(X_1, \ldots, X_n; \theta) = \prod_{i=1}^{n} p_\theta(X_i),$$

wenn P_θ diskret mit Wahrscheinlichkeitsfunktion p ist, und

$$L(X_1, \ldots, X_n; \theta) = \prod_{i=1}^{n} f_\theta(X_i),$$

wenn P_θ stetig mit Dichte f ist.

Definition 5.7: Maximum-Likelihood-Methode

Der Maximum-Likelihood (ML-) Schätzer ist der Wert von θ, der die Likelihoodfunktion maximiert.

Das heißt in der **Praxis:**

- Likelihoodfunktion logarithmieren (machts später einfacher)
- Die Funktion nach allen Parametern (z.B µ und σ) **partiell ableiten**
- gleich **0** setzen (aka. Maximieren)
- Gleichungen lösen

Satz 5.1: Cramér-Rao

Wenn p_θ bzw. f_θ zweimal nach θ differenzierbar ist und zusätzliche Regularitätsvoraussetzungen erfüllt, dann gilt für jeden Erwartungstreuen Schätzer $\hat{\theta}_n$

$$\mathbb{V}(\hat{\theta}_n) \geq \frac{1}{I_n(\theta)} = \frac{1}{nI(\theta)}.$$

Dabei ist

$$I_n(\theta) = \mathbb{E}_\theta((\frac{\partial \log(L(X_1, \ldots, X_n; \theta))}{\partial \theta})^2) = -\mathbb{E}_\theta(\frac{\partial^2}{\partial \theta^2} \log(L(X_1, \ldots, X_n; \theta)))$$

und $I(\theta) = I_1(\theta)$, also

$$I(\theta) = -\mathbb{E}_\theta((\frac{\partial \log(f_\theta(X))}{\partial \theta})^2) = -\mathbb{E}_\theta(\frac{\partial^2}{\partial \theta^2} \log(f_\theta(X)))$$

bzw.

$$I(\theta) = -\mathbb{E}_\theta((\frac{\partial \log(p_\theta(X))}{\partial \theta})^2) = -\mathbb{E}_\theta(\frac{\partial^2}{\partial \theta^2} \log(p_\theta(X))).$$

$I(\theta)$ bzw. $I_n(\theta)$ wird als die Fisherinformation bezeichnet.

Die Fisher-Information gibt an wie gut eine **Observation** (Stichprobe z.B.) der Zufallsvariable den wahren Wert des Parameters lokalisiert. (https://www.youtube.com/watch?v=pneluWj-U-o)

Also gibt die **Cramer-Rao** Schranke eine untere Schranke für die Varianz eines **Schätzers**.

Definition 5.8

Die Statistik $T = T(X_1, \ldots, X_n)$ ist *suffizient* für den Parameter θ, wenn die bedingte Verteilung von $(X_1, \ldots, X_n)$ unter T nicht von θ abhängt.

Satz 5.2: Faktorisierungssatz von Neyman

T ist genau dann suffizient für θ, wenn die Likelihoodfunktion in der Form

$$L(X_1, \ldots, X_n; \theta) = g(T, \theta) h(X_1, \ldots, X_n)$$

dargestellt werden kann.

Beispiel 5.12

Für die Gleichverteilung $U(0, \theta)$ kann die Likelihoodfunktion als

$$L(X_1, \ldots, X_n, \theta) = \begin{cases} \frac{1}{\theta^n} & X_i \geq 0, i = 1, \ldots, n, \max(X_1, \ldots, X_n) \leq \theta, \\ 0 & \text{sonst} \end{cases}$$

geschrieben werden. Hier ist $T = \max(X_1, \ldots, X_n)$ suffizient.

Intervallschätzung

Definition 5.9

$a = a(X_1, \ldots, X_n) \leq b = b(X_1, \ldots, X_n)$ seien zwei Statistiken. Das Intervall $[a, b]$ heißt Konfidenzintervall für θ mit Überdeckungswahrscheinlichkeit γ, wenn

$$\mathbb{P}_\theta(a \leq \theta \leq b) \geq \gamma.$$

Wenn in dieser Ungleichung Gleichheit gilt, sprechen wir von einem exakten Konfidenzintervall.

Konfidenzintervall Beispiel Normalverteilung: Es wird die Verteilung der **Stichprobenmittelwerte** genommen, dort wird dann mit der Verteilung ein **symmetrisches** Intervall (z.B. 95%) aufgestellt. Das heißt mit 95% Wsl. ist dort **µ** drinnen.

Wichtig für später: Konfidenzintervalle und **Hypothesentests** sind fast dasselbe, bei dem **Konfidenzintervall** ist das Intervall um den **Mittelwert der Stichprobe**, bei den **Hypothesentests** ist das Intervall um die Annahme (z.B. µ = 200) und es wird dann verglichen, ob der **Mittelwert der Stichprobe** im **Hypothesenintervall** ist.

Deswegen sind die Formeln für die Hypothesentests und Konfidenzintervalle fürs µ fast ident.

Satz 5.4

$X - 1, \ldots, X_n$ seien unabhängig $N(\mu, \sigma^2)$-verteilt. Dann sind $\bar{X}_n$ und S_n^2 unabhängig, $\bar{X}_n \sim N(\mu, \sigma^2/n)$ und $(n-1)S_n^2/\sigma^2 \sim \chi^2_{n-1}$ (Chi-Quadrat mit $n-1$ Freiheitsgraden).

$$T = \frac{\bar{X}_n - \mu}{\sqrt{S_n^2/n}}$$

hat dann eine t-Verteilung mit $n-1$ Freiheitsgraden.

Dieser Satz besagt einfach, dass (wie z.B. die µ-Mittelwerte N(µ, σ/sqrt(n)) verteilt sind) die **normierte Varianz** Chi-Quadrat verteilt ist. Und damit das **Konfidenzintervall** für **µ t-verteilt** ist.

Satz 5.5

Konfidenzintervalle für die Normalverteilung $N(\mu, \sigma^2)$:
Für μ, wenn σ^2 bekannt ist:

$$[\bar{X}_n - z_{\frac{1+\gamma}{2}}\sqrt{\frac{\sigma^2}{n}}, , \bar{X}_n + z_{\frac{1+\gamma}{2}}\sqrt{\frac{\sigma^2}{n}}],$$

Für μ, wenn σ^2 unbekannt ist:

$$[\bar{X}_n - t_{n-1;\frac{1+\gamma}{2}}\sqrt{\frac{S_n^2}{n}}, , \bar{X}_n + t_{n-1;\frac{1+\gamma}{2}}\sqrt{\frac{S_n^2}{n}}],$$

für σ^2:

$$[\frac{(n-1)S_n^2}{\chi^2_{n-1;\frac{1+\gamma}{2}}}, \frac{(n-1)S_n^2}{\chi^2_{n-1;\frac{1-\gamma}{2}}}].$$

Mithilfe der Normalverteilung können auch **approximative Konfidenzintervalle** erstellt werden.

Beispiel Alternativverteilung: Var(X) = p*(1-p) E(X) = p

⇨ In die Formel des Normalen Konf. Intervalls einsetzen:

Satz 5.6

$(X_1, \ldots, X_n)$ sei eine Stichprobe einer Alternativverteilung $A(p), 0 < p < 1$.

$$\hat{p} = \bar{X}_n$$

sei der (erwartungstreue und effiziente) Schätzer für p. Dann ist ein approximatives Konfidenzintervall gegeben durch

$$[\hat{p} - z_{\frac{1+\gamma}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, , \hat{p} + z_{\frac{1+\gamma}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}].$$

Hypothesentests

Definition 5.10

Eine *Hypothese* ist eine Teilmenge des Parameterraums Θ.

Enthält die Hypothese nur einen Wert (also 99% der Fälle) z.B. **c** als Grenze, nennen wir ihn **einfach**.

Definition 5.11

Ein Test heißt vom Niveau α, wenn die Wahrscheinlichkeit für einen Fehler erster Art (die bei zusammengesetzten Hypothesen eine Funktion davon θ ist) nicht größer als α ist.

Fehler 1. Art: Nullhypothese wird verworfen, obwohl sie stimmt.

Fehler 2. Art: Nullhypothese wird akzeptiert, obwohl sie nicht zutrifft.

Logischerweise kann man durch ein sehr niedriges **alpha** kaum einen **Fehler 1. Art** machen, jedoch wird die Fehlerwahrscheinlichkeit eines **Fehlers 2. Art höher**.

Kochrezept für Hypothesentests:

1) **Nullhypothese** und **Alternative** formulieren
2) Teststatistik auswählen die **Nullhypothese** in **Frage stellen** könnte
3) **Signifikanzniveau** (Fehlerwsl. 1. Art) festlegen
4) Bestimmung der **Verteilung** der Statistik
5) **Stichprobe** ziehen und **Teststatistik** berechnen
6) Prüfen ob **Wert** (z.B. **Mittelwert**) im **Intervall** liegt, **wenn nicht**, **verwerfen**, sonst **akzeptieren**.

Definition 5.12: p-Wert

T sei eine Teststatistik, T_{stp} sei der Wert von T, der aus einer konkreten Stichprobe berechnet wurde. Der p-Wert ist die Wahrscheinlichkeit, einen Wert für T zu erhalten, der mindestens so stark gegen die Nullhypothese spricht wie der aus der Stichprobe. Was das konkret bedeutet, hängt von der Semantik von T ab:

- Wenn T einseitig ist, und H_0 verworfen wird, wenn $T > t_c$ ist, dann ist der p-Wert $\mathbb{P}(T \geq T_{stp})$,
- Wenn T in der anderen Richtung einseitig ist, also H_0 verworfen wird, wenn $T < t_c$ ist, dann ist der p-Wert $\mathbb{P}(T \leq T_{stp})$,
- Wenn T zweiseitig ist, dann ist der p-Wert $2\min(\mathbb{P}(T \geq T_{stp}), \mathbb{P}(T \leq T_{stp}))$.

Die Nullhypothese wird auf Niveau α verworfen, wenn der p-Wert kleiner als α ist.

Für den Mittelwert einer Normalverteilung, wenn σ^2 unbekannt ist:

$$T = \frac{\bar{X}_n - \mu_0}{\sqrt{S_n^2/n}}$$

$H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$: verwerfen, wenn $T > t_{n-1;1-\alpha}$.
$H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$: verwerfen, wenn $T < -t_{n-1;1-\alpha}$.
$H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$: verwerfen, wenn $|T| > t_{n-1;1-\alpha/2}$.
Für die Varianz einer Normalverteilung:

$$T = \frac{(n-1)S_n^2}{\sigma_0^2}$$

$H_0 : \sigma^2 \geq \sigma_0^2$ gegen $H_1 : \sigma^2 < \sigma_0^2$: verwerfen, wenn $T < \chi^2_{n-1;\alpha}$.
$H_0 : \sigma^2 = \sigma_0^2$ gegen $H_1 : \sigma^2 \neq \sigma_0^2$: verwerfen, wenn $T > \chi^2_{n-1;1-\alpha/2}$ oder $T < \chi^2_{n-1;\alpha/2}$.
$H_0 : \sigma^2 \leq \sigma_0^2$ gegen $H_1 : \sigma^2 > \sigma_0^2$: verwerfen, wenn $T > \chi^2_{n-1;1-\alpha}$.
Für Anteilswerte:

$$T = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}.$$

$H_0 : p \leq p_0$ gegen $H_1 : p > p_0$: verwerfen, wenn $T > z_{1-\alpha}$.
$H_0 : p \geq p_0$ gegen $H_1 : p < p_0$: verwerfen, wenn $T < -z_{1-\alpha}$.
$H_0 : p = p_0$ gegen $H_1 : p \neq p_0$: verwerfen, wenn $|T| > z_{1-\alpha/2}$.

Definition 5.13

Die Likelihoodquotientenstatistik ist

$$\ell = \frac{\sup_{\theta \in H_0} L(X_1, \ldots, X_n, \theta)}{\sup_{\theta \in H_1} L(X_1, \ldots, X_n, \theta)}$$

bzw.

$$\ell = \frac{\sup_{\theta \in H_0} L(X_1, \ldots, X_n, \theta)}{\sup_{\theta \in \Theta} L(X_1, \ldots, X_n, \theta)}$$

(die zweite Form ist oft einfacher zu berechnen, und für große Stichprobenumfänge sind die Tests identisch).

Der Likelihoodquotiententest verwirft die Nullhypothese, wenn ℓ kleiner ist als ein kritischer Wert.

In einem speziellen Fall ist der Likelihoodquotiententest optimal:

Satz 5.7: Neyman-Pearson

Falls sowohl H_0 als auch H_1 einfach ist,dann ist der Likelihoodquotiententest optimal, d.h., er hat unter allen Tests mit demselben Niveau die minimale Wahrscheinlichkeit für einen Fehler zweiter Art (in diesem Fall wird der Likelihoodquotient einfach als

$$L(X_1, \ldots X_n, \theta_0)/L(X_1, \ldots X_n, \theta_1)$$

berechnet)

Intuition: Es werden die **Likelihood Funktionen** genommen, und verglichen, ob der Parameter **in H0** oder der Parameter in **H1** eine größere **Likelihood** hat. (https://www.youtube.com/watch?v=Tn5y2i_MqQ8)

Chi-Quadrat Anpassungstest (https://www.youtube.com/watch?v=1Ldl5Zfcm1Y&ab_channel=MathMeeting)

Hier wird überprüft ob eine **Stichprobe** aus einer bestimmten Verteilung **stammen** kann.

Dabei ist:

- **H0:** X ~ P
- **H1:** X ≁ P

H0 wird z.B. verworfen (mit Signifikanz = 0.05), wenn: $\sum_i^x \frac{(x_i - E(X))^2}{E(X)} \geq \chi_{n-1,0.95}$

Wahrscheinlichkeit und Stochastik WS2021

1. Woraus besteht ein statistisches Modell?

Eine **Menge** von **Verteilungen**, die durch endlich viele Parameter beschrieben werden können.

$$\mathcal{P} = \{P_\theta, \theta \in \Theta\}$$

2. Was ist der Unterschied zwischen einem parametrischen und einem nichtparametrischen Modell?

Wenn der Parameterraum Θ ⊆ R^d dann ist es ein **parametrisches Modell. Sonst nicht.**

3. Geben Sie Beispiele für statistische Modelle an.

Normalverteilung: (μ,σ^2) ⊆ Θ = R^2

Alternativverteilung: (p) ⊆ Θ = [0,1]

4. Was ist eine Statistik?

Eine **Zufallsvariable**, die aus der **Stichprobe** (Folge von unabhängigen Zufallsvariablen) berechnet werden kann.

Die unbekannten Parameter dürfen hier NICHT vorkommen

5. Was ist ein Schätzer?

Ist eine **Folge** von **Statistiken** (Also Parameter die wir schätzen)

6. Welche Eigenschaften können wir von einem Schätzer verlangen?

Schwach konsistent => Parameter konvergiert in Wahrscheinlichkeit

Stark konsistent => Parameter konvergiert mit Wahrscheinlichkeit 1

Erwartungstreu => E(θn) = θ.

Effizient => Erwartungstreu und **niedrigste Varianz** aller Schätzern.

7. Wie sind Stichprobenmittel und Stichprobenvarianz definiert? Warum wird bei der Stichprobenvarianz durch n – 1 dividiert und nicht durch n?

Stichprobenmittel => normaler Mittelwert

Stichprobenvarianz => Mittlere Quadratische Abweichung aber durch n-1 dividiert

Warum durch n-1? Wenn man sich den **Beweis** anschaut ergibt es sich **rechnerisch**, da die normale Varianz Formel **nicht Erwartungstreu** ist und somit **biased**.

8. Wie ist die Likelihoodfunktion definiert?

$$L(X_1, \ldots, X_n; \theta) = \prod_{i=1}^{n} p_\theta(X_i),$$

9. Welche Methoden zur Konstruktion von Schätzern kennen Sie?

Momentenschätzer => Oft einfacher und schneller, mithilfe von Umkehrfunktionen wird gelöst

Maximum Likelihood => Oft komplizierter und zach

$$E_\theta(X) = \frac{\theta}{\theta+1} \Rightarrow \bar{X}_n = \frac{\hat{\theta}_n}{\hat{\theta}_n+1} \Rightarrow \text{umformen nach } \hat{\theta}_n$$

$$\bar{X}_n \cdot \hat{\theta}_n + \bar{X}_n = \hat{\theta}_n$$

$$\bar{X}_n = \hat{\theta}_n - \bar{X}_n \cdot \hat{\theta}_n$$

$$\bar{X}_n = \hat{\theta}_n (1-\bar{X}_n) \Rightarrow \hat{\theta}_n = \frac{\bar{X}_n}{1-\bar{X}_n} \quad \text{Momentschätzer}$$

10. Wie klein kann die Varianz eines erwartungstreuen Schätzers werden?

Die Cramer-Rao Schranke gibt hier mithilfe der Fisher Information eine untere Grenze für die minimale Varianz an.

Kehrwert der **Fisher Information der Likelihood Funktion** ist die untere Grenze.

Wahrscheinlichkeit und Stochastik WS2021

11. Wie ist ein Konfidenzintervall definiert?

Mit welcher Wahrscheinlichkeit ein Parameter in einem Intervall sich befindet.

12. Geben Sie ein Konfidenzintervall für den Mittelwert μ einer Normalverteilung an.

Konfidenzintervall des Mittelwerts kann in der **Stichprobenmittelwertverteilung** die Varianz n*σ^2 hat.

Konfidenzintervalle für die Normalverteilung $N(\mu, \sigma^2)$:

Für μ, wenn σ^2 bekannt ist:

$$[\bar{X}_n - z_{\frac{1+\gamma}{2}}\sqrt{\frac{\sigma^2}{n}}, , \bar{X}_n + z_{\frac{1+\gamma}{2}}\sqrt{\frac{\sigma^2}{n}}],$$

Für μ, wenn σ^2 unbekannt ist:

$$[\bar{X}_n - t_{n-1;\frac{1+\gamma}{2}}\sqrt{\frac{S_n^2}{n}}, , \bar{X}_n + t_{n-1;\frac{1+\gamma}{2}}\sqrt{\frac{S_n^2}{n}}],$$

für σ^2:

$$[\frac{(n-1)S_n^2}{\chi^2_{n-1;\frac{1+\gamma}{2}}}, \frac{(n-1)S_n^2}{\chi^2_{n-1;\frac{1-\gamma}{2}}}].$$

13. Geben Sie ein Konfidenzintervall für die Varianz σ einer Normalverteilung an.

14. Geben Sie ein approximatives Konfidenzintervall für die Erfolgswahrscheinlichkeit p einer Alternativverteilung an.

Das kann man sich leicht konstruieren mithilfe des Zentral Grenzwertsatzes. Varianz ist oben in der Wurzel eingesetzt, Mittelwert ebenfalls.

$$[\hat{p} - z_{\frac{1+\gamma}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, , \hat{p} + z_{\frac{1+\gamma}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}].$$

15. Was ist ein statistischer Test?

Man versucht mithilfe einer Stichprobe eine genaueren Einblick über eine Hypothese über einen bestimmten Parameter zu geben.

16. Was ist eine Hypothese?

Eine Teilmenge des Parameterraums (=also eine Aussage über den Parameter z.B. μ>1)

17. Welche Typen von Hypothesen gibt es?

Nullhypothese (H0), also das was wir annehmen das stimmen sollte.

Alternativhypothese (H1), das Gegenteil.

18. Welche Fehler können beim Testen auftreten?

Fehler 1. Art: Nullhypothese wird verworfen, obwohl sie stimmt.

Fehler 2. Art: Nullhypothese wird akzeptiert, obwohl sie nicht zutrifft.

19. Was ist das Niveau (Signifikanzniveau) eines Tests?

Der Fehler 1. Art

20. Was ist der p-Wert und wie wird er verwendet?

In Programmen gibt's oft keine Niveaus, die man setzen kann, hier wird einfach der p-Wert zurückgegeben.

z.B ist der **p-Wert** beim einseitigen **Hypothesentest** (z.B H0 = μ < 1) und Stichprobe ist 1.5: P(X > 1.5) die **Restwsl**. **sozusagen**

21. Was versteht man unter einem Anpassungstest?

Man überprüft wie wahrscheinlich eine Stichprobe von einer bestimmten Verteilung genommen wurde. (z.B. ist der Würfel fair?)

22. Wie wird der Chi-Quadrat-Anpassungstest durchgeführt?

Siehe letzten Abschnitt

23. Wie hängen Tests und Konfidenzintervalle zusammen?

Sind sehr ähnlich, beim Konfidenzintervall bildet sich das Intervall um die Stichprobe, beim Hypothesentest bildet sich das Intervall um die Aussage (z.B μ = 1), und dann wird geschaut, ob der Mittelwert in diesem Intervall liegt.

Stochastik

Definition 3.1

Ein stochastischer Prozess ist eine Familie $(X_t, t \in T)$ von Zufallsvariablen. Die Indexmenge T wird Parameterraum genannt und soll eine Teilmenge der reellen Zahlen sein. Der Wertebereich Ω_X von X_t heißt Zustandsraum oder Phasenraum. Wenn T endlich oder abzählbar (etwa $\mathbb{N}$) ist, sprechen wir von einem Prozess in diskreter Zeit, wenn T ein ganzes (endliches oder unendliches) Intervall ist, von einem Prozess in stetiger Zeit,

Definition 3.2

(T_n) sei eine Folge von unabhängigen identisch verteilten Zufallsvariablen mit $T_n > 0$. Wir setzen $S_0 = 0$, $S_n = T_1 + \ldots + T_n$, und für $n = 0, \ldots$

$$X(t) = n \text{ für } S_n \leq t < S_{n+1}.$$

Wir nennen X einen Erneuerungsprozess.

Erneuerungsprozess Beispiel: Glühbirne geht kaputt

T1 = **Zeitpunkt** 1. Glühbirne geht **kaputt**, T2 = **Zeitpunkt** 2. Glühbirne …

n-te Glühbirne **beginnt** ab Sn-1 = T1 + T2 + … + Tn-1 zu arbeiten und geht kaputt bei Sn

X(t) (Oder X_t) gibt an wv Glühbirnen schon kaputt gegangen sind, dazu muss man nur schauen zwischen welchen zwei Summen man sich befindet.

Definition 3.3

$X_t(\omega)$ sei ein stochastischer Prozess. Wir nennen die Abbildung $t \mapsto X_t(\omega)$ als (als Funktion von t für fixes ω) einen Pfad, eine Trajektorie bzw. eine Realisation des Prozesses.

Definition 3.4

Der Prozess $X_t, t \in T$ heißt stationär, wenn wenn für $t_1 < t_2 < \ldots < t_n$ und $h > 0$ die gemeinsame Verteilung von $(X(t_1), \ldots, X(t_n))$ mit der von $(X(t_1 + h), \ldots, X(t_n + h))$ übereinstimmt.

Heißt wir können bei einer Folge von (X1, X2, …, Xn) einfach die ersten drei wegnehmen und haben dann (X4, X5, … Xn) und es bleibt **gleich verteilt**.

Satz 3.1: Ergodensatz von Birkhoff

Wenn die Folge (X_n) stationär ist und endlichen Erwartungswert $m = \mathbb{E}(X_n)$ hat, dann existiert

$$X_\infty = \lim_{n\to\infty} \bar{X}_n$$

mit Wahrscheinlichkeit 1 und

$$\mathbb{E}(X_\infty) = m.$$

Also sowas ähnliches wie das starke Gesetz der großen Zahlen.

Definition 3.5

Der Prozess $(X_t, t \in T)$ heißt Prozess mit unabhängigen Zuwächsen, wenn für $t_1 < t_2 < \ldots < t_n$ die Zufallsvariablen

$$X(t_1), X(t_2) - X(t_1), X(t_3) - X(t_2), \ldots, X(t_n) - X(t_{n-1})$$

unabhängig sind.

Daraus folgt die Markov Eigenschaft, also das die Zukunft nur vom letzten Zustand abhängt.

Definition 3.6

Der Prozess $X(t)$ heißt Markovprozess, wenn für $t_1 < t_2 < \ldots < t_n$

$$\mathbb{P}(X(t_n) \leq x_n | X(t_1) = x_1, \ldots, X(t_{n-1}) = x_{n-1}) = \mathbb{P}(X(t_n) \leq x_n | X(t_{n-1}) = x_{n-1}).$$

https://www.youtube.com/watch?v=6stYmO_N7LI&ab_channel=MITOpenCourseWare Gute Reihe für Poisson Prozesse

Definition 3.7

Der Poissonprozess mit Rate $\lambda > 0$ ist ein Prozess mit unahängigen Zuwächsen, mit $X(0) = 0$ und $X(t) - X(s) \sim P(\lambda(t-s))(s < t)$.

Beispiel: Wartezeiten für Bus oder ähnliches.

Markovketten

Definition 3.8

Die Wahrscheinlichkeiten

$$\mathbb{P}(X_{n+1} = j | X_n = i)$$

nennen wir die Übergangswahrscheinlichkeiten der Markovkette. Wenn diese nicht von n

abhängen, sprechen wir von einer homogenen Markovkette und setzen

$$p_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i)$$

Die Wahrscheinlichkeiten

$$p_{ij}(t) = \mathbb{P}(X_{n+t} = j | X_n = i)$$

nennen wir die t-stufigen Übergangswahrscheinlichkeiten.

Meisten Markov Ketten die wir verwenden sind **homogen**, da sich die **Wahrscheinlichkeit nicht ändert**.

Definition 3.9

Wir sagen, dass die Folge (X_n) ein (stochastisches) dynamisches System bildet, wenn es eine Folge (Y_n) von unabhängigen Zufallsvariablen und eine Funktion $f : \Omega_X \times \mathbb{R} \to \Omega_X$ gibt, sodass

$$X_{n+1} = f(X_n, Y_n)$$

gilt.

Satz 3.2: Chapman-Kolmogorov Gleichungen

$$p_{ij}(s+t) = \sum_{k \in \Omega_X} p_{ik}(s) p_{kj}(t).$$

in Matrixnotation mit den *t-stufigen Übergangsmatrizen*

$$P(t) = (p_{ij}(t))_{\Omega_X \times \Omega_X}$$

lauten die Chapman-Kolmogorov Gleichungen

$$P(t+s) = P(t)P(s).$$

Das heißt wenn man eine Übergangsmatrix in **t-Schritten** hat, und eine Übergangsmatrix mit **s-Schritten** hat, kann man die Matrizen multiplizieren, um die **(t+s)-Schritt** Matrix zu haben.

Definition 3.10

Der einfache Random Walk (auch: einfache Irrfahrt) ist der Partialsummenprozess

$$X(t) = X(0) + \sum_{i=1}^{t} Y_i,$$

wobei $(Y_1, Y_2, \ldots)$ unabhängige Zufallsvariable mit

$$\mathbb{P}(Y_i = 1) = p, \mathbb{P}(Y = -1) = 1 - p$$

sind. Der Spezialfall $p = 1/2$ heißt einfacher symmetrischer Random Walk.

Random Walk geht die Zustände random durch, zählt wie oft er in welchem Zustand war. Diesen dann durch die Anzahl der Durchgänge rechnen und dann hat man seine ungefähren Wahrscheinlichkeiten.

Klasseneigenschaften

Definition 3.11

Der Zustand j heißt Nachfolger von i ($i \to j$), wenn j von i aus mit positiver Wahrscheinlichkeit erreicht werden kann, also, wenn es ein $t \geq 0$ gibt, sodass $p_{ij}(t) > 0$.

Wenn sowohl $i \to j$ als auch $j \to i$ gilt, dann heißen i und j verbunden oder kommunizierend.

Definition 3.12

Eine Eigenschaft heißt Klasseneigenschaft, wenn sie entweder für alle Zustände einer Klasse oder für keinen gilt.

Definition 3.13

Die Periode eines Zustandes ist

$$d(i) = \mathrm{ggT}\{t \geq 0 : p_{ii}(t) > 0\}.$$

Wenn $d(i) = 1$ gilt, dann heißt der Zustand i aperiodisch, sonst periodisch.

Satz 3.3

Die folgenden Bedingungen sind äquivalent:

1. $\mathbb{P}_i(\tau_i < \infty) = 1$,
2. $\mathbb{P}_i(\nu_i = \infty) = 1$,
3. $\mathbb{E}_i(\nu_i) = \infty$,
4. $\sum_t p_{ii}(t) = \infty$.

Wenn diese Bedingungen erfüllt sind, dann heißt i rekurrent, sonst transient. Rekurrenz und Transienz sind Klasseneigenschaften.

Rekurrent sind sozusagen mehrere **Absorbierende Zustände**.

Absorbierende Zustände sind Zustände die nicht mehr verlassen werden können.

Transient sind **Zustände**, die **verlassen** werden können, und wo es eine Möglichkeit gibt, **nie** zurückzukommen.

Definition 3.14

i sei ein rekurrenter Zustand. Wenn

$$\mathbb{E}_i(\tau_i) < \infty$$

gilt, dann heißt i positiv rekurrent, sonst nullrekurrent.

Definition 3.15

$(\pi_i, i \in \Omega_X)$ heißt stationäre Verteilung, wenn

$$\pi_i \geq 0,$$

$$\sum_i \pi_i = 1$$

und

$$\sum_i \pi_i p_{ij} = \pi_j.$$

Wahrscheinlichkeiten, die sich unabhängig von den Zuständen nicht verändern und ein allgemeines Bild über den Prozess geben. (Eigenvektor der Übergangsmatrix)

Satz 3.4

Wenn (X_n) irreduzibel und aperiodisch ist, dann existieren die Grenzwerte

$$\lim_{n\to\infty} p_{ij}(t) = \pi_j = \frac{1}{\mathbb{E}_j(\tau_j)}.$$

Im periodischen Fall gilt

$$\pi_j = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} p_{ij}(t).$$

Eine Markov Kette ist dann irreduzibel, wenn man von jedem Zustand zu jedem Zustand gelangen kann.

Satz 3.5

Die Absorptionswahrscheinlichkeiten sind die kleinste nichtnegative Lösung des Gleichungssystems

$$a_{i_0} = 1,$$

$$a_i = \sum_j p_{ij} a_j, i \neq i_0.$$

Markov Ketten in stetiger Zeit

Gelten weiterhin:

- Markov Eigenschaft
- Chapman-Kolmogorov Gleichungen

Satz 3.6

Wenn die Übergangswahrscheinlichkeiten bei 0 stetig sind, dann existieren die Grenzwerte

$$q_{ij} = \lim_{t \to 0} \frac{p_{ij}(t) - p_{ij}(0)}{t}.$$

Die Zahlen q_{ij} heißen die infinitesimalen Parameter, die Matrix

$$Q = (q_{ij})_{\Omega_X \times \Omega_X}$$

die infinitesimale Matrix oder der infinitesimale Erzeuger der Markovkette.

Definition 3.16: Kolmogorovsche Differentialgleichungen

Die Gleichung

$$P'(t) = QP(t)$$

heißt die Rückwärtsgleichung,

$$P'(t) = P(t)Q$$

die Vorwärtsgleichung von Kolmogorov.

Definition 3.17

Die Markovkette X_t bzw. ihr Erzeuger Q heißt konservativ, wenn für alle i

$$q_{ii} > -\infty$$

und

$$\sum_j q_{ij} = 0$$

gilt.

Satz 3.7

Wenn die Markovkette X_t konservativ ist, dann erfüllen die Übergangswahrscheinlichkeiten $P(t)$ die Rückwärtsgleichung.

Wenn die infinitesimalen Parameter beschränkt sind, dann erfüllen die Übergangswahrscheinlichkeiten $P(t)$ die Vorwärtsgleichung.

Definition 3.18

Ein Geburts- und Todesprozess ist eine Markovkette mit Zustandsraum $\{0, 1, \ldots\}$ und infinitesimalen Parametern

$$q_{00} = -\lambda_0, q_{01} = \lambda_0,$$

$$q_{i,i-1} = \mu_i, q_{ii} = -(\lambda_i + \mu_i), q_{i,i+1} = \lambda_i.$$

Die Zahlen λ_i heißen die Geburtsraten, μ_i die Todesraten. Sind alle $\mu_i = 0$, dann sprechen wir von einem reinen Geburtsprozess, sind alle $\lambda_i = 0$, von einem reinen Todesprozess.

Definition 3.19

Eine Markovkette in stetiger Zeit heißt regulär, wenn sie in endlicher Zeit nur endlich viele Sprünge ausführt, im anderen Fall singulär..

Satz 3.8

Ist X regulär, dann gelten die Vorwärts- und die Rückwärtsgleichung, und sie sind eindeutig lösbar.

Wahrscheinlichkeit und Stochastik WS2021

1. Was ist ein stochastischer Prozess?

Eine Familie von Zufallsvariablen Xt, die unabhängig oder abhängig von einander sein können. t ϵ T ist der **Parameterraum** (z.B. ℕ)

2. Was ist ein Erneuerungsprozess?

Familie von unabhängigen Zufallsvariablen.

Erneuerungsprozess Beispiel: Glühbirne geht kaputt

T1 = **Zeitpunkt** 1. Glühbirne geht **kaputt**, T2 = **Zeitpunkt** 2. Glühbirne …

n-te Glühbirne **beginnt** ab Sn-1 = T1 + T2 + … + Tn-1 zu arbeiten und geht kaputt bei Sn

X(t) (Oder X_t) gibt an wv Glühbirnen schon kaputt gegangen sind, dazu muss man nur schauen zwischen welchen zwei Summen man sich befindet.

Es gilt: X(t) = n für Sn < t < Sn+1

3. Wie ist ein stationärer Prozess definiert?

Wenn einzelne **Teilmengen** vom Prozess Xt auch **unabhängig** verteilt sind**. (Xt1, …, Xtn) unabhängig von (Xt4, …, Xtn+4)**

4. Wie ist ein Prozess mit unabhängigen Zuwächsen definiert?

Ein Prozess der unabhängig von der Vergangenheit ist. Also X(t1), X(t2) – X(t1), X(t3) – X(t2) unabhängig.

5. Wie ist die Markoveigenschaft definiert?

$$\mathbb{P}(X(t_n) \leq x_n | X(t_1) = x_1, \ldots, X(t_{n-1}) = x_{n-1}) = \mathbb{P}(X(t_n) \leq x_n | X(t_{n-1}) = x_{n-1})$$

Die Wahrscheinlichkeit hängt nur vom derzeitigen Zustand ab und nicht von der Vergangenheit.

6. Was ist eine Markovkette?

Stochastischer Prozess mit **diskretem Zustandsraum** welcher die Markoveigenschaft erfüllt.

7. Wie lauten die Chapman-Kolmogorov-Gleichungen?

$$p_{ij}(s+t) = \sum_{k \in \Omega_X} p_{ik}(s) p_{kj}(t).$$

in Matrixnotation mit den *t-stufigen Übergangsmatrizen*

$$P(t) = (p_{ij}(t))_{\Omega_X \times \Omega_X}$$

lauten die Chapman-Kolmogorov Gleichungen

$$P(t+s) = P(t)P(s).$$

8. Wie ist der einfache Random Walk definiert, und wie lauten seine Übergangswahrscheinlichkeiten?

Partialsummenprozess: Xt = X0 + Summe(Yi) wobei P(Yi = 1) = p und P(Yi = -1) = 1-p

Die Übergangswahrscheinlichkeiten pij sind

- **p wenn j = i+1**
- **1-p wenn j = i-1**
- **Sonst 0**

9. Wann heißt ein Zustand Nachfolger eines anderen?

j ist Nachfolger von i, wenn pij(t) > 0

10. Wann heißen zwei Zustände kommunizierend?

Wenn pij(t) > 0 und pji(t) > 0

11. Was ist eine Klasseneigenschaft?

Eine Eigenschaft die für eine ganze Klasse an Zuständen zutrifft.

12. Wann heißt eine Markovkette irreduzibel?

Wenn alle Zustände von allen Zuständen erreicht werden können

13. Wie ist die Periode eines Zustands definiert?

$$d(i) = \mathrm{ggT}\{t \geq 0 : p_{ii}(t) > 0\}.$$

Wenn d = 1 ist der Zustand aperiodisch.

14. Wann heißt ein Zustand rekurrent?

Rekurrent sind sozusagen mehrere **Absorbierende Zustände**. (Also wenn eine Klasse an Zuständen nur mehr untereinander erreichbar sind) bzw. eigentlich einfach nur das der Zustand mit Wsl 1 wieder erreicht wird.

15. Wie ist positive Rekurrenz definiert?

Wenn der Erwartungswert der Wartezeit bis der Zustand wieder erreicht wird endlich ist. (Sozusagen das die Markov Kette recht schnell den Zustand wieder erreicht)

Bei endlichen Zuständen, sind rekurrente Zustände immer positiv rekurrent

16. Wie verhalten sich die Übergangswahrscheinlichkeiten pij (t) in einer irreduziblen aperiodischen Markovkette für t → ∞?

Sie konvergieren gegen die Stationäre Verteilung.

17. Was ist eine stationäre Verteilung?

Der Eigenwert der Übergangsmatrix sozusagen. Gibt die Wahrscheinlichkeit in welchem Zustand man sich im nächsten Schritt befinden wird unabhängig vom derzeitigen Zustand.

18. Wie kann man die positive Rekurrenz anders charakterisieren? (Keine Ahnung was die Frage heißt)

Das der rekurrente Zustand schneller erreicht wird.

19. Was kann man über die Rekurrenz des Random Walk aussagen?

Das ein bestimmter Zustand wieder erreicht wird? **-Keine Ahnung**

20. Was kann man über die Rekurrenzeigenschaften von endlichen Markovketten haben.

Eine Stationäre Verteilung? **-Keine Ahnung**

21. Wie sind Absorptionswahrscheinlichkeiten definiert?

Das man von einem Zustand **i** von einem Zustand **i0** absorbiert wird, also sozusagen in ein „Loch" fällt.

$$a_i = \mathbb{P}_i(\tau_{i_0} < \infty) = \mathbb{P}_i(X \text{wird in } i_0 \text{ absorbiert}).$$

Wahrscheinlichkeit und Stochastik WS2021

22. Wie kann man die Absorptionswahrscheinlichkeiten berechnen?

Für einen Absorbierenden Zustand i0:

$$a_{i_0} = 1,$$
$$a_i = \sum_j p_{ij} a_j, i \neq i_0.$$

23. Welche speziellen Annahmen machen wir für Markovketten in stetiger Zeit?

Das die t-stufigen Übergangswahrscheinlichkeiten bei t=0 stetig sind. (Also Pij(0) = 1 wenn i = j)

24. Wie kann man Rekurrenz, Transienz, positive Rekurrenz und Nullrekurrenz für Markovketten in stetiger Zeit definieren?

-

25. Was sind die infinitesimalen Parameter einer Markovkette?

Die Matrix Q, zeitliche Ableitungen der Übergangswahrscheinlichkeiten, zeigen an wie oft zwischen Zuständen gewechselt wird.

26. Wie lauten die Kolmogorovschen Differentialgleichungen?

Die Gleichung

$$P'(t) = QP(t)$$

heißt die Rückwärtsgleichung,

$$P'(t) = P(t)Q$$

die Vorwärtsgleichung von Kolmogorov.

27. Was kann man über die Gültigkeit der Kolmogorovschen Differentialgleichungen sagen?

Sie hängt davon ab ob die Markovkette konservativ (qii > -unendlich, Summe(qij) über j = 0) ist oder nicht. Wenn ja, erfüllt sie die Rückwärtsgleichung, sonst wenn sie beschränkt sind, die Vorwärtsgleichung.

28. Wie kann man für endliche Markovketten die Übergangsmatrizen aus dem infinitesimalen Erzeuger erhalten?

$$P(t) = e^{Qt} = E\Lambda^t E^{-1} = E \begin{pmatrix} e^{t\lambda_1} & & \\ & \ddots & \\ & & e^{t\lambda_d} \end{pmatrix} E^{-1}$$

29. Wie kann man die infinitesimalen Parameter anschaulich (?) interpretieren?

Wie oft die Kette zwischen Zuständen hin und her springen wird.

30. Wie kann mithilfe des infinitesimalen Erzeugers die Absorptionswahrscheinlichkeiten berechnen?

$$\sum_j q_{ij} a_j = 0.$$

Wobei ai = 1 und **i** der absorbierende Zustand ist.

31. Wie kann mithilfe des infinitesimalen Erzeugers die stationäre Verteilung berechnen?

Übergangsmatrizen bestimmten, potenziereren mit unendlich