Zusammenfassung Theoretische Informationstechnik

28. Januar 2012

1 Allgemeines

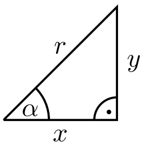
$(u \cdot v)' = u' \cdot v + u \cdot v'$ $\quad (\frac{u}{v})' = \frac{u' \cdot v - u \cdot v'}{v^2}$ $\quad \int u' \cdot v \, dx = u \cdot v - \int u \cdot v' \, dx$ $\quad \frac{dy(x(t))}{dt} = y'(x(t)) \cdot x'(t)$

$\frac{d\sin(x)}{dt} = \cos(x)$ $\quad \frac{d\cos(x)}{dt} = -\sin(x)$ $\quad \frac{d\tan(x)}{dx} = \frac{1}{\cos^2(x)}$ $\quad \frac{d\ln(|x|)}{dx} = \frac{1}{x}$

$e^{jx} = \cos(x) + j\sin(x)$ $\quad j = \sqrt{-1}$ $\quad |z_1 \cdot z_2| = |z_1| \cdot |z_2|$ mit $z_1, z_2 \in \mathbb{C}$

$e^{jx} + e^{-jx} = 2 \cdot \cos(x)$ $\quad e^{jx} - e^{-jx} = 2j \cdot \sin(x)$ $\quad e^x + e^{-x} = 2 \cdot \cosh(x)$ $\quad e^x - e^{-x} = 2 \cdot \sinh(x)$

$\log_b r = \frac{\log_a r}{\log_a b}$ $\quad \log(ab) = \log(a) + \log(b)$ $\quad \log(\frac{a}{b}) = \log(a) - \log(b)$ $\quad \log(a^b) = b \cdot \log(a)$



$\cos(\alpha) = \frac{x}{r}$ $\quad \sin(\alpha) = \frac{y}{r}$ $\quad \tan(\alpha) = \frac{\sin(\alpha)}{\cos(\alpha)} = \frac{y}{x}$

$\cos(x) = \cos(-x)$ $\quad -\sin(x) = \sin(-x)$ $\quad -\arctan(x) = \arctan(-x)$ $\quad \mathrm{si}(x) = \frac{\sin(x)}{x}$ $\quad 1rad = \frac{360°}{2\pi}$

$\cos(a \pm b) = \cos(a)\cos(b) \mp \sin(a)\sin(b)$

$\sin(a) \cdot \sin(b) = \frac{1}{2}(\cos(a-b) - \cos(a+b))$ $\quad \cos(a) \cdot \cos(b) = \frac{1}{2}(\cos(a-b) + \cos(a+b))$

$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a \cdot d - c \cdot b$ $\quad \det\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = aei + bfg + cdh - gec - hfa - idb$ $\quad \det(\mathbf{A}^{-1}) = \frac{1}{\det(\mathbf{A})}$

transponierte Matrix: $\mathbf{A}^T$ $\quad$ transponierte und konjugiert komplexe Matrix: $\mathbf{A}^*$

1.1 Unbestimmte Integrale

$\int e^{at} \, dt = \frac{1}{a} e^{at}$ $\quad \int t e^{at} \, dt = \frac{e^{at}}{a^2}(at - 1)$ $\quad \int t^2 e^{at} \, dt = \frac{e^{at}}{a}(t^2 - \frac{2t}{a} + \frac{2}{a^2})$

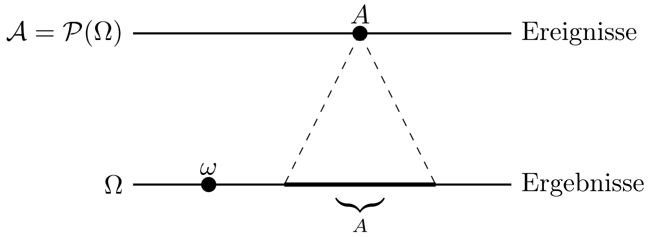
$\int e^{at} \sin(bt) \, dt = \frac{e^{at}}{a^2+b^2}(a\sin(bt) - b\cos(bt))$ $\quad \int e^{at} \cos(bt) \, dt = \frac{e^{at}}{a^2+b^2}(a\cos(bt) + b\sin(bt))$

2 Stochastische Modellierung

Ergebnismenge Ω:
Alle möglichen, einander ausschliessenden, nicht weiter zerlegbaren Ausgänge eines bestimmten Zufallsversuches. Es muss dabei genau ein Ergebnis eintreten.
z.B.: $\underbrace{\{0,1\}}_{\text{einmaliger Münzwurf}}, \{0,1\}^{\mathbb{N}}, \underbrace{\{1,\dots,6\}}_{\text{einmaliges Würfeln}}, \underbrace{\{1,\dots,6\}^n}_{\text{mehrmaliges Würfeln mit Reihenfolge}}, \mathbb{N}, \mathbb{R}, \mathbb{R}^n, \mathbb{C}^n$

Ereignismenge $\mathcal{A}$:
Alle relevanten Ereignisse fasst man zur Ereignismenge zusammen. Sie besteht aus den interessierenden

Einzelergebnissen (*Elementarereignisse*) und ist eine Teilmenge von Ω, oft gilt $\mathcal{A} = \mathcal{P}(\Omega)$. Ein Ereignis tritt ein, wenn eins seiner Ergebnisse eintritt.
z.B.: bei n Münzwürfen fällt k-mal Zahl



Ereignis $A \in \mathcal{A}$, $A \subseteq \Omega$
Ergebnis $\omega \in \Omega$

Beispiel:
Eine Münze wird immer zwei mal geworfen, daher die Ergebnismenge $\Omega = \{(K,K),(K,Z),(Z,K),(Z,Z)\}$. Die Ergebnismenge ist nicht einfach $\Omega = \{K,Z\}$, da beim Zufallsexperiment die Münze zwei mal geworfen wird, also muss das Ergebnis den Ausgang beider Münzwürfe beinhalten!
z.B. Kopf beim ersten Wurf: A $= \{(K,K),(K,Z)\}$

Wahrscheinlichkeitsmaß: $P\colon \mathcal{A} \to [0,1]$
Zu jedem Ereignis A wird eine Maßzahl $P(A) \in [0,1]$ gesucht, die den Grad der Wahrscheinlichkeit des Ereignisses A angibt. Die Wahl von P ist am heikelsten, da diese die eigentliche relevante Information über das Zufallsgeschehen enthält.

es gilt:
Normierung: $P(\Omega) = 1$
σ-Additivität: bei unvereinbaren Ereignissen addieren sich die Wahrscheinlichkeiten
$P\left(\bigcup\limits_{i\geq 1} A_i\right) = \sum\limits_{i\geq 1} P(A_i)$ mit $A_i \cap A_j = \emptyset$ für $i \neq j$ mit $i,j \in \mathbb{N}$ (paarweise disjunkt)
$P(\emptyset) = 0$
$P(A\cup B) + P(A\cap B) = P(A) + P(B)$ ggf. $P(A\cap B) = 0$ wegen $A \cap B = \emptyset$
$P(A) + P(A^C) = 1$ mit $P(A^C) = P(\Omega \setminus A)\ \forall A \in \mathcal{A}$
$A \subseteq B \Rightarrow \begin{cases} P(A) \leq P(B) \\ P(B\setminus A) = P(B) - P(A) \end{cases}$
$P(A) = \frac{\text{Anzahl günstiger Ergebnisse}}{\text{Anzahl möglicher Ergebnisse}} = \frac{|A|}{|\Omega|}$ mit $A \in \mathcal{A}$ (bei gleicher Wahrscheinlichkeit der Ergebnisse)

Interpretation von Wahrscheinlichkeitsmaßen:

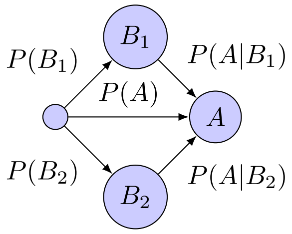
- $P(A)$ ist der Grad der Sicherheit, mit der A eintritt
- $P(A)$ ist die relative Häufigkeit, mit der A unter den gleichen äußeren Bedingungen einzutreten pflegt

Mengen: $A, B \in \mathcal{A}$
$A\cup B$: Vereinigung, «A oder B treten ein» (mindestens eins tritt ein)
$A \cap B$: Schnitt, «A und B treten ein»
A^C: Komplement, «A tritt nicht ein»
$\emptyset$: leere Menge
$|A|$: Anzahl der Elemente in A (Kardinalität von A)
$A \subseteq B$: A ist Teilmenge von B
Regeln von de Morgan: $P(A\cup B) = P((A^C \cap B^C)^C)$ und $P(A\cap B) = P((A^C \cup B^C)^C)$

bedingte Wahrscheinlichkeit: $A, B \in \mathcal{A}$ mit $P(B) \neq 0$
Die Wahrscheinlichkeit, dass das Ereignis A eintritt, wenn B schon eingetreten ist.
$P(A|B) = \begin{cases} 0 & \text{für } P(B) = 0 \\ \frac{P(A\cap B)}{P(B)} & \text{sonst} \end{cases}$

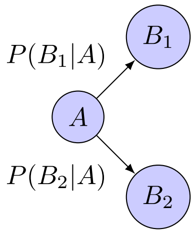
Beispiel:
Wahrscheinlichkeit, dass eine 2 gewürfelt wurde, wenn das Ergebnis des Wurfes eine gerade Zahl ist:
$A = \{2\}, B = \{2,4,6\}$, dann $P(2|\text{gerade Zahl}) = \frac{P(\{2\}\cap\{2,4,6\})}{P(\text{gerade Zahl})} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$

Satz von der totalen Wahrscheinlichkeit:
Alle Wege zum Ereignis A.
$P(A) = \sum_{n=1}^{\infty} P(A \cap B_n) = \sum_{n=1}^{\infty} P(A|B_n) \cdot P(B_n)\ \forall A \in \mathcal{A}$ mit $B_i \cap B_j = \emptyset$ für $i \neq j$ mit $i, j \in \mathbb{N}$



Bayes-Formel: $P(A) > 0$
Bedingte Wahrscheinlichkeit des Ereignisses B_n unter der Voraussetzung, dass A bereits eingetreten ist.
(Wahrscheinlichkeit des Weges über B_n im Verhältnis zur Gesamtwahrscheinlichkeit aller Wege nach A)
(Wie wahrscheinlich ist es, dass man über B_n nach A gekommen ist.)
$P(B_n|A) = \frac{P(A\cap B_n)}{P(A)} = \frac{P(A|B_n)\cdot P(B_n)}{P(A)} = \frac{P(A|B_n)\cdot P(B_n)}{\sum_{k=1}^{\infty} P(A|B_k)\cdot P(B_k)}$

und $\sum_{n=1}^{\infty} P(B_n|A) = 1$ (alle wegführenden Wege vom Ereignis A)



stochastische Unabhängigkeit: *«A und B sind stochastisch unabhängig»*
Das Eintreten von A hat keine Auswirkung auf die Wahrscheinlichkeit, dass B eintritt.
$P(A \cap B) = P(A) \cdot P(B) \Rightarrow P(A^C \cap B^C) = P(A^C) \cdot P(B^C)$

Kombinatorik:
Anzahl der Möglichkeiten beim Ziehen aller n Kugeln aus einer Urne ohne Zurücklegen mit Berücksichtigung der Reihenfolge: $n!$

Anzahl verschiedener Auswahlen von k Elementen aus n Elementen:

k aus n	mit Berücksichtigung der Reihenfolge	ohne Berücksichtigung der Reihenfolge
mit Wiederholung	n^k, Zahlenschloss (n Ziffern, k Stellen)	$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$
ohne Wiederholung	$\frac{n!}{(n-k)!}$	$\binom{n}{k} = \frac{n!}{k!\cdot(n-k)!}$, Lotto «6 aus 49»

Zufallsvariable: $X\colon \Omega \to \mathbb{R}$
Oft interessiert nicht das gesamte Modell, sondern nur gewisse Teilgrößen. Jedem Ergebnis eines Zufallsexperimentes wird eine Zahl zugeordnet. Die Menge aller Ergebnisse deren Realisation unterhalb eines bestimmten Wertes liegt muss ein Ereignis bilden.
Ergebnis $\to$ Zahl
Ereignis $\to$ Intervall

$X^{-1}((-\infty, x]) = \underbrace{\{\omega \in \Omega \mid X(\omega) \leq x\}}_{\text{Ereignis}} = \{X \leq x\} = A \in \mathcal{A}$

Wahrscheinlichkeit eines Ereignisses: $P(\{\omega \in \Omega \mid X(\omega) \leq x\}) = P(X \leq x)$ mit $x \in \mathbb{R}$

Beispiel:
Steht etwa Ω für die Menge der möglichen Ausgänge eines Glücksspiels, so könnte $X(\omega)$ der Gewinn sein, den ein Spieler beim Ausgang ω des Spiels erhält, wobei ein negativer Wert einen Verlust darstellt.

Verteilungsfunktion einer Zufallsvariablen: $F_X(x) = P(X \leq x)$ mit $0 \leq F_X(x) \leq 1 \; \forall \, x \in \mathbb{R}$
Die Verteilungsfunktion einer Zufallsvariablen X ist die Wahrscheinlichkeit dafür, dass X einen Wert annimmt, der kleiner oder gleich einer beliebigen Zahl $x \in \mathbb{R}$ ist.
Die Verteilung jeder Zufallsvariablen X ist eindeutig durch die zugehörige Verteilungsfunktion beschrieben.

Eigenschaften: (*Die Umkehrung gilt auch* $\Longleftrightarrow$)

- $F_X(x)$ ist monoton steigend $(x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2))$ (da die Wahrscheinlichkeit nicht sinkt)
- unsicheres Ereignis: $\lim\limits_{x \to -\infty} F_X(x) = 0$
- sicheres Ereignis: $\lim\limits_{x \to \infty} F_X(x) = 1$
- $F_X(x)$ ist rechtsseitig stetig: $F_X(x) = \lim\limits_{h \to 0^+} F_X(x+h)$

absolut-stetige Zufallsvariable:
$F_X(x) = P(X \leq x) = \int\limits_{-\infty}^{x} f_X(t)\, dt \; \forall \, x \in \mathbb{R}$

Verteilungsdichte: $f_X \colon \mathbb{R} \to \mathbb{R}^+$ (ist $\neq$ der Wahrscheinlichkeit!)
mit $\int\limits_{-\infty}^{\infty} f_X(x)\, dx = 1$ und $f(x) \geq 0 \; \forall \, x \in \mathbb{R}$, dann $f_X(x) = \frac{dF_X(x)}{dx}$ wenn $F_X(x)$ differenzierbar ist

Berechnung von Wahrscheinlichkeiten mit Hilfe der Verteilungsdichte:
$P(X > x) = 1 - P(X \leq x) = 1 - F_X(x) = 1 - \int\limits_{-\infty}^{x} f_X(t)\, dt$

$P(x_1 < X \leq x_2) = P(X \leq x_2) - P(X \leq x_1) = F_X(x_2) - F_X(x_1) = \int\limits_{x_1}^{x_2} f_X(t)\, dt$

diskrete Zufallsvariable:
$X \colon \Omega \to T$ mit dem Träger der Zufallsvariablen $T = \{t_1, t_2, \ldots, t_n\} \subset \mathbb{R}$

diskrete Zähldichte:
$f_X(t_i) \colon T \to [0,1]$ mit $f_X(t_i) = P(X = t_i)$
es gilt $\sum\limits_{t_1 \leq k \leq t_n} f_X(k) = 1$

diskrete Verteilungsfunktion:
$F_X(t_i) = P(X \leq t_i) = \sum\limits_{k \leq t_i} P(X = k)$ mit $t_i, k \in T$
$P(a \leq X \leq b) = \sum\limits_{a \leq k \leq b} f_X(k)$ mit $a, b, k \in T$

diskrete Verteilungen:

a) Diskrete Gleichverteilung $X \sim \mathrm{U}(\{1, \ldots, n\})$: $P(X = i) = \frac{1}{n} \quad \forall \, i = 1, \ldots, n$
z.B. Würfel- oder Münzwurf

b) Bernoulli-Verteilung $X \sim \mathrm{Ber}(p)$: $P(X = 1) = p$ und $P(X = 0) = 1 - p \quad$ mit $p \in [0, 1]$
z.B. Münze mit Kopf (p) und Zahl $(1 - p)$

c) Binomialverteilung $X \sim \mathrm{Bin}(n,p)$: $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$ mit $k = 0, 1, \ldots, n$ und $p \in [0,1]$
z.B. bei n Münzwürfen k-mal Kopf oder Zahl (p)

d) Geometrische Verteilung $X \sim \mathrm{Geo}(p)$: $P(X=k) = p(1-p)^k$ mit $k \in \mathbb{N}_0$ und $p \in (0,1]$
z.B. k Fehlversuche vor dem ersten Erfolg (p)

e) Poisson-Verteilung $X \sim \mathrm{Poi}(\lambda)$: $P(X=k) = \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda}$ mit $k \in \mathbb{N}_0$ und $\lambda > 0$, $E(X) = \lambda$ und $Var(X) = \lambda$, mehrfaches Bernoulli-Experiment

absolut-stetige Verteilungen:

a) Normalverteilung $X \sim \mathrm{N}(\mu, \sigma^2)$: $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $\mu \in \mathbb{R}, \sigma^2 > 0$
z.B. Abweichungen der (Mess-) Werte vieler natur- und ingenieurwissenschaftlicher Vorgänge vom Mittelwert

b) Gleichverteilung $X \sim \mathrm{R}(a,b)$: $f_X(x) = \frac{1}{b-a} \cdot \mathbb{I}_{[a,b]}(x) = \begin{cases} \frac{1}{b-a} & \text{falls } a \le x \le b \\ 0 & \text{sonst} \end{cases}$, $a, b \in \mathbb{R}$

c) Exponentialverteilung $X \sim \mathrm{Exp}(\lambda)$: $f_X(x) = \lambda \mathrm{e}^{-\lambda x} \cdot \mathbb{I}_{[0,\infty)}(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x} & \text{falls } x \ge 0 \\ 0 & \text{sonst} \end{cases}$, $\lambda > 0$
$E(X) = \frac{1}{\lambda}$ und $Var(X) = \frac{1}{\lambda^2}$
z.B. Wahrscheinlichkeit des Ausfalls einer Komponente nach der Zeit $x > 0$

d) Rayleigh-Verteilung $R \sim \mathrm{Ray}(\sigma^2)$: $f_R(r) = \frac{r}{\sigma^2} \mathrm{e}^{-\frac{r^2}{2\sigma^2}} \cdot \mathbb{I}_{[0,\infty)}(r) = \begin{cases} \frac{r}{\sigma^2} \mathrm{e}^{-\frac{r^2}{2\sigma^2}} & \text{falls } r \ge 0 \\ 0 & \text{sonst} \end{cases}$, $\sigma^2 > 0$
$E(R) = \sqrt{\frac{\pi}{2}}\sigma$, $Var(R) = (2 - \frac{\pi}{2})\sigma^2$
z.B. Betrag der Komponenten eines zweidimensionalen Zufallsvektors, die normalverteilt und stochastisch unabhängig sind

e) Rice-Verteilung $R \sim \mathrm{Rice}(\mu, \sigma^2)$: $f_R(r) = \frac{r}{\sigma^2} \mathrm{e}^{-\frac{r^2+\mu^2}{2\sigma^2}} \cdot \frac{1}{\pi} \int\limits_0^\pi \mathrm{e}^{\frac{r\mu}{\sigma^2}\cos(\vartheta)}\, d\vartheta$, $\mu > 0, \sigma^2 > 0$
Verallgemeinerung der Rayleigh-Verteilung

f) Lognormal-Verteilung $Y \sim \mathrm{LogN}(\mu, \sigma^2)$: $f_Y(y) = \begin{cases} \frac{1}{y\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(\ln(y)-\mu)^2}{2\sigma^2}} & \text{falls } y > 0 \\ 0 & \text{sonst} \end{cases}$, $\mu \in \mathbb{R}, \sigma^2 > 0$
Verteilung einer Zufallsvariablen Y, wenn $X = \ln(Y)$ normalverteilt ist

Erwartungswert: mittlerer (typischer) Wert der Zufallsvariablen, Lagemaß
$E(g(X)) = \sum\limits_{i=1}^{\infty} g(x_i) \cdot f_X(x_i) = \sum\limits_{i=1}^{\infty} g(x_i) \cdot P(X = x_i)$ falls $\sum\limits_{i=1}^{\infty} |g(x_i)| \cdot f_X(x_i) < \infty$
$E(g(X)) = \int\limits_{-\infty}^{\infty} g(x) \cdot f_X(x)\, dx$ falls $\int\limits_{-\infty}^{\infty} |g(x)| \cdot f_X(x)\, dx < \infty$

k-tes Moment von X: $E(X^k)$ mit $k \in \mathbb{N}_0$

k-tes zentrales Moment von X: $E((X - E(X))^k)$

Varianz:
mittlere (typische) quadratische Abweichung der Zufallsvariablen vom Erwartungswert, Streuungsmaß
$\sigma^2 = Var(X) = E((X - E(X))^2) = E(X^2) - E^2(X)$

Standardabweichung: $\sigma = \sqrt{Var(X)}$

Kovarianz: Korrekturterm bei der Berechnung der Varianz von Summen
$Cov(X,Y) = E((X - E(X)) \cdot (Y - E(Y))) = E(X \cdot Y) - E(X) \cdot E(Y)$

Korrelation: Maß für den linearen Zusammenhang zwischen Zufallsvariablen
$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}}$

Rechenregeln: $a, b, c \in \mathbb{R}$

- Linearität des Erwartungswertes: $E(a \cdot X + b \cdot Y) = a \cdot E(X) + b \cdot E(Y)$
- Erwartungswert einer Konstanten: $E(a) = a$
- Erwartungswert zweier stochastisch unabhängiger Zufallsvariablen: $E(X \cdot Y) = E(X) \cdot E(Y)$
- Monotonie des Erwartungswertes: $X \leq Y \Rightarrow E(X) \leq E(Y)$
- Markow-Ungleichung: $P(|X| > c) \leq \frac{E(|X|)}{c}\ \forall c > 0$
- Tschebyschow-Ungleichung: $P(|X - E(X)| \geq c) \leq \frac{Var(X)}{c^2}\ \forall c > 0$
- $Var(X + Y) = Var(X) + Var(Y) + 2 \cdot Cov(X,Y)$
- $Var(a \cdot X + b) = a^2 \cdot Var(X) \geq 0$
- $Cov(X,X) = Var(X)$
- $Cov(a \cdot X + b, c \cdot Y + d) = ac \cdot Cov(X,Y)$
- $Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$
- $|Cov(X,Y)| \leq \sqrt{Var(X) \cdot Var(Y)}$ also $|Corr(X,Y)| \leq 1$

Zufallsvektor:
$\mathbf{X} = (X_1, \ldots, X_n)^T \colon \Omega \to \mathbb{R}^n$
mit den Zufallsvariablen $X_1, \ldots, X_n$ aus dem selben Wahrscheinlichkeitsraum.
Das Ziel ist die Beschreibung von gemeinsamen zufälligen Ausgängen, z.B. ein zufälliges Signal $R \cdot e^{j \cdot \varphi}$ mit $\mathbf{X} = (R, \varphi)^T$.

gemeinsame Verteilungsfunktion: $(\mathbf{X} \sim F_{\mathbf{X}})$
$F_{\mathbf{X}}(x_1, \ldots, x_n) = P(\{\omega \in \Omega \mid X_1(\omega) \leq x_1, \ldots, X_n(\omega) \leq x_n\}) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$
Beschreibt die Verteilung des Zufallsvektors eindeutig.

Verteilungsdichte: $f_{\mathbf{X}} \colon \mathbb{R}^n \to \mathbb{R}^+$ $(\mathbf{X} \sim f_{\mathbf{X}})$
$F_{\mathbf{X}}(x_1, \ldots, x_n) = \int\limits_{-\infty}^{x_n} \ldots \int\limits_{-\infty}^{x_1} f_{\mathbf{X}}(t_1, \ldots, t_n)\, dt_1 \ldots dt_n\ \forall x_1, \ldots, x_n \in \mathbb{R}$

$\mathbf{X}$ ist absolut-stetig: $P(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2) = \int\limits_{a_2}^{b_2} \int\limits_{a_1}^{b_1} f_{\mathbf{X}}(x_1, x_2)\, dx_1 dx_2$

n-dimensionale Normalverteilung: $\mathrm{N}_n(\boldsymbol{\mu}, \mathbf{C})$
Erwartungswerte: $\boldsymbol{\mu} \in \mathbb{R}^n$, Kovarianzen: $\mathbf{C} \in \mathbb{R}^{n \times n}$ symmetrisch positiv definit $(\mathbf{x}^T \cdot \mathbf{C} \cdot \mathbf{x} > \mathbf{0}\ \forall \mathbf{x} \neq \mathbf{0})$
$f_{\mathbf{X}}(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \cdot \sqrt{|\det(\mathbf{C})|}} \exp(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \cdot \mathbf{C}^{-1} \cdot (\mathbf{x} - \boldsymbol{\mu}))$ mit $\mathbf{x} \in \mathbb{R}^n$

z.B. sei $\mathbf{C} = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{pmatrix} = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{12}^2 & \sigma_{22}^2 \end{pmatrix}$, dann $\mathbf{C}^{-1} = \frac{1}{\det(\mathbf{C})} \cdot \begin{pmatrix} \sigma_{22}^2 & -\sigma_{12}^2 \\ -\sigma_{21}^2 & \sigma_{11}^2 \end{pmatrix} = \frac{1}{\det(\mathbf{C})} \cdot \begin{pmatrix} \sigma_{22}^2 c & -\sigma_{12}^2 \\ -\sigma_{12}^2 & \sigma_{11}^2 \end{pmatrix}$

Erwartungswertvektor:
$$E(\mathbf{X}) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix} \in \mathbb{R}^n$$

Kovarianzmatrix: (symmetrisch)
$\mathbf{C} = Cov(\mathbf{X}) = (Cov(X_i, X_j))_{1 \leq i,j \leq n} \in \mathbb{R}^{n \times n}$ (i senkrecht, j waagerecht)

Rechenregeln: $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{b} \in \mathbb{R}^m$ (m senkrecht, n waagerecht)

- $E(\mathbf{A} \cdot \mathbf{X} + \mathbf{b}) = \mathbf{A} \cdot E(\mathbf{X}) + \mathbf{b} \in \mathbb{R}^m$
- $Cov(\mathbf{A} \cdot \mathbf{X} + \mathbf{b}) = \mathbf{A} \cdot Cov(\mathbf{X}) \cdot \mathbf{A}^T \in \mathbb{R}^{m \times m}$

stochastische Unabhängigkeit: $\mathbf{X} = (X_1, \ldots, X_n)^T$
$X_1, \ldots, X_n$ sind stochastisch unabhängig, wenn
$F_{\mathbf{X}}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_n}(x_n) = \prod\limits_{i=1}^{n} P(X_i \leq x_i) \; \forall \, x_1, \ldots, x_n \in \mathbb{R}$

X ist stochastisch unabhängig und hat die selbe Verteilung:
«stochastisch unabhängig und identisch verteilt»

gemeinsame Dichte, wenn **X** absolut-stetig ist:
$f_{\mathbf{X}}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n) \; \forall \, (x_1, \ldots, x_n)^T \in \mathbb{R}^n \Rightarrow$ stochastisch unabhängig

$Cov(\mathbf{C}) = diag(\sigma_1^2, \ldots, \sigma_n^2) = diag(Var(X_1), \ldots, Var(X_n))$

die diskreten Zufallsvariablen $X_1, \ldots, X_n$ sind stochastisch unabhängig
$\Leftrightarrow P(X_1 = x_1, \ldots, X_n = x_n) = \prod\limits_{i=1}^{n} P(X_i = x_i) \; \forall \, x_i \in T_i$

gemeinsame Zähldichte:
$f_{(X,Y)}(x, y) = P(X = x, Y = y)$ mit $(x, y) \in T_X \times T_Y$

diskrete bedingte Zähldichte:
$$f_{X|Y}(x|y) = \begin{cases} \frac{P(X=x,Y=y)}{P(Y=y)} = \frac{f_{(X,Y)}(x,y)}{f_Y(y)} & \text{falls } f_Y(y) > 0 \\ f_X(x) & \text{falls } f_Y(y) = 0 \end{cases}$$

diskreter bedingter Erwartungswert:
$E(g(X)|Y = y) = \sum\limits_{x \in T_X} g(x) \cdot f_{X|Y}(x|y)$ mit $y \in T_Y$

$E(g(X)) = \sum\limits_{y \in T_Y} E(g(X)|Y = y) \cdot f_Y(y)$

diskrete Randdichte:
$f_X(x) = P(X = x) = \sum\limits_{y \in T_Y} f_{X|Y}(x|y) \cdot f_Y(y) = \sum\limits_{y \in T_Y} f_{X,Y}(x, y)$ mit $x \in T_X$

bedingte Dichte:
$$f_{X|Y}(x|y) = \begin{cases} \frac{f_{(X,Y)}(x,y)}{f_Y(y)} & \text{falls } f_Y(y) > 0 \\ f_X(x) & \text{falls } f_Y(y) = 0 \end{cases}$$

bedingter Erwartungswert:
$E(g(X)|Y = y) = \int\limits_{-\infty}^{\infty} g(x) \cdot f_{X|Y}(x|y) \, dx$

$E(g(X)) = \int\limits_{-\infty}^{\infty} E(g(X)|Y = y) \cdot f_Y(y) \, dy$

bedingte Randdichte:
$f_X(x) = \int\limits_{-\infty}^{\infty} f_{X|Y}(x|y) \cdot f_Y(y) \, dy = \int\limits_{-\infty}^{\infty} f_{(X,Y)}(x, y) \, dy$

Transformation von Zufallsvektoren: $\mathbf{M} = \{\mathbf{x} \in \mathbb{R}^n \mid f_{\mathbf{X}}(\mathbf{x}) > \mathbf{0}\} \subseteq \mathbb{R}^n$, offen
Transformation T: $\mathbf{M} \to \mathbb{R}^n$, injektiv (umkehrbar), mit $\left| \left(\frac{\partial T_i(x_1, \ldots, x_n)}{\partial x_j} \right)_{1 \leq i,j \leq n} \right| > 0 \; \forall \, \mathbf{x}^T \in \mathbf{M}$

X ist absolut-stetig, $\mathbf{Y} = T(\mathbf{X})$

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\left| \left(\frac{\partial T_i(\mathbf{x})}{\partial x_j} \right)_{\mathbf{x} = T^{-1}(\mathbf{y})} \right|} \cdot f_{\mathbf{X}}(T^{-1}(\mathbf{y})) = \left| \left(\frac{\partial T_i^{-1}(\mathbf{y})}{\partial y_j} \right) \right| \cdot f_{\mathbf{X}}(T^{-1}(\mathbf{y})) \; \forall \, \mathbf{y} \in T(\mathbf{M})$$

Darauf achten, welche Determinante (von T oder T^{-1}) leichter zu berechnen ist!
Jacobi-Matrix: i senkrecht, j waagerecht

bei nicht-injektiver Transformation T:
Partition $I_1, \ldots, I_K$ von $\mathbb{R}$, so dass die Abbildungen T_k: $I_k \to \mathbb{R}\ \forall x \in I_k$ jeweils injektiv und stetig differenzierbar sind.

$f_Y(y) = \sum\limits_{k=1}^{K} \left| \frac{dT_k^{-1}(y)}{dy} \right| \cdot f_X(T_k^{-1}(y))$

Summe von Zufallsvariablen: $\mathbf{X} = (X_1, X_2)^T$ mit $f_{\mathbf{X}}(x_1, x_2)$, $Y = X_1 + X_2$
$f_Y(y) = \int\limits_{-\infty}^{\infty} f_{\mathbf{X}}(t, y-t)\, dt$ (Faltung der Verteilungsdichten)

und wenn X_1 und X_2 stochastisch unabhängig sind: $f_Y(y) = \int\limits_{-\infty}^{\infty} f_{X_1}(t) \cdot f_{X_2}(y-t)\, dt$

Faltungsstabilität der Γ- und Normalverteilungen:

- $X_1 \sim \Gamma(\alpha, \lambda), X_2 \sim \Gamma(\beta, \lambda)$: $X_1 + X_2 \sim \Gamma(\alpha + \beta, \lambda)$
- $X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2)$: $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Mehrwegeausbreitung:
empfangenes Signal: $r(t) = \underbrace{\mathrm{e}^{j2\pi ft}}_{s(t)} \cdot (X(t) + j \cdot Y(t))$

Empfangsleistung: $|R|^2$, mittlere Empfangsleistung: $E(|R|^2)$

$X(t), Y(t) \sim \mathrm{N}(0, \frac{\sigma}{2})$ unkorreliert, stochastisch unabhängig und gleichverteilt

Signal zufällig gedämpft durch $R = \sqrt{X^2(t) + Y^2(t)}$
Signal zufällig phasenverschoben durch $\Phi = \measuredangle(X(t) + j \cdot Y(t))$ bei der Zeit t, fest

«Als Fading (Schwund) bezeichnet man durch Interferenz oder Abschattung verursachte Schwankungen der Empfangsfeldstärke bei Funkübertragungen.»

Rayleigh-Fading-Modell:
keine direkte Sichtverbindung, daher $E(X) = E(Y) = 0$.
$R \sim \mathrm{Ray}(\tau^2)$, $\Phi \sim \mathrm{R}(0, 2\pi)$ mit $\tau^2 = \frac{\sigma^2}{2}$
Empfangsleistung: $Z \sim \mathrm{Exp}(\frac{1}{\sigma^2})$ mit der erwarteten Empfangsleistung σ^2

Rice-Fading-Modell:
Zufällige, schnelle Fading-Effekte, wenn neben der Mehrwegeausbreitung auch eine direkte Sichtverbindung zwischen Sender und Empfänger besteht.
$X \sim \mathrm{N}(\mu_1, \sigma^2)$, $Y \sim \mathrm{N}(\mu_2, \sigma^2)$ dann $R \sim \mathrm{Rice}(\sqrt{\mu_1^2 + \mu_2^2}, \sigma^2)$

Lognormal-Fading-Modell:
Langsame Fading-Effekte durch die Bewegung einer Mobilstation. $\mathrm{LogN}(\mu, \sigma^2)$

$|z|^2 = z \cdot z^* = a^2 + b^2$ mit $z = a + jb \in \mathbb{C}$

n-dimensionale komplexe Normalverteilung: $\mathbf{X} \sim \mathrm{CN}$
Der Zufallsvektor $\mathbf{X} \sim \mathbf{U} + j \cdot \mathbf{V} \in \mathbb{C}^n$ ist n-dimensional komplex normalverteilt,
wenn $\begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \in \mathbb{R}^{2n}$ $2n$-dimensional normalverteilt ist.

zirkulär symmetrisch komplex normalverteilt: $\mathbf{X} \sim \mathrm{SCN}(\boldsymbol{\mu}, \mathbf{Q})$
Eindeutig durch $E(\mathbf{X}) = \boldsymbol{\mu}$ und die Kovarianzmatrix $\mathbf{Q}$ bestimmt.
Ein komplexer Zufallsvektor, der gegenüber Drehungen im Komplexen invariant ist, d.h.

$(\mathbf{X}-\boldsymbol{\mu})\mathrm{e}^{j\vartheta} \sim \mathbf{X}-\boldsymbol{\mu}\ \forall\,\vartheta \in [0,2\pi)$

oder
hermitesche Matrix $\mathbf{Q} \in \mathbb{C}^{n\times n}$, so dass $Cov\begin{pmatrix}\mathbf{U}\\ \mathbf{V}\end{pmatrix} = \frac{1}{2}\begin{pmatrix}\mathrm{Re}(\mathbf{Q}) & -\mathrm{Im}(\mathbf{Q})\\ \mathrm{Im}(\mathbf{Q}) & \mathrm{Re}(\mathbf{Q})\end{pmatrix}$

$\mathbf{Q} = \mathbf{Q}^*$, also $q_{ji} = q_{ij}^*$

$\mathbf{Q} = E((\mathbf{X}-E(\mathbf{X}))\cdot(\mathbf{X}-E(\mathbf{X}))^*)$

$\mathbf{Q}$ nicht-negativ definit, d.h. $\mathbf{x}^*\cdot\mathbf{Q}\cdot\mathbf{x} \geq 0\ \forall\,\mathbf{x}\in\mathbb{C}^n$

Dichte:
$\mathbf{X} \sim \mathrm{SCN}(\boldsymbol{\mu},\mathbf{Q})$, $\boldsymbol{\mu}\in\mathbb{C}^n$, $\mathbf{Q}\in\mathbb{C}^{n\times n}$ regulär, hermitesch, nicht-negativ definit
dann $f_{\mathbf{X}}(\mathbf{x}) = (\det(\pi\cdot\mathbf{Q}))^{-1}\cdot\exp(-(\mathbf{x}-\boldsymbol{\mu})^*\cdot\mathbf{Q}^{-1}\cdot(\mathbf{x}-\boldsymbol{\mu}))$

$\mathbf{X} \sim \mathrm{SCN}(\boldsymbol{\mu},\mathbf{Q}), \mathbf{A}\in\mathbb{C}^{m\times n} \Rightarrow \mathbf{A}\cdot\mathbf{X} \sim \mathrm{SCN}(\mathbf{A}\cdot\boldsymbol{\mu}, \mathbf{A}\cdot\mathbf{Q}\cdot\mathbf{A}^*)$

$\mathbf{X} \sim \mathrm{SCN}(\boldsymbol{\mu}_1,\mathbf{Q}_1)$ und $\mathbf{Y} \sim \mathrm{SCN}(\boldsymbol{\mu}_2,\mathbf{Q}_2)$ stochastisch unabhängig $\Rightarrow \mathbf{X}+\mathbf{Y} \sim \mathrm{SCN}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2, \mathbf{Q}_1+\mathbf{Q}_2)$

2.1 Stochastische Prozesse

Eine Familie von reellen oder komplexen Zufallsvariablen $\{X(t) \mid t\in T\subseteq\mathbb{R}\}$ aus einem gemeinsamen Wahrscheinlichkeitsraum.

$X(t,\omega)$: $T\times\Omega\to\mathbb{R}$ oder $\mathbb{C}$, Zufallsvariable für festes $t\in T$ *oder* Funktion von t für festes $\omega\in\Omega$

gemeinsame Verteilungsfunktion: von $(X(t_1),\ldots,X(t_n))^T$
$F_{(X(t_1),\ldots,X(t_n))}(x_1,\ldots,x_n)$ für n Zeitpunkte $t_1,\ldots,t_n\in T$

Erwartungswertfunktion: $\mu_X(t) = E(X(t))$
falls $X(t) = U(t)+jV(t) \Rightarrow E(X(t)) = E(U(t))+jE(V(t))$

Autokorrelationsfunktion: $R_{XX}(t_1,t_2) = E(X(t_1)\cdot X^*(t_2))$
$R_{XX}(t_1,t_2) = R_{XX}^*(t_2,t_1)$
$R_{XX}(t,t) = E(|X(t)|^2) \geq 0$
R_{XX} nicht-negativ definit, dann ist die Matrix $(R_{XX}(t_i,t_j))_{1\leq i,j\leq n}$ nicht-negativ definit

Autokovarianzfunktion: $C_{XX}(t_1,t_2) = R_{XX}(t_1,t_2) - \mu_X(t_1)\cdot\mu_X^*(t_2)$

Gaußprozess: $\{X(t) \mid t\in T\subseteq\mathbb{R}\}$
$(X(t_1),\ldots,X(t_n))^T$ ist n-dimensional normalverteilt $\forall\,t_1,\ldots,t_n\in T$

strikt stationärer stochastischer Prozess:
Alle endlich-dimensionalen Randverteilungen sind zeitinvariant.
$F_{(X(t_1),\ldots,X(t_n))} = F_{(X(t_1+s),\ldots,X(t_n+s))}\ \forall\,t_1,\ldots,t_n,t_1+s,\ldots,t_n+s\in T$

strikt stationär $\Rightarrow$ schwach stationär

schwach stationärer stochastischer Prozess:
Der Erwartungswert ist konstant und damit von t unabhängig.
Der Wert der Autokorrelationsfunktion ist nur von der Differenz t_1-t_2 abhängig.
$E(X(t)) = \mu\ \forall\,t\in T$
$E(X(t_1)\cdot X^*(t_2)) = R_{XX}(t_2-t_1) = R_{XX}(t_1,t_1+t_{diff})\ \forall\,t_1,t_2\in T$ und $\forall\,t_{diff}=t_2-t_1$

Lineare zeitinvariante (LTI) Systeme: $y(t) = T(x(t)) = x(t) * h(t)$

$$\xrightarrow{\;x(t)\;} \boxed{h(t)} \xrightarrow{\;y(t)\;}$$

Lineare zeitinvariante Filter mit der Impulsantwort $h(t)$ werden eingesetzt um Signale zu glätten und bestimmte Frequenzen herauszuschneiden.

Faltung: $y(t) = \int\limits_{-\infty}^{\infty} h(u) \cdot x(t-u)\,du = \int\limits_{-\infty}^{\infty} h(t-u) \cdot x(u)\,du = h(t) * x(t)$

Ein schwach stationärer Prozess $\{X(t)\}$ als Eingabe für ein LTI-System, dann $Y(t) = \int\limits_{-\infty}^{\infty} h(u) \cdot X(t-u)\,du$

Dann ist $\{Y(t)\}$ auch schwach stationär mit

$$\mu_Y(t) = E(Y(t)) = \mu_X(t) \cdot \underbrace{\int_{-\infty}^{\infty} h(u)\,du}_{\text{oft } =1}$$

$$R_{YY}(t) = \int\limits_{-\infty}^{\infty} h(u) \cdot R_{XY}(t-u)\,du = h(t) * R_{XY}(t) = h(t) * \underbrace{h^*(-t) * R_{XX}(t)}_{R_{XY}(t)}$$

$$R_{XY}(t) = \int\limits_{-\infty}^{\infty} h^*(v) \cdot R_{XX}(t+v)\,dv = h^*(-t) * R_{XX}(t)$$

Kreuzkorrelationsfunktion zweier stochastischer Prozesse:
$R_{XY}(t) = E(X(t_1) \cdot Y^*(t_2))$

Energiespektrum: $|G(f)|^2$
Energie $E = \int\limits_{-\infty}^{\infty} |g(t)|^2\,dt = \int\limits_{-\infty}^{\infty} |G(f)|^2\,df$ (Satz von Parseval)

näherungsweise Signalenergie im Band $[f, f+\Delta f]$: $|G(f)|^2 \cdot \Delta f$

zu erwartende Momentanleistung zum Zeitpunkt t: $E(|X(t)|^2) = E(X(t) \cdot X^*(t)) = R_{XX}(0) = \int\limits_{-\infty}^{\infty} S_{XX}(f)\,df$

$S_{XX}(f) = \int\limits_{-\infty}^{\infty} R_{XX}(t) \cdot \mathrm{e}^{-j2\pi ft}\,dt \qquad R_{XX}(t) = \int\limits_{-\infty}^{\infty} S_{XX}(f) \cdot \mathrm{e}^{j \cdot 2\pi ft}\,df$ (Fourier-Transformation)

Leistungsdichtespektrum:
schwach stationärer stochastischer Prozess $\{X(t)\}$ mit Autokorrelationsfunktion $R_{XX}(t) \;\circ\!\!-\!\!\bullet\; S_{XX}(f)$

$S_{XX}(f) = S_{XX}(-f)$ für $R_{XX}(t) \in \mathbb{R}$

$S_{XX}(f) \geq 0 \in \mathbb{R}\ \forall f \in \mathbb{R}$

$S_{YY}(f) = |H(f)|^2 \cdot S_{XX}(f)$ bei LTI-Filterung

erwartete Leistung im Band $[f, f+\Delta f]$: $S_{XX}(f) \cdot \Delta f$

Weißes Rauschen: schwach stationärer stochastischer Prozess $\{W(t)\}$
Hat ein konstantes Leistungsdichtespektrum $S_{XX}(f)$, d.h. alle Frequenzen haben die gleiche, positive Energie, d.h. die Autokorrelationsfunktion $R_{XX}(t) = 0\ \forall t \neq 0$
Physikalisch ist dies nicht realisierbar, da dafür eine unendlich hohe Leistung benötigt werden würde: $\int_{-\infty}^{\infty} S_{WW}(f)\,df = \infty$

Dirac-Impuls $\delta(t)$ mit $\int\limits_{-\infty}^{\infty} \delta(t)\,dt = 1$ und $\int\limits_{-\infty}^{\infty} \delta(t) \cdot f(t)\,dt = f(0)$

Leistungsdichtespektrum $S_{WW}(f) = \frac{N_0}{2} \; \forall f \in \mathbb{R}$

Autokorrelationsfunktion $R_{WW}(t) = \frac{N_0}{2}\delta(t)$

Gaußsches Weißes Rauschen:
Der Zufallsvektor $(W(t_1), \ldots, W(t_n))^T$ ist gemeinsam (n-dimensional) normalverteilt $\forall\, t_1, \ldots, t_n$

Farbiges Rauschen: $\{N(t)\}$ mit $N(t) = h(t) * W(t)$, also gefiltertes Weißes Rauschen

Leistungsdichtespektrum $S_{NN}(f) = |H(f)|^2 \cdot S_{WW}(f) = |H(f)|^2 \, \frac{N_0}{2}$

matched Filter: *signalangepasstes Filter*
empfangen: $X(t) = s(t) + W(t)$ mit dem deterministischen Signal $s(t)$

gefiltert mit dem LTI-System $h(t)$: $Y(t) = h(t) * X(t) = h(t) * s(t) + h(t) * W(t) = y_s(t) + \underbrace{N(t)}_{\text{farbiges Rauschen}}$

Der Signal-zu-Rausch Abstand $\text{SNR} = \frac{\text{Signalleistung}}{\text{Rauschleistung}} = \frac{|y_s(t_0)|^2}{E(|N(t)|^2)}$

wird durch $h(t) = s(t_0 - t) \quad \circ\!\!-\!\!\bullet \quad H(f) = S^*(f) \cdot \mathrm{e}^{-j \cdot 2\pi f t_0}$ maximiert.

3 Elemente der Informationstheorie

Informationsgehalt eines Ereignisses: $I_X(x_j) = -\log(P(X = x_j))$
Eine Maßzahl für die Unbestimmtheit des Ausgangs des Zufallsexperimentes oder eine Maßzahl für den Informationsgewinn nach dem Ausgang des Zufallsexperimentes.

Entropie: erwarteter (mittlerer) Informationsgehalt aller Ereignisse
$H(X) = -\sum_{j=1}^{m} P(X = x_j) \cdot \log(P(X = x_j)) = E(-\log(X)) \geq 0$

gemeinsame Entropie:
$H(X, Y) = -\sum_{i,j} P(X = x_i, Y = y_j) \cdot \log(P(X = x_i, Y = y_j))$
$H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$

bedingte Entropie: mittlere Entropie bezüglich der bedingten Verteilung
$H(X|Y) = -\sum_{j} P(Y = y_j) \cdot \sum_{i} P(X = x_i|Y = y_j) \cdot \log(P(X = x_i|Y = y_j))$
$= -\sum_{i,j} P(X = x_i, Y = y_j) \cdot \log(P(X = x_i|Y = y_j)) \leq H(X)$

m *Trägerpunkte/Buchstaben*
$H(X) = 0$: X ist einpunktverteilt ($P(X = x_i) = 1$ für ein x_i)
$H(X) = \log(m)$: X ist gleichverteilt ($P(X = x_i) = \frac{1}{m} \forall\, i$)
$H(X|Y) = 0$: X ist total abhängig von Y ($P(X = x_i|Y = y_j) = 1 \;\forall\, i, j$ mit $P(X = x_i, Y = y_j) > 0$)
$H(X|Y) = H(X)$: X und Y sind stochastisch unabhängig
$H(X, Y) = H(X)$: Y ist total abhängig von X
$H(X, Y) = H(X) + H(Y)$: X und Y sind stochastisch unabhängig

Transinformation:
Misst, wie stark die Unbestimmtheit von X im Mittel sinkt, wenn das Ergebnis von Y bekannt ist.
Auch ein Maß für die Güte der Informationsübertragung im Kanal.
$I(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X, Y) \geq 0$
$I(X, (Y, Z)) = I(X, Z) + I(X, (Y|Z))$

Kullback-Leibler-Distanz: *(relative Entropie)* $\mathbf{p}, \mathbf{q}$ diskrete Verteilungen

$D(\mathbf{p}||\mathbf{q}) = \sum\limits_{i=1}^{k} p_i \cdot \log(\frac{p_i}{q_i}) \geq 0$

Quellencodierung:
Wird benutzt um die Symbole einer Quelle möglichst dicht zu komprimieren, wobei die ursprünglichen Symbole eindeutig rekonstruiert werden können, auch Konkatenationen dieser.
z.B. Quellalphabet $\mathcal{X} = \{A, \ldots, Z\}$ mit dem Kodealphabet $\mathcal{Y} = \{0, 1\}$
Eine wichtige Klasse eindeutig dekodierbarer Kodes: *präfixfreie* Kodes (kein Kodewort ist Präfix eines anderen Kodewortes, z.B. Huffman-Kode für die optimale, kürzeste Kodelänge)

erwartete Kodewortlänge: $\overline{n} = \sum\limits_{j=1}^{m} n_j \cdot P(X = x_j)$

Noiseless Coding Theorem:
für alle eindeutig dekodierbaren Kodes g gilt: $\frac{H(X)}{\log(d)} \leq \overline{n}(g)$
es existiert ein präfixfreier Kode g mit: $\overline{n}(g) \leq \frac{H(X)}{\log(d)} + 1$

Kanalkapazität: $C = \max\limits_{(p_1,\ldots,p_m)} \{I(X,Y)\}$ bei gegebenem $P(Y|X)$

Binärer symmetrischer Kanal: $C = 1 + (1-\varepsilon) \cdot \mathrm{ld}(1-\varepsilon) + \varepsilon \cdot \mathrm{ld}(\varepsilon)$ bei Gleichverteilung der Eingabe

ME-Dekodierung: *(minimum error)*
Es wird die Wahrscheinlichkeit maximiert, $\mathbf{c}_j$ gesendet zu haben, wenn $\mathbf{b}_N$ empfangen wird.
$\mathbf{c}_j = h_N(\mathbf{b}_N) \Rightarrow P(\mathbf{X}_N = \mathbf{c}_j | \mathbf{Y}_N = \mathbf{b}_N) \geq P(\mathbf{X}_N = \mathbf{c}_i | \mathbf{Y}_N = \mathbf{b}_N) \;\forall\, i = 1, \ldots, M$

Eingabekode: $\mathcal{C}_N = \{\mathbf{c}_1, \ldots, \mathbf{c}_M\} \subseteq \mathcal{X}^N$

Dekodierregel: $h_N \colon \mathcal{Y}^N \to \mathcal{C}_N \colon \mathbf{b}_N \mapsto h_N(\mathbf{b}_N)$

ML-Dekodierung: *(maximum likelihood)*
Es wird die Wahrscheinlichkeit maximiert, $\mathbf{b}_N$ zu empfangen, wenn $\mathbf{c}_j$ gesendet wurde.
$\mathbf{c}_j = h_N(\mathbf{b}_N) \Rightarrow P(\mathbf{Y}_N = \mathbf{b}_N | \mathbf{X}_N = \mathbf{c}_j) \geq P(\mathbf{Y}_N = \mathbf{b}_N | \mathbf{X}_N = \mathbf{c}_i) \;\forall\, i = 1, \ldots, M$

Noisy-channel Coding Theorem, Shannonscher Fundamentalsatz:
diskreter gedächtnisloser Kanal mit $0 < R < C$. Es gelte $M_N < 2^{R \cdot N}$. Dann existiert eine Folge von (M_N, N)-Kodes mit $\hat{e}(\mathcal{C}_N) \to 0$ für $N \to \infty$. (M_N Kodewörter, Kodewortlänge N)

$\frac{\log(M_N)}{N} < R < C$

Ist die Quellenrate R kleiner als die Kapazität C, so existieren für große Blocklängen Kodes mit beliebig kleiner Fehlerwahrscheinlichkeit. Dies führt aber zu aufwändigen Kodier- und Dekodieralgorithmen.

differentielle Entropie:
$H(X) = -\int\limits_{-\infty}^{\infty} f_X(x) \cdot \log(f_X(x))\, dx = E(-\log(f_X(x))$

gemeinsame differentielle Entropie:
$H(X_1, \ldots, X_n) = -\int\limits_{-\infty}^{\infty} \ldots \int\limits_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, \ldots, x_n) \cdot \log(f_{\mathbf{X}}(x_1, \ldots, x_n))\, dx_1 \ldots dx_n$

bedingte differentielle Entropie:
$H(X|Y) = -\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} f_{\mathbf{X}}(x,y) \cdot \log(f_{\mathbf{X}}(x|y))\, dx\, dy$

n-dimensionale Normalverteilung: $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \mathbf{C})$ mit $H(\mathbf{X}) = \frac{1}{2} \cdot \ln((2\pi \mathrm{e})^n \cdot |\det(\mathbf{C})|)$
hat die größte Entropie unter allen absolut-stetigen Verteilungen

Kullback-Leibler-Distanz:

$D(f||g) = \int\limits_{-\infty}^{\infty} f_{\mathbf{X}}(\boldsymbol{x}) \cdot \log(\frac{f_{\mathbf{X}}(\boldsymbol{x})}{g_{\mathbf{X}}(\boldsymbol{x})})\, d\boldsymbol{x}$

$I(X,Y) = D(f_{(X,Y)}(x,y)||f_X(x) \cdot f_Y(y)) \geq 0$

$\mathbf{X}$ absolut-stetiger Zufallsvektor, $\mathbf{A} \in \mathbb{R}^{n\times n}$ regulär, $\mathbf{b} \in \mathbb{R}^n$
$H(\mathbf{A} \cdot \mathbf{X} + \mathbf{b}) = H(\mathbf{X}) + \log(|\det(\mathbf{A}|)$

4 Kanäle und ihre Kapazität