Cheatsheet WuS

Nicolas Wehrli

June 2023

1 Grundbegriffe

1.1 Wahrscheinlichkeitsraum

Axiome von Kolmogorov

Das Tuple $(\Omega, \mathcal{A}, \mathbb{P})$ ist ein **Wahrscheinlichkeitsraum** mit

I. **Grundraum** Ω mit $\Omega \neq \emptyset$, wobei $\omega \in \Omega$ ein Elementarereignis ist.

II. **σ-Algebra** $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ wobei gilt:
   1. $\Omega \in \mathcal{A}$
   2. $A \in \mathcal{A} \implies A^{\complement} \in \mathcal{A}$
   3. $A_1, A_2, \dots \in \mathcal{A} \implies \bigcup_i A_i \in \mathcal{A}$

III. **Wahrscheinlichkeitsmass** $\mathbb{P}$ auf $(\Omega, \mathcal{A})$ ist eine Abbildung $\mathbb{P} : \mathcal{A} \mapsto [0,1]$, wobei gilt:
   1. $\mathbb{P}(\Omega) = 1$
   2. $A_1, A_2, \dots \in \mathcal{A}, \forall i \neq j : A_i \cap A_j = \emptyset$
      $\implies \mathbb{P}(\bigcup_i A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

De-Morgan
Sei $(A_i)_{i\geq 1}$ eine Folge von beliebigen Mengen. Dann gilt

$$\left(\bigcup_{i=1}^{\infty} A_i\right)^{\complement} = \bigcap_{i=1}^{\infty} (A_i)^{\complement}$$

Daraus folgt
1. $A_1, A_2, \dots \in \mathcal{A} \implies \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$
2. $A, B \in \mathcal{A} \implies (A \cup B), (A \cap B) \in \mathcal{A}$

und für $A, B \in \mathcal{A}$
1. $\mathbb{P}(A^{\complement}) = 1 - \mathbb{P}(A)$
2. $A \subseteq B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Sei $A_1, A_2, \dots \in \mathcal{A}$, dann gilt:
Union Bound

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

Siebformel

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} \sum_{1\leq i_1 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k})$$

Atome

Sei Ω nicht leer und diskret. Sei $\mathcal{F}$ eine beliebige σ-Algebra über Ω.
Eine nichtleere Menge $A \in \mathcal{F}$ heisst **atomare** Mengee von $\mathcal{F}$ falls für alle $B \in \mathcal{F}$ gilt:

$$B \subseteq A \implies B = \emptyset \lor B = A$$

(Intuitiv: A ist die kleinste nichtleere Menge bezüglich der Inklusion in $\mathcal{F}$)
Die Menge der atomaren Mengen von $\mathcal{F}$ bezeichnen wir mit $\text{Atom}(\mathcal{F})$.
Jedes Element von $\mathcal{F}$ lässt sich als abzählbare Vereinigung von Elementen aus $\text{Atom}(\mathcal{F})$ schreiben.

1.2 Bedingte Wahrscheinlichkeiten

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum.

Bedingte Wahrscheinlichkeit

Sei $A, B \in \mathcal{A}$ und $\mathbb{P}(B) > 0$, dann ist die **bedingte Wahrscheinlichkeit von A gegeben B**

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

Satz der totalen Wahrscheinlichkeit
Sei $(B_i)_{i\in I}$ eine Partition von Ω. Dann gilt für jedes beliebige $A \in \mathcal{A}$

$$\mathbb{P}(A) = \sum_{i:\ \mathbb{P}(B_i)>0} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

Satz von Bayes
Aus der Definition der bedingten W'keit folgt sofort die Bayessche Formel, welche den Zusammenhang zwischen $\mathbb{P}(A|B)$ und $\mathbb{P}(B|A)$ beschreibt:

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

Mit dem **Satz der totalen W'keit** können wir $\mathbb{P}(A)$ umschreiben und kommen auf folgende Form:
Sei $(B_i)_{i\in I}$ eine **Partition** von Ω. Dann gilt für jedes beliebige $A \in \mathcal{A}, \mathbb{P}(A) > 0$

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)}{\sum_{j:\ \mathbb{P}(B_j)>0} \mathbb{P}(A|B_j) \cdot \mathbb{P}(B_j)}$$

Intuition Bayessche Statistik
In dieser Form würde man A als das **eingetretene Ereignis** und die B_i als die verschiedene **Hypothesen** verstehen.
In der Bayesschen Statistik versucht man die Hypothese zu finden, so dass $\mathbb{P}(B_i|A)$ **maximiert** wird.
(Wurde in der Vorlesung nicht weiter behandelt)

1.3 Unabhängigkeit

Unabhängigkeit von zwei Ereignissen

Zwei Ereignisse $A, B \in \mathcal{A}$ heissen **unabhängig**, wenn

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

- $\mathbb{P}(A) \in \{0,1\} \implies A$ zu jedem Ereignis unabhängig
- A zu sich selbst unabhängig $\implies \mathbb{P}(A) \in \{0,1\}$
- A, B unabhängig $\implies A, B^{\complement}$ unabhängig

Wenn $\mathbb{P}(A) > 0, \mathbb{P}(B) > 0$ gilt:
A, B unabhängig $\iff \mathbb{P}(A|B) = \mathbb{P}(A) \iff \mathbb{P}(B|A) = \mathbb{P}(B)$

Wir können die Definition der Unabhängigkeit auf beliebige Mengen von Ereignissen erweitern.

Allgemeine Unabhängigkeit

Eine Kollektion von Ereignissen $(A_i; i \in I)$ heisst **(stochastisch) unabhängig**, wenn

$$J \subseteq I \text{ endlich} \implies \mathbb{P}\left(\bigcap_{i\in J} A_i\right) = \prod_{i\in J} \mathbb{P}(A_i)$$

2 Zufallsvariablen

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum.

Zufallsvariable

Eine (reellwertige) **Zufallsvariable** auf Ω ist eine messbare Funktion $X : \Omega \to \mathbb{R}$.

$$X : \Omega \to \mathbb{R} \textbf{ messbar} \iff \forall x \in \mathbb{R} : X^{-1}(\{x\}) \in \mathcal{A}$$

Die Eigenschaft **messbar** ist bezüglich dem Wahrscheinlichkeitsmass $\mathbb{P}$ relevant (i.e. dann ist $\mathbb{P}(X = x) := \mathbb{P}(\{\omega \in \Omega \mid X(\omega) = x\})$ wohldefiniert).

Diese Definition von **messbar** ist für diskrete Ω äquivalent zu derjenigen der Vorlesung, die die rechte Seite vom ' $\iff$ ' für alle abgeschlossenen Teilmengen $B \subset \mathbb{R}$ fordert.
Für die Messbarkeit von X ist nur $X(\Omega) \subseteq \mathbb{R}$ entscheidend und jede Teilmenge $A \subseteq X(\Omega)$ ist abzählbar (da Ω abzählbar). Somit kann $X^{-1}(A)$ als abzählbare Vereinigung von $\bigcup_{x\in A} X^{-1}(\{x\})$ geschrieben werden.
($\implies X^{-1}(A) \in \mathcal{A}$ per Def. σ-Algebra)

2.1 Verteilungsfunktion

Die **Verteilungsfunktion** ist die Abbildung $F_X : \mathbb{R} \to [0,1]$ definiert durch:

$$F_X(t) := \mathbb{P}(X \leq t), \forall t \in \mathbb{R}$$

Die Funktion erfüllt folgende Eigenschaften:

1. F_X ist monoton wachsend
2. F_X ist rechtsstetig, i.e. $\lim_{h\downarrow 0} F_X(x+h) = F_X(x)$
3. $\lim_{x\to-\infty} F_X(x) = 0$ und $\lim_{x\to\infty} F_X(x) = 1$
4. $\forall a, b \in \mathbb{R}, a < b : \mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$

Linksstetigkeit
Die Verteilungsfunktion ist nicht immer linksstetig.
Sei $F_X(a-) := \lim_{h\downarrow 0} F_X(a-h)$ für $a \in \mathbb{R}$ beliebig.
Dann gilt:

$$\mathbb{P}(X = a) = F_X(a) - F_X(a-)$$

Intuitiv folgt daraus

- Wenn F_X in einem Punkt $a \in \mathbb{R}$ nicht stetig ist, dann ist die "Sprunghöhe" $F_X(a) - F_X(a-)$ gleich der Wahrscheinlichkeit $\mathbb{P}(X = a)$.
- Falls F_X stetig in einem Punkt $a \in \mathbb{R}$, dann gilt $\mathbb{P}(X = a) = 0$.

Unabhängigkeit von Zufallsvariablen

Seien $X_1, ..., X_n$ Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. Dann heissen $X_1, ..., X_n$ **unabhängig**, falls

$$\forall x_1, ..., x_n \in \mathbb{R} :$$
$$\mathbb{P}(X_1 \leq x_1, ..., X_n \leq x_n) = \mathbb{P}(X_1 \leq x_1) \cdot ... \cdot \mathbb{P}(X_n \leq x_n).$$

Unendlich viele Bernoulli-Experimente

TBD

2.2 Diskrete Zufallsvariablen

Sei $A \in \mathcal{F}$ ein Ereignis.
Wir sagen A tritt **fast sicher (f.s.)** ein, falls $\mathbb{P}(A) = 1$.

Seien $X, Y : \Omega \to \mathbb{R}$ Zufallsvariablen:
$X \leq Y$ f.s. $\iff \mathbb{P}(X \leq Y) = 1$

Diskrete Zufallsvariable

Eine Zufallsvariable $X : \Omega \to \mathbb{R}$ heisst **diskret**, falls eine endliche oder abzählbare Menge $W \subset \mathbb{R}$ existiert, sodass

$$\mathbb{P}(X \in W) = 1$$

Falls Ω endlich oder abzählbar ist, dann ist X immer diskret.

Die **Verteilungsfunktion** einer diskreten ZV X:

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{y\in W} p(y) \cdot \mathbb{1}_{y\leq x}$$

Die **Gewichtsfunktion** einer diskreten ZV X:

$$\forall x \in X(\Omega) : p(x) = \mathbb{P}(X = x) \text{ wobei } \sum_{x\in X(\Omega)} p(x) = 1$$

2.3 Diskrete Verteilungen

Bernoulli-Verteilung ($X \sim \text{Ber}(p)$):
$X(\Omega) = \{0, 1\}$ und die Gewichtsfunktion ist definiert durch

$$p(1) := \mathbb{P}(X = 1) = p \text{ und } p(0) := \mathbb{P}(X = 0) = 1 - p.$$

Binomialverteilung ($X \sim \text{Bin}(n, p)$):
Wiederholung von n unabhängigen Bernoulli-Experimenten mit gleichem Parameter p.

$$p(k) := \mathbb{P}(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \quad \forall k \in \{0, 1, \ldots, n\}$$

Geometrische Verteilung ($X \sim \text{Geo}(p)$):
Warten auf den 1-ten Erfolg.

$$p(k) := \mathbb{P}(X = k) = (1-p)^{k-1} \cdot p \quad \forall k \in \mathbb{N} \setminus \{0\}$$

Poisson-Verteilung ($X \sim \text{Poisson}(\lambda)$):
Grenzwert der Binomialverteilung für grosse n und kleine p.

$$p(k) := \mathbb{P}(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda} \quad \forall k \in \mathbb{N}_0, \lambda > 0$$

1. Für $X_n \sim \text{Bin}(n, \frac{\lambda}{n})$ gilt $\lim_{n\to\infty} \mathbb{P}(X_n = k) = \mathbb{P}(Y = k)$ wobei $Y \sim \text{Poisson}(\lambda)$.
2. Seien $X_1 \sim \text{Poisson}(\lambda_1)$ und $X_2 \sim \text{Poisson}(\lambda_2)$ unabhängig. Dann gilt $(X_1 + X_2) \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

2.4 Stetige Zufallsvariablen

Stetige Zufallsvariablen, Dichte

Eine Zufallsvariable $X : \Omega \to \mathbb{R}$ heisst **stetig**, wenn ihre Verteilungsfunktion F_X wie folgt geschrieben werden kann

$$F_X(a) = \int_{-\infty}^{a} f(x)\, dx = \text{ für alle } a \in \mathbb{R}.$$

wobei $f : \mathbb{R} \to \mathbb{R}^+$ eine nicht-negative Funktion ist. f wird dann als **Dichte** von X benannt.
Wenn $f : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ messbar ist, ist die Zufallsvariable X **absolut stetig**.

Intuition: $f(x)\, dx$ ist die Wahrscheinlichkeit, dass $X \in [x, x + dx]$.

Von F_X zu f:
Sei X eine Zufallsvariable mit stetiger Verteilungsfunktion F_X und F_X stückweise $\mathcal{C}^1$, d.h. es gibt $x_0 = -\infty < \ldots < x_{n-1} < x_n = +\infty$, sodass F_X auf jedem Intervall (x_i, x_{i+1}) Element von $\mathcal{C}^1$ ist.
Dann ist X eine stetige Zufallsvariable und die Dichte f kann wie folgt konstruiert werden:

$$\forall x \in (x_i, x_{i+1}) \quad f(x) = F'_X(x).$$

TODO: Beispielrechnung -Dichte finden

2.5 Stetige Verteilungen

Gleichverteilung ($X \sim \mathcal{U}([a, b])$):
Die Dichte ist auf dem Intervall $[a, b]$ gleich.

$$f_{a,b}(x) = \begin{cases} 0 & x \notin [a, b] \\ \frac{1}{b-a} & x \in [a, b] \end{cases}$$

Exponentialverteilung ($T \sim \text{Exp}(\lambda)$):
Lebensdauer oder Wartezeit eines allg. Ereignisses (Stetiges Äquivalent zur Geometrischen Verteilung).

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

Normalverteilung ($X \sim \mathcal{N}(m, \sigma^2)$):
Häufig verwendete Verteilung. Undefiniert für $\sigma = 0$.

$$f_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

1. Seien $X_1, \ldots, X_n$ **unabhängige** normalverteilte ZV mit Parametern $(m_1, \sigma_1^2), \ldots, (m_n, \sigma_n^2)$, dann ist
$$Z = m_0 + \lambda_1 X_1 + \ldots + \lambda_n X_n$$
eine normalverteilte ZV mit Parametern $m = m_0 + \lambda_1 m_1 + \ldots + \lambda_n m_n$ und $\sigma^2 = \lambda_1^2 \sigma_1^2 + \ldots + \lambda_n^2 \sigma_n^2$.
2. Sei $Z \sim \mathcal{N}(0, 1)$ eine **standardnormalverteilte** Zufallsvariable. Dann gilt für $X \sim \mathcal{N}(m, \sigma^2)$
$$X = m + \sigma \cdot Z$$

3 Erwartungswert

Erwartungswert - Diskrete ZV

Sei $X : \Omega \to \mathbb{R}$ eine diskrete Zufallsvariable, $W_X := X(\Omega)$ und $\phi : \mathbb{R} \to \mathbb{R}$ eine Abbildung. Falls die Summe wohldefiniert ist, gilt:

$$\mathbb{E}(\phi(X)) := \sum_{x\in W_X} \phi(x) \cdot \mathbb{P}(X = x)$$

Wenn $X : \Omega \to \mathbb{N}_0$, kann man auch den Erwartungswert als

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} \mathbb{P}(X > n)$$

schreiben.

Erwartungswert - Stetige ZV

Sei $X:\Omega\to\mathbb{R}$ eine stetige Zufallsvariable mit Dichte f. Sei $\phi:\mathbb{R}\to\mathbb{R}$ eine Abbildung, sodass $\phi(X)$ eine Zufallsvariable ist. Dann gilt

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

solange das Integral wohldefiniert ist.
Sei X eine stetige ZV mit $X\geq 0$ f.s., dann gilt:

$$\mathbb{E}(X) = \int_0^{\infty} \mathbb{P}(X>x)\,dx$$

3.1 Rechnen mit Erwartungswerten

Linearität des Erwartungswertes:
Seien $X,Y:\Omega\to\mathbb{R}$ ZV mit $\lambda\in\mathbb{R}$, Falls die Erwartungswerte wohldefiniert sind, gilt:

$$\mathbb{E}(\lambda\cdot X+Y) = \lambda\cdot\mathbb{E}(X)+\mathbb{E}(Y)$$

Falls X,Y **unabhängig**, dann gilt auch:

$$\mathbb{E}(X\cdot Y) = \mathbb{E}(X)\cdot\mathbb{E}(Y)$$

Eine generellere Form wäre folgende Äquivalenz:
$X_1,X_2,...,X_n$ unabhängig
$\iff$
Für jede $\phi_1:\mathbb{R}\to\mathbb{R},\dots,\phi_n:\mathbb{R}\to\mathbb{R}$ stückweise stetig, beschränkt gilt

$$\mathbb{E}(\phi_1(X_1)\cdots\phi_n(X_n)) = \mathbb{E}(\phi_1(X_1))\cdots\mathbb{E}(\phi_n(X_n))$$

3.2 Ungleichungen

Monotonie
Seien X,Y ZV mit $X\leq Y$ f.s., dann gilt:

$$\mathbb{E}(X)\leq\mathbb{E}(Y)$$

Markov Ungleichung
Sei X eine ZV und ferner $g:X(\Omega)\to[0,+\infty)$ eine wachsende Funktion. Für jedes $c\in\mathbb{R}$ mit $g(c)>0$ gilt dann

$$\mathbb{P}(X\geq c)\leq\frac{\mathbb{E}(g(X))}{g(c)}$$

Einfache Version:
Sei X eine ZV mit $X\geq 0$ f.s., dann gilt für jedes $t>0$:

$$\mathbb{P}(X\geq t)\leq\frac{\mathbb{E}(X)}{t}$$

Chebyshev Ungleichung
Sei Y eine ZV mit endlicher Varianz. Für jedes $b>0$ gilt dann

$$\mathbb{P}(|Y-\mathbb{E}(Y)|\geq b)\leq\frac{\mathrm{Var}(Y)}{b^2}$$

Jensen Ungleichung
Sei X eine ZV und $\phi:\mathbb{R}\to\mathbb{R}$ eine konvexe Funktion, dann gilt:

$$\phi(\mathbb{E}(X))\leq\mathbb{E}(\phi(X))$$

3.3 Varianz

Varianz

Sei X eine ZV, sodass $\mathbb{E}(X^2)<\infty$. Die **Varianz** von X ist definiert durch

$$\mathrm{Var}(X) = \sigma_X^2 = \mathbb{E}((X-m)^2) = \mathbb{E}(X^2)-\mathbb{E}(X)^2$$

wobei $m=\mathbb{E}(X)$. Dabei wird σ_X als **Standardabweichung** von X bezeichnet und beschreibt den Erwartungswert für die Distanz von X zu $\mathbb{E}(X)$.

1. Sei X ein ZV, sodass $\mathbb{E}(X^2)<\infty$ und $\lambda\in\mathbb{R}$:
$$\mathrm{Var}(a\cdot X+b) = a^2\cdot\mathrm{Var}(X)$$
2. Seien $X_1,...,X_n$ paarweise unabhängig. Dann gilt
$$\mathrm{Var}(X_1+\ldots+X_n) = \mathrm{Var}(X_1)+\ldots+\mathrm{Var}(X_n)$$

Kovarianz

Seien X,Y ZV mit $\mathbb{E}(X^2)<\infty,\mathbb{E}(Y^2)<\infty$. Wir definieren die **Kovarianz** zwischen X und Y durch

$$\mathrm{Cov}(X,Y) := \mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y)$$

1. $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$
2. X,Y unabhängig $\implies$ $\mathrm{Cov}(X,Y)=0$ (Die Umkehrung ist falsch!)
3. $\mathrm{Var}(X+Y) = \mathrm{Var}(X)+\mathrm{Var}(Y)+2\mathrm{Cov}(X,Y)$

3.4 Bedingter Erwartungswert

Sei $(\Omega,\mathcal{A},\mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum und $X:\Omega\to\mathbb{R}$ eine Zufallsvariable.
Für ein beliebiges $B\in A,\mathbb{P}(B)>0$ definieren wir den **bedingten Erwartungswert** X bedingt durch B als

$$\mathbb{E}(X\mid B) = \frac{\mathbb{E}(\mathbb{1}_B X)}{\mathbb{P}(B)} = \sum_{x\in X(\Omega)} x\mathbb{P}(X=x\mid B) = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\}\mid B)$$

TODO: Beispielrechnung
Bedingter Erwartungswert als Zufallsvariable
Wir betrachten eine Partition $\mathcal{B}=(B_i)_{i\in I}$ von Ω (B_i sind disjunkt und nichtleer, I abzählbar).
Dann definieren wir die **Zufallsvariable**

$$\mathbb{E}(X\mid\mathcal{B})(\omega) = \sum_{i\in I,\mathbb{P}(B_i>0)} \mathbb{E}(X\mid B_i)\mathbb{1}_{B_i}(\omega)$$

1. **Intuition:** Die Information, die durch die Partition gegeben ist, ist dass eines der B_i eintreten wird. Bei der Realisierung durch das Eintreten des Elementarereignisses ω wird $\mathbb{E}(X\mid\mathcal{B})$ zu dem $\mathbb{E}(X\mid B_i)$ realisiert, bei welchem $\omega\in B_i$.
2. **Bemerkung:** Das $\mathcal{B}$ hat in der Vorlesung 2 verschiedene Bedeutungen. Es wird als Variable für sowohl die Borelsche σ-Algebra als auch die Partition von Ω verwendet.

TODO: Beispielrechnung

4 Mehrere Zufallsvariablen

Gemeinsame Verteilung

Die **gemeinsame Verteilungsfunktion** von n Zufallsvariablen $X_1,\dots,X_n$ (stetig oder diskret) ist die Abbildung $F:\mathbb{R}^n\to[0,1]$,

$$(x_1,\dots,x_n)\mapsto F(x_1,\dots,x_n) := \mathbb{P}(X_1\leq x_1,\dots,X_n\leq x_n)$$

4.1 Diskreter Fall - Gewichtsfunktion

Für n diskrete ZV $X_1,\dots,X_n$ definieren wir ihre **gemeinsame Gewichtsfunktion** $p:\mathbb{R}^n\to[0,1]$ durch

$$p(x_1,\dots,x_n) := \mathbb{P}(X_1=x_1,\dots,X_n=x_n)$$

Aus der gemeinsamen Gewichtsfunktion p bekommt man die gemeinsame Verteilungsfunktion mit

$$\begin{aligned}F(x_1,\dots,x_n) &= \mathbb{P}(X_1\leq x_1,\dots,X_n\leq x_n)\\ &= \sum_{y_1\leq x_1,\dots,y_n\leq x_n} \mathbb{P}(X_1=y_1,\dots,X_n=y_n)\\ &= \sum_{y_1\leq x_1,\dots,y_n\leq x_n} p(y_1,\dots,y_n)\end{aligned}$$

Konstruktion einer ZV

Seien $X_1,\dots,X_n$ diskrete Zufallsvariablen in $(\Omega,\mathcal{F},\mathbb{P})$, sodass $X_1\in W_1,\dots,X_n\in W_n$ f.s. für $W_1,\dots,W_n\subset\mathbb{R}$ endlich oder abzählbar.
Für $\phi:\mathbb{R}^n\to\mathbb{R}$ beliebig, ist $Z=\phi(X_1,\dots,X_n)$ eine diskrete Zufallsvariable mit $Z\in W=\phi(W_1\times\ldots\times W_n)$ f.s. .
Die Gewichtsfunktion von Z ist gegeben durch $p_Z:W\to[0,1]$:

$$p_Z(t) := \mathbb{P}(Z=t) = \sum_{\substack{x_1\in W_1,\dots,x_n\in W_n\\ \phi(x_1,\dots,x_n)=t}} p(x_1,\dots,x_n)$$

1. Mit dem vorherigen Satz können wir aus der gemeinsamen Verteilung die **Randverteilung** einer Zufallsvariablen extrahieren (wegsummieren). Wir verwenden dafür einfach die Funktion
$$\phi(x_1,\dots,x_n) = x_i$$
2. Der Erwartungswert des Bildes der Funktion $\phi:\mathbb{R}^n\to\mathbb{R}$ ist
$$\mathbb{E}(\phi(X_1,\dots,X_n)) = \sum_{x_1,\dots,x_n} \phi(x_1,\dots,x_n)p(x_1,\dots,x_n)$$

3. Wir haben eine Äquivalenz:

$$X_1, \dots, X_n \text{unabhängig}$$
$$\iff$$
$$\forall x_1 \in W_1, \dots, x_n \in W_n$$
$$p(x_1, \dots, x_n) = \mathbb{P}(X_1 = x_1) \cdot \dots \cdot \mathbb{P}(X_n = x_n)$$

4.2 Stetiger Fall - Gemeinsame Dichte

Gemeinsame Dichte

Falls die gemeinsame Verteilungsfunktion von n Zufallsvariablen $X_1, \dots, X_n$ sich schreiben lässt als

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1, \dots, t_n)\, dt_n \dots dt_1$$

für eine Funktion $f : \mathbb{R}^n \to [0, \infty)$, so heisst $f(x_1, \dots, x_n)$ die **gemeinsame Dichte** von $X_1, \dots X_n$.

1. $f(x_1, \dots, x_n) \geq 0$, und $= 0$ ausserhalb von $\mathcal{W}(X_1, \dots, X_n)$.
2. $$\mathbb{P}((X_1, \dots, X_n) \in A) = \int \cdots \int_{(x_1, \dots, x_n) \in A} f(x_1, \dots, x_n)\, dx_n \dots dx_1$$
   für $A \subseteq \mathbb{R}^n$ beliebig.
3. Haben X, Y die gemeinsame Verteilungsfunktion $F_{X,Y}$, so ist $F_X : \mathbb{R} \to [0, 1]$,
   $$F_X(x) := \mathbb{P}(X \leq x) = \mathbb{P}(X \leq x, Y \leq \infty) = \lim_{y \to \infty} F_{X,Y}(x, y)$$
   die Verteilungsfunktion der *Randverteilung* von X. Analoges gilt für F_Y.
4. Falls X, Y eine gemeinsame Dichte $f(x, y)$ haben, so haben auch die Randverteilungen von X und Y Dichten $f_X : \mathbb{R} \to [0, \infty)$ und $F_Y : \mathbb{R} \to [0, \infty)$.
   $$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \ \text{ bzw. } f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$
   Die **Dichtefunktion** einer Randverteilung (Randdichte) entsteht aus der gemeinsamen Dichtefunktion durch "Wegintegrieren" der anderen Variable(n).

Wenn $X_1, \dots, X_n$ stetige ZV mit Dichten $f_1, \dots, f_n$, dann sind die folgenden Aussagen äquivalent:

- $X_1, \dots, X_n$ unabhängig
- $(X_1, \dots, X_n)$ ist stetig mit gemeinsamer Dichte
  $$f(x_1, \dots, x_n) = f_1(x_1) \cdot \dots \cdot f_n(x_n)$$
- Für alle $\phi_1 : \mathbb{R} \to \mathbb{R}, \dots, \phi_n : \mathbb{R} \to \mathbb{R}$ die stückweise stetig und beschränkt sind, gilt
  $$\mathbb{E}(\phi_1(X_1) \cdot \dots \cdot \phi_n(X_n)) = \mathbb{E}(\phi_1(X_1)) \cdot \dots \cdot \mathbb{E}(\phi_n(X_n))$$

4.3 Transformation von Zufallsvariablen

Transformationssatz

Sei Z ein n-dimensionaler Zufallsvektor und $g : (\mathbb{R}^n, \mathcal{B}^n) \to (\mathbb{R}^m, \mathcal{B}^m)$ eine messbare Abbildung. Dann ist

$$H(\omega) = g(Z(\omega))$$

ein m-dimensionaler Zufallsvariable und ferner gilt

$$\mathbb{P}(H \in A) = \mathbb{P}(X \in g^{-1}(A)).$$

Wenn g linear und umkehrbar (i.e. $g(x) = m + Bx$ mit $\det(B) \neq 0$) und unter Vorraussetzung, dass die Verteilung von Z absolut stetig ist, dann ist H auch absolut stetig und es gilt:

$$f_H(x) = \frac{1}{|\det(B)|} f_Z(B^{-1}(x - m)).$$

Beispielrechnung
$Z = (X, Y)$ 2-dim Zufallsvektor. Wir wollen die Dichte von $X + Y$ berechnen.
Man wäre versucht die Matrix B und den Vektor m wie folgt zu definieren:

$$B = \begin{pmatrix} 1 & 1 \end{pmatrix} \text{ und } m = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$
$$\implies g((X, Y)) = \begin{pmatrix} 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \end{pmatrix} = X + Y$$

Dann wäre aber B (und somit g) nicht invertierbar! Deshalb wollen wir B so wählen, dass $g((X, Y)) = (X, X + Y)$:

$$\begin{pmatrix} X \\ X + Y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$
$$\det(B) = 1 \neq 0 \implies B \text{ invertierbar}$$
$$B^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$$

Nach dem Transformationssatz gilt

$$\begin{aligned} f_{X,X+Y}(x, z) &= \frac{1}{|\det(B)|} f_{X,Y}\left(B^{-1} \cdot \begin{pmatrix} x \\ z \end{pmatrix}\right) \\ &= 1 \cdot f_{X,Y}\left(\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix}\right) \\ &= f_{X,Y}(x, z - x) \end{aligned}$$

Aus der gemeinsamen Dichte $f_{X,X+Y}$ können wir die Dichte f_{X+Y} bestimmen.

$$\begin{aligned} f_{X+Y}(z) &= \int_{-\infty}^{\infty} f_{X,X+Y}(x, z)\, dx \\ &= \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)\, dx \end{aligned}$$

Falls X und Y unabhängig

$$= \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z - x)\, dx$$

Transformation von Zufallsvariablen

Sei $\phi : \mathbb{R}^n \to \mathbb{R}$ eine Abbildung und $X_1, \dots, X_n$ ZV mit gemeinsamer Dichte f. Dann lässt sich $\mathbb{E}(Z)$ für die Zufallsvariable $Z = \phi(X_1, \dots, X_n)$ mit

$$\mathbb{E}(Z) = \int \cdots \int_{\mathbb{R}^n} \phi(x_1, \dots, x_n) \cdot f(x_1, \dots, x_n)\, dx_n \dots dx_1$$

berechnen.

Beispielrechnung
Wir können diese Art den Erwartungswert zu berechnen nutzen, um die Dichte einer transformierten Zufallsvariable zu berechnen.
Seien X und Y zwei Zufallsvariablen mit gemeinsamer Dichtefunktion

$$f(x, y) = \begin{cases} \frac{1}{x^2 y^2} & \text{für } x \geq 1, y \geq 1 \\ 0 \text{ sonst.} \end{cases}$$

Bestimme die Dichtefunktion f_V der Zufallsvariable $V = XY$.
Wir definieren $\phi(x, y) = xy$ und berechnen

$$\begin{aligned} \mathbb{E}(V) = \mathbb{E}(\phi(X, Y)) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(x, y) f(x, y)\, dx\, dy \\ &= \int_1^{\infty} \int_1^{\infty} \phi(x, y) \frac{1}{x^2 y^2}\, dx\, dy \\ &\text{Substition } v = xy, dv = y\, dx \\ &= \int_1^{\infty} \int_y^{\infty} \frac{v}{y} \frac{1}{v^2}\, dv\, dy \end{aligned}$$

$$\begin{aligned} A &= \{(v, y) \in \mathbb{R}^2 \mid 1 \leq y < \infty, y \leq v < \infty\} \\ &= \{(v, y) \in \mathbb{R}^2 \mid 1 \leq y \leq v, 1 \leq v < \infty\} \end{aligned}$$

Zeichnung hilft ;)

$$\begin{aligned} &= \int_1^{\infty} \int_1^{v} \frac{1}{vy}\, dy\, dv \\ &= \int_1^{\infty} \frac{\ln(v)}{v}\, dv \\ &= \int_{-\infty}^{\infty} v \cdot \frac{\ln(v)}{v^2} \mathbb{1}_{v \in [1, \infty)}\, dv \end{aligned}$$

$$\implies f_V(t) = \frac{\ln(v)}{v^2} \mathbb{1}_{v \in [1, \infty)}$$

5 Konvergenz in Wahrscheinlichkeitsräumen

Unabhängigkeit einer Folge und iid./uiv.

Eine Folge von ZV $X_1, X_2, \dots$ ist unabhängig, wenn $X_1, \dots, X_n$ unabhängig ($\forall n \in \mathbb{N}$, nach der Definition in 2.1).
Sie ist zudem **uiv./iid.**, falls $F_{X_i} = F_{X_j}, \forall i, j \in \mathbb{N}$.

In einem Wahrscheinlichkeitsraum können wir für eine Folge von Zufallsvariablen $X_1, X_2, \dots$ und einer ZV Z zwischen 3 Arten von Konvergenz unterscheiden:

1. **schwache Konvergenz / Konvergenz in Verteilung**
   Wir definieren $X_n \xrightarrow{d} Z$ (d for distribution) als
   $$\lim_{n\to\infty} \mathbb{P}(X_n \leq x) = \lim_{n\to\infty} F_{X_n}(x) = F_Z(x) = \mathbb{P}(Z \leq x)$$
   für jede Stetigkeitsstelle $x \in \mathbb{R}$ von F_Z.
2. **Konvergenz in Wahrscheinlichkeit**
   Wir definieren $X_n \xrightarrow{\mathbb{P}} Z$ als
   $$\forall \varepsilon > 0 \quad \lim_{n\to\infty} \mathbb{P}(|X_n - Z| > \varepsilon) = 0$$
3. **Fast-sichere Konvergenz**
   Wir definieren $X_n \xrightarrow{\textbf{f.s.}} Z$ als
   $$\mathbb{P}(\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = Z(\omega)\}) = 1$$

Wir haben dann auch
$$X_n \xrightarrow{\textbf{f.s.}} Z \implies X_n \xrightarrow{\mathbb{P}} Z \implies X_n \xrightarrow{d} Z$$
Die Umkehrung der Implikationen gilt nicht, wie folgende Beispiele zeigen:

1. $X_n \xrightarrow{d} Z \not\Rightarrow X_n \xrightarrow{\mathbb{P}} Z$
   Sei $\Omega = \{0,1\}$ und für alle $n \in \mathbb{N}$
   $$\mathbb{P}(X_n = 0) = \mathbb{P}(X_n = 1) = \frac{1}{2}, \; X_n(\omega) = \begin{cases} 0 & \omega = 0 \\ 1 & \omega = 1 \end{cases}$$
   und
   $$\mathbb{P}(Z = 0) = \mathbb{P}(Z = 1) = \frac{1}{2}, \; Z(\omega) = \begin{cases} 1 & \omega = 0 \\ 0 & \omega = 1 \end{cases}$$
   .
   Aus $F_{X_n} = F_Z$ folgt direkt $X_n \xrightarrow{d} Z$.
   Da aber $|X_n(\omega) - Z(\omega)| = 1, \forall \omega \in \Omega$ und demzufolge $\lim_{n\to\infty} \mathbb{P}(|X_n - Z| > \epsilon) \not\to 0$ für $n \to \infty$, gilt
   $$X_n \overset{\mathbb{P}}{\not\to} Z$$
   .
2. $X_n \xrightarrow{\mathbb{P}} Z \not\Rightarrow X_n \xrightarrow{\textbf{f.s.}} Z$
   Wir betrachten den Wahrscheinlichkeitsraum $([0,1], \mathcal{B}, \mathbb{P})$. Für ein beliebiges $n \in \mathbb{N}$ sei $k = \lfloor \log_2(n) \rfloor$ und $j \in \{0, \dots, 2^k - 1\}$, sodass $n = 2^k + j$.
   Dann definieren wir
   $$X_n(\omega) = \mathbb{1}_{\left[\frac{j}{2^k}, \frac{j+1}{2^k}\right]}(\omega).$$
   und
   $$Z(\omega) = 0 \quad \forall \omega \in \Omega = [0,1]$$
   Zur Visualisierung würde die Folge so aussehen
   $$X_1 = \mathbb{1}_{[0,1]}, X_2 = \mathbb{1}_{[0,\frac{1}{2}]}, X_3 = \mathbb{1}_{[\frac{1}{2},1]}, X_4 = \mathbb{1}_{[0,\frac{1}{4}]} \text{ etc.}$$
   Wir hätten dann
   $$\forall \varepsilon > 0 \quad \lim_{n\to\infty} \mathbb{P}(|X_n - Z| > \varepsilon) = 0 \implies X_n \xrightarrow{\mathbb{P}} Z$$
   Aber für jedes $\omega \in [0,1]$ finden wir unendlich viele X_n mit $X_n(\omega) = 1$ und deshalb
   $$\mathbb{P}(\{\omega \in [0,1] \mid \lim_{n\to\infty} X_n(\omega) = Z(\omega)\}) = 0 \implies X_n \overset{\textbf{f.s.}}{\not\to} Z$$

5.1 Gesetz der grossen Zahlen

starkes Gesetz der grossen Zahlen

Sei $X_1, X_2, \dots$ eine Folge von uiv. Zufallsvariablen. Sei $\mathbb{E}(|X_1|) < \infty$ und $\mu = \mathbb{E}(X_1)$. Für
$$\overline{X}_n = \frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^n X_i$$
gilt dann
$$\overline{X}_n \longrightarrow \mu \textbf{ f.s.}$$

Dies ist eine fast-sichere Konvergenz.

schwaches Gesetz der grossen Zahlen

Sei $X_1, X_2, ...$ eine Folge von paarweise unkorrelierten Zufallsvariablen, die alle den gleichen Erwartungswert $\mathbb{E}(X_i) = \mu$ und die gleiche Varianz $\mathrm{Var}(X_i) = \sigma^2$ haben. Sei
$$\overline{X}_n = \frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^n X_i$$
Dann konvergiert $\overline{X}_n$ für $n \to \infty$ in Wahrscheinlichkeit gegen $\mu = \mathbb{E}(X_i)$, d.h.
$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| > \varepsilon) = 0 \text{ i.e. } \overline{X}_n \xrightarrow{\mathbb{P}} \mu$$

Bemerkung:
Zur Erinnerung:
$$X_i, X_j \text{ unkorreliert} \iff \mathrm{Cov}(X_i, X_j) = 0$$
Wir haben auch
$$X_i, X_j \text{ unabhängig} \implies X_i, X_j \text{ unkorreliert}$$

5.2 Zentraler Grenzwertsatz

Zentraler Grenzwertsatz

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von iid. Zufallsvariablen mit $\mathbb{E}(X_i) = \mu < \infty$ und $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Dann gilt
$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x) \quad \forall x \in \mathbb{R}$$
also
$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$$

Bemerkungen:
Man verwendet auch oft die Form für $\overline{X}_n = \frac{1}{n} S_n$ als
$$\frac{X_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$$

beziehungsweise
$$S_n \xrightarrow{d} \mathcal{N}(n\mu, n\sigma^2) \text{ und } \overline{X}_n \xrightarrow{d} \mathcal{N}\left(\mu, \frac{1}{n}\sigma^2\right)$$

Beispielrechnung:
Seien $(X_i)_{i\geq 1}, (Y_i)_{i\geq 1}$ und $(Z_i)_{i\geq 1}$ Folgen von iid. ZV mit
$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$$
und analog für Y_1 und Z_1. Wir definieren
$$S_n^{(x)} := \sum_{i=1}^n X_i, \quad S_n^{(y)} := \sum_{i=1}^n Y_i, \quad S_n^{(z)} := \sum_{i=1}^n Z_i$$
Die Folge $\left((S_n^{(x)}, S_n^{(y)}, S_n^{(z)})\right)_{n\geq 1}$ wird zufällige Irrfahrt in $\mathbb{Z}^3$ genannt. Sei $\alpha > \frac{1}{2}$. Zeige, dass
$$\mathbb{P}\left(\left\|(S_n^{(x)}, S_n^{(y)}, S_n^{(z)})\right\|_2 \leq n^\alpha\right) \longrightarrow 1 \text{ für } n \to \infty,$$
wobei $\|(x,y,z)\|_2 := \sqrt{x^2 + y^2 + z^2}$ die euklidische Norm ist.
Schritt 1: Für alle $\alpha > 1/2$ zeigen wir $\mathbb{P}(|S_n^{(x)}| \leq n^\alpha) \overset{n\to\infty}{\longrightarrow} 1$.
Da $\mathbb{E}(X_i) = 0$ und $\mathrm{Var}(X_i) = 1$ folgt für $a \in \mathbb{R}$ beliebig per ZGS
$$\mathbb{P}\left(S_n^{(x)} \leq a\sqrt{n}\right) = \mathbb{P}\left(\frac{S_n^{(x)}}{\sqrt{n}} \leq a\right) \overset{n\to\infty}{\longrightarrow} \Phi(a)$$
und somit auch
$$\mathbb{P}\left(|S_n^{(x)}| \leq a\sqrt{n}\right) = \mathbb{P}\left(S_n^{(x)} \leq a\sqrt{n}\right) - \mathbb{P}\left(S_n^{(x)} \leq -a\sqrt{n}\right)$$
$$\overset{n\to\infty}{\longrightarrow} \Phi(a) - \Phi(-a) = 2\Phi(a) - 1$$
Sei $\alpha = 1/2 + \beta, \beta > 0$. Dann instanzieren wir mit $a = n^\beta$.
$$\mathbb{P}\left(|S_n^{(x)}| \leq n^\alpha\right) = \mathbb{P}\left(|S_n^{(x)}| \leq n^\beta\sqrt{n}\right) \longrightarrow \lim_{n\to\infty}(2\Phi(n^\beta) - 1) = 1$$
Dies gilt analog für $S_n^{(y)}$ und $S_n^{(z)}$.
Schritt 2: $\forall \alpha > 1/2, \mathbb{P}\left(\left\|\left(S_n^{(x)}, S_n^{(y)}, S_n^{(z)}\right)\right\|_2 \leq n^\alpha\right) \overset{n\to\infty}{\longrightarrow} 1$.
Sei $\alpha' \in (1/2, \alpha)$. Dann folgt
$$\left\{|S_n^{(x)}| \leq n^{\alpha'} \wedge |S_n^{(y)}| \leq n^{\alpha'} \wedge |S_n^{(z)}| \leq n^{\alpha'}\right\}$$
$$\subseteq \left\{\left\|\left(S_n^{(x)}, S_n^{(y)}, S_n^{(z)}\right)\right\|_2 \leq \sqrt{3} \cdot n^{\alpha'}\right\}$$
Da $n^\alpha \geq \sqrt{3} n^{\alpha'}$ für grosse n, folgt
$$\lim_{n\to\infty} \mathbb{P}\left(\left\|\left(S_n^{(x)}, S_n^{(y)}, S_n^{(z)}\right)\right\|_2 \leq n^\alpha\right)$$
$$\geq \lim_{n\to\infty} \mathbb{P}\left(\left\|\left(S_n^{(x)}, S_n^{(y)}, S_n^{(z)}\right)\right\|_2 \leq \sqrt{3} \cdot n^{\alpha'}\right)$$
$$\geq \lim_{n\to\infty} \mathbb{P}\left(|S_n^{(x)}| \leq n^{\alpha'}, |S_n^{(y)}| \leq n^{\alpha'}, |S_n^{(z)}| \leq n^{\alpha'}\right) = 1$$