Stochastik für Informatiker

I - Wahrscheinlichkeiten

σ-Algebren

- Eine σ-Algebra $\mathcal{F}$ ist eine Menge von Teilmengen einer Menge Ω, sodass:
  - $\Omega \in \mathcal{F}$
  - Falls $A \in \mathcal{F}$, dann auch $\Omega \backslash A \in \mathcal{F}$
  - Falls $A_1, A_2, \ldots \in \mathcal{F}$, dann auch $A_1 \cup A_2 \cup \ldots \in \mathcal{F}$. Sind also die Mengen $\{1\}$ und $\{2\}$ enthalten, so auch $\{1,2\}$, was dem Ereignis "1 oder 2" entspricht.
- Bei vielen Anwendungen ist eine sinnvolle Wahl für $\mathcal{F}$ die Potenzmenge $\mathcal{P}(A)$.

Wahrscheinlichkeitsmaße

- Ein Wahrscheinlichkeitsmaß ist eine Funktion P, welche die in $\mathcal{F}$ enthaltenen Teilmengen (genannt **Ereignisse**) auf den Intervall $[0,1]$ abbildet (Also ihnen eine Wahrscheinlichkeit zuweist), sodass folgende Eigenschaften erfüllt sind:
  - $P(\Omega) = 1$
  - Für $A_1, A_2, \ldots \in \mathcal{F}$ mit $A_i \cap A_j = \emptyset$ gilt:

    $$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$$

    Diese Eigenschaft nennt man die **σ-Additivität** von P.

    Die Wahrscheinlichkeit des Ereignisses $\{1,2\}$ ("1 oder 2") ist also einfach $P(1) + P(2)$.
- Das folgende Wahrscheinlichkeitsmaß auf einem endlichen Raum Ω ist bekannt als **diskrete Gleichverteilung** oder **Laplace-Wahrscheinlichkeitsmaß**:

  $$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der günstigen Ereignisse}}{\text{Anzahl der möglichen Ereignisse}}$$

- Ein Wahrscheinlichkeitsmaß P heißt **diskret auf den Träger** Ω', falls es eine Menge $\Omega' \in \Omega$ gibt, sodass

  $$\sum_{\omega \in \Omega'} P(\omega) = 1$$

  Die Ereignisse ω nennen wir dann **Elementarereignisse**.

- Hieraus folgt, dass alle Ereignisse, welche keine Obermengen von Elementarereignissen sind, die Wahrscheinlichkeit 0 haben. Es gilt also

$$P(A) = \sum_{a \in A \cap \Omega'} P(a) \tag{1}$$

- Sei P ein diskretes Wahrscheinlichkeitsmaß mit Träger Ω'. Wir nennen das Wahrscheinlichkeitsmaß

$$p : \Omega' \to [0,1] p(\omega) = P(\omega)$$

die Zähldichte von p. Nach Lemma 1 wird P eindeutig von p bestimmt.

- Wichtige diskrete Verteilungen:
  - **Bernoulli-Verteilung** (Parameter $p \in (0,1)$)

$$\Omega' = \{0,1\}, \quad P(1) = p$$

  - **Binomialverteilung** (Parameter $N \in \mathbb{N}$ und $p \in (0,1)$)

$$\Omega' = \{0,1,\dots,N\}, \quad P(k) = \binom{N}{k} p^k (1-p)^{n-k}$$

  - **Geometrische Verteilung** ($p \in (0,1)$)

$$\Omega' = \mathbb{N}_1, \quad P(k) = p(1-p)^{k-1}$$

  - **Poisson-Verteilung** (Parameter $\lambda > 0$)

$$\Omega' = \mathbb{N}_0, \quad P(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

- Einige weitere wichtige Eigenschaften:
  - $P(\emptyset) = 0$
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - $P(\Omega \backslash A) = 1 - P(A)$
  - $P(A \cap \neg B) = P(A) - P(A \cap B)$
  - $A \subseteq B \implies P(A) \leq P(B)$
  - $P(\bigcup A_i) = \sum P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \dots$ (Also immer plus die ANDs mit ungerader Länge und minus die ANDs gerader Länge). Diese Formel ist bekannt als die **Inklusions-Exklusions-Formel**.

- Jedes Wahrscheinlichkeitsmaß ist **Stetig von Unten**, d.h. für eine Folge $A_n \subset A_{n+1}$ von Ereignissen gilt

$$\lim_{n\to\infty} P(A_n) = P\left(\bigcup A_i\right)$$

- Jedes Wahrscheinlichkeitsmaß ist **Stetig von Oben**, d.h. für eine Folge $A_{n+1} \subset A_n$ von Ereignissen gilt

$$\lim_{n\to\infty} P(A_n) = P\left(\bigcap A_i\right)$$

Wahrscheinlichkeitsräume

- Ein Wahrscheinlichkeitsraum ist definiert als das Tupel $(\Omega, \mathcal{F}, P)$, bestehend aus:
  - Einer beliebigen Menge Ω, gennant der **Grundraum**,
  - Einer σ-Algebra $\mathcal{F}$,
  - Einem Wahrscheinlichkeitsmaß P.
- Das N-fache Werfen eines fairen sechsseitigen Würfels sieht dementsprechend folgendermaßen aus:
  - $\Omega = \{1, 2, 3, 4, 5, 6\}^N$
  - $\mathcal{F} = \mathcal{P}(\Omega)$
  - P ist die Gleichverteilung.
  - Werfen wir einen unfairen Würfel, so bleibt Ω gleich, nur P verändert sich.

Borel σ-Algebra

- Falls Ω ein überabzählbarer Intervall $[a, b]$ ist, so ist eine geeignete σ-Algebra die **Borel σ-Algebra** $\mathcal{B}([a, b])$:

$$\mathcal{B}([a, b]) = \bigcap \{\mathcal{A} : \mathcal{A} \text{ ist } \sigma\text{-Algebra auf [a,b], die alle Teilintervalle } [c, d] \text{ enthält}\}$$

  Analog kann $\mathcal{B}(\mathbb{R})$ definiert werden. Es gilt

$$\mathcal{B}([a, b]) = \mathcal{B}(\mathbb{R}) \cap [a, b]$$

- **Satz 1.12:** Seien P und Q Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathcal{B}(\mathbb{R})$. So gilt:

$$\begin{aligned} &P([a, b]) = Q([a, b]) \forall a, b \in \mathbb{Q} \\ \implies &P(A) = Q(A) \forall A \in \mathcal{B}(\mathbb{R}) \end{aligned}$$

- Die stetige Gleichverteilung auf $[a, b] \in \mathbb{R}$ ist gegeben durch

$$P([c, d]) = \frac{d - c}{b - a}$$

Dichtefunktionen

- Eine **Dichtefunktion** ist eine integrierbare Funktion $f : \mathbb{R} \to \mathbb{R}$ mit folgenden Eigenschaften:
  - $f \geq 0$
  - $\int_{-\infty}^{\infty} f(x)dx = 1$
- Ein Wahrscheinlichkeitsmaß P wird **stetiges Maß P mit Dichtefunktion f** genannt, falls:

  $$P([a,b]) = \int_a^b f(x)dx$$

  Für jede Dichtefunktion f ist dieses Maß eindeutig.
- Wichtige stetige Verteilungen:
  - **Normalverteilung** (Parameter $\mu \in \mathbb{R}, \sigma^2 > 0$:

    $$f(x) = \frac{1}{2\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

  - **Exponentialverteilung** (Parameter $\lambda > 0$):

    $$f(x) = \frac{1}{\lambda} e^{-\lambda x} \mathbb{1}_{\{x \geq 0\}}$$

II - Rechnen mit Wahrscheinlichkeiten

I - Ziehen mit Zurücklegen und mit Ordnung

- Zieht man aus einer Urne N **Kugeln** mit m **Farben** mit Zurücklegen. Die Reihenfolge zählt, die gezogenen Kugel werden also als Tupel statt als Menge kodiert.
- Dann gibt es

  $$m^N$$

  mögliche Tupel.

II - Ziehen ohne Zurücklegen mit Ordnung

- Wir nehmen an dass jede Farbe nur einmal vertreten ist.
- Dann gibt es für eine Menge N gezogene Kugeln mit m Farben

  $$\frac{m!}{(m-N)!}$$

  Möglichkeiten.

III - Ziehen ohne Zurücklegen ohne Ordnung

- Hier gibt es
$$\binom{m}{N}$$
Möglichkeiten. Die Intuition dahinter ist folgende: Es gibt
$$\frac{m!}{(m-N)!}$$
geordnete Tupel. Jedes dieser Tupel hat $N!$ Permutationen. Nach ignorieren der Ordnung werden also je $N!$ Tupel auf die gleiche Menge abgebildet, also wird die Anzahl der Möglichkeiten durch $N!$ dividiert. Wir erhalten also
$$\frac{m!}{N!(m-N)!} = \binom{m}{N}$$

IV - Permutationen von m Kugeln in r Gruppen

- Es gibt m_n Kugeln der Gruppe n. Wenn alle Kugeln unterscheidbar wären gäbe es $m!$ Permutationen. Innerhalb von jeder Gruppe n gibt es $m_n!$ nicht unterscheidbare Teilpermutationen, die jeweils die Anzahl der Möglichkeiten reduzieren, genau wie im letzten Beispiel. Letzendlich gibt es also
$$\frac{m!}{m_1!m_2!\dots m_r!}$$
Möglichkeiten

V - Ziehen mit Zurücklegen ohne Ordnung

- Dies lässt sich auf das letzte Beispiel reduzieren, indem wir die gezogenen Kugeln folgendermaßen kodieren: Wir schreiben eine 0 für jede gezogene Kugel von Farbe 1, dann eine 1, dann eine 0 für jede gezogene Kugel von Farbe 2, dann wieder eine 0, etc. Nun entspricht das Beispiel IV, mit N 0ern und $m-1$ 1ern. Die Anzahl der Möglichkeiten ist also
$$\frac{(N+m-1)}{N!(m-1)!} = \binom{N+m+1}{N}$$

Beispiel

- Wie viele Möglichkeiten gibt es, bei 6 Würfen eines Sechsseitigen Würfels drei Paare zu würfeln?
- Es gibt nach III $\binom{6}{3} = 20$ Möglichkeiten, drei Zahlen Auszuwählen.
- Es gibt nach IV $\frac{6!}{2!2!2!}$ Möglichkeiten, diese Menge aus drei Paaren zu permutieren.
- Letztendlich gibt es also $20 \cdot \frac{6!}{2!2!2!} = 20 \cdot 90 = 1800$ Möglichkeiten, also eine Wahrscheinlichkeit von $\frac{1800}{6^6} = \frac{1800}{46.656}$

II.1 - Zufallsvariablen und Verteilungen

Zufallsvariablen

- Eine Abbildung $X : \Omega \to \mathbb{R}^d$ heißt **Zufallsvariable** oder **Zufallsvektor**, falls
$$\forall A \in \mathcal{B}(\mathbb{R}^d)(\{X(\omega) \in A\} \in \mathcal{F})$$
- Wenn Ω abzählbar ist ist jede Abbildung $\Omega \to \mathbb{R}^d$ eine Zufallsvariable.
- Dies gilt ebenfalls für alle stetigen X mit $\Omega \in \mathcal{B}(\mathbb{R}^d)$.
- Seien X, Y und g Zufallsvektoren. Dann ist $g(X, Y)$ ein Zufallsvektor. Insbesondere ist für $a, b \in \mathbb{R}$ die Abbildung $aX + bY$ ein Zufallsvektor.
- Die Wahrscheinlichkeit $P(X \in A)$ wird auch $P^X(A)$ geschrieben. Wir nennen sie die **Verteilung** von X
- Wir nennen die Funktion $F_X : x \to P(X \leq x)$ ***die*** **Verteilungsfunktion** von X.
- Für eine diskrete Zufallsvariable X mit Zähldichte p ist dies
$$F_X(x) = \sum_{k \leq x} p(k)$$
- Für eine stetige Zufallsvariable X mit Dichtefunktion P ist es
$$F_X(x) = \int_{-\infty}^{x} f(t)dt$$
- **Satz 2.12:** Eine Funktion $F : \mathbb{R} \to \mathbb{R}$ ist genau dann eine Verteilungsfunktion, wenn sie folgende Eigenschaften erfüllt:
  - F ist monoton wachsend
  - F ist rechtsseitig stetig
  - $\lim_{x \to \infty} F(x) = 1$, $\lim_{x \to -\infty} F(x) = 0$
- **Proposition 2.13:** Verteilungsfunktionen haben ebenfalls folgende Eigenschaften:
  - $P(X \geq x) = 1 - F_X(x)$
  - $P(x < X \leq y) = F_X(y) - F_X(x)$
- **Quantiltransformation:** Zu jeder Verteilungsfunktion F existiert eine Zufallsvariable X auf dem Wahrscheinlichkeitsraum
$$(\Omega = (0,1), \mathcal{F} = \mathcal{B}((0,1)), P = \text{Uni}((0,1))),$$
so dass $F = F_X$, nämlich die **Quantiltransformation** oder **Linksinverse**:
$$X(u) := \inf \{y \in \mathbb{R} : F(y) \geq u\}, u \in (0,1)$$

II.2 - Gemeinsame Verteilungen

- Ein Wahrscheinlichkeitsmaß P heißt stetig, falls es eine Funktion $f : \mathbb{R}^d \to \mathbb{R}_+$ gibt, sodass
$$P([a_1, b_1] \times \ldots \times [a_d, b_d]) = \int_{a_1}^{b_1} dx_1 \ldots \int_{a_d}^{b_d} dx_d f(x_1, \ldots, x_d)$$
für alle $a_n < b_n$

- Ein Zufallsvektor $X = (X_1, \ldots, X_d)$ heißt stetig, falls seine Verteilung F_X stetig ist. Die Dichtefunktion f_X heißt dann auch **gemeinsame Dichtefunktion** von $X_1, \ldots, X_d$.

- Es gilt
$$F_{X_k}(x_k) = F_X(\infty, \ldots, \infty, x_k, \infty, \ldots, \infty)$$

- Falls X eine Stetige Verteilung F_x (Und somit eine Dichtefunktion f_x hat, so auch jedes X_k, wobei gilt:
$$f_{X_k}(x_k) = \int_{-\infty}^{\infty} dx_1 \ldots \int_{-\infty}^{\infty} dx_{k-1} \ldots \int_{-\infty}^{\infty} dx_{k+1} \ldots \int_{-\infty}^{\infty} dx_d f_X(x_1, \ldots, x_d)$$
Diese Dichtefunktionen nennt man **Randdichten** oder **Marginaldichten**

- Analog gilt für Zähldichten p_X
$$p_{X_k}(x_k) = \sum_{x_1} \ldots \sum_{x_{k-1}} \sum_{x_{k+1}} \ldots \sum_{x_d} p_X(x_1, \ldots, x_d),$$
genannt **Randzähldichten** oder **Marginalzähldichten**.

- $X = (X_1, X_2)$ ist nicht umbedingt stetig nur weil X_1 und X_2 stetig sind, z.B. ist $X = (Y, Y)$ nicht stetig (da es sich effektiv um eine unendlich dünne Kurve handelt)

II.3 - Bedingte Wahrscheinlichkeiten

- Die Bedingte Wahrscheinlichkeit $P(A|B)$ (A, angenommen dass B gilt) ist
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- **Gesetz der Totalen Wahrscheinlichkeit:** Angenommen $\Omega = \bigcup B_i$ für paarweise disjunkte B_i. Dann gilt
$$P(A) = \sum P(B_i) P(A_{B_i})$$

- **Formel von Bayes:**
$$P(B|A) = \frac{P(A|B) P(B)}{P(A)}$$

II.4 - Unabhängigkeit von Ereignissen

- A ist unabhängig von B, wenn
$$P(A|B) = P(A).$$
- Dies ist genau der Fall wenn
$$P(A \cap B) = P(A)P(B)$$
- Analog für größere Mengen von Ereignissen - Die Familie A_i ist unabhängig, falls
$$P\left(\bigcap A_i\right) = \prod P(A_i)$$
für jede endliche Teilfamilie.
- Eine Menge von Ereignissen A_i heißt paarweise unabhängig, falls je zwei Ereignisse voneinander unabhängig sind. Alle unabhängigen Familien sind paarweise unabhängig, aber nicht alle paarweise unabhängigen Familien sind unabhängig.

II.5 - Unabhängigkeit von Zufallsvariablen

- Sei für $i \in I$ X_i ein d-dimensionaler Zufallsvektor. Die Familie $\{X_i\}_{i \in I}$ ist unabhängig gdw. für jede Teilmenge $J \subset I$ und alle $A_i \in \mathcal{B}(\mathbb{R}^d)$ gilt
$$P\left(\bigcap_{j \in J} \{X_j \in A_i\}\right) = \prod_{j \in J} P\left(\{X_j \in A_i\}\right)$$
- Diese Formel ist in der Praxis meistens zu aufwändig. Praktischere Kriterien:
  - Eine Familie $X = (X_1, \ldots, X_n)$ von **diskreten Zufallsvariablen** mit gemeinsamer Zähldichte p^X ist unabhängig gdw.
  $$p^X(x_1, \ldots, x_n) = \prod_{k=1}^{n} p^{X_k}(x_k)$$
  - Eine Familie $X = (X_1, \ldots, X_n)$ von **stetigen Zufallsvariablen** ist genau dann unabhängig, wenn
  $$\prod_{k=1}^{n} f^{X_k}(x_k)$$
  eine Dichtefunktion von X ist.

II.6.1 - Erwartungswert

- Sei g eine Zufallsvariable und X eine **Zufallsvariable**. Der **Erwartungswert** $E[g(X)]$ ist:
  - Falls X diskret ist:
$$E[g(X)] = \sum_x g(x)p^X(x)$$
  - Falls X stetig ist:
$$E[g(X)] = \int_{-\infty}^{\infty} dx_1 \dots \int_{-\infty}^{\infty} dx_d \; g(x_1, \dots, x_d) f_X(x_1, \dots, x_d)$$
- Sei X ein Zufallsvektor. Für eine Menge $A \in \mathcal{B}(\mathbb{R}^d)$ gilt:
$$P(X \in A) = E[\mathbb{1}_A(X)]$$
- $g(X)$ ist diskret, falls X diskret ist, aber für stetiges X ist $g(X)$ nicht umbedingt stetig.
- **Recheneigenschaften des Erwartungswerts**:
  - Falls X und Y gleiche Verteilungen haben gilt $E[X] = E[Y]$
  - $E[cX + Y] = cE[X] + E[Y]$
  - $|E[X]| \leq E[|X|]$
  - $P(X \leq Y) = 1 \implies E[X] \leq E[Y]$
  - $P(X_n \geq 0) = 1 \implies E[\sum X_n] = \sum E[X_n]$
  - Falls $X_n \to X$, dann $E[X_n] \to E[X]$
- **Markovs Ungleichung**:
$$\forall \epsilon > 0 \left(P\left(|X| \geq \epsilon\right) \leq \frac{E[|X|]}{\epsilon} \right)$$
- Seien X, Y Zufallsvariablen. Dann gilt:
  - $P(X \geq 0) = 1, E[X] = 0 \implies P(X = 0) = 1$
  - Wenn $P(X \leq Y) = 1$ dann $E[X] = E[Y]$ gdw. $P(X = Y) = 1$
- Sei X eine Zufallsvariable mit nichtnegativen Wertebereich. Dann gilt:
$$E[X] = \int_0^{\infty} (1 - F^X(x)) dx$$

II.6.2 - Varianz

- Sei X eine Zufallsvariable mit $|E[X]| < \infty$. Die **Varianz von** X ist definiert als

$$\text{Var}[X] := E[(X - E[X])^2]$$

- Es gilt:

$$\text{Var}[X] = E[X^2] - E[X]^2$$

- Einige Eigenschaften der Varianz:
  - $\text{Var}[X] \geq 0$
  - $\text{Var}[X] = 0 \iff P(X = E[X]) = 1$
  - Sei $f(x) = E[(X - x)^2]$. Die se Funktion hat ihr Minimum bei $x_0 = E[X]$, dieser Minimalwert ist genau die Varianz von X.

II.6.3 - Kovarianz

- Seien $X_1, \dots, X_N$ unabhängig mit wohldefiniertem Erwartungswert. Dann gilt

$$E\left[\prod_{k=1}^{N} X_k\right] = \prod_{k=1}^{N} E[X_k]$$

- Seien X und Y Zufallsvariablen. Die **Kovarianz** ist definiert als

$$\text{Cov}[X, Y] := E[(X - E[X])(Y - E[Y])]$$

- Die Kovarianz ist symmetrisch und bilinear, d.h.:
  - $\text{Cov}[X, Y] = \text{Cov}[Y, X]$
  - $\text{Cov}[aX+bY, cZ+dV] = ac\text{Cov}[X, Z]+ad\text{Cov}[X, V]+bc\text{Cov}[Y, Z]+bd\text{Cov}[Y, V]$
- Es gilt:

$$\text{Var}\left[\sum X_k\right] = \sum \text{Var}[X_k] + \sum \text{Cov}[X_i, X_j]$$

- Es gilt:

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y]$$

  Daraus folgt:

$$\text{Cov}[X, X] = \text{Var}[X]\text{Cov}[X, c] = 0$$

  für konstantes $c \in \mathbb{R}$

- Seien X und Y unabhängig. Dann gilt
$$\text{Cov}[X, Y] = 0$$
- Es gilt
$$(\text{Cov}[X, Y])^2 \leq \text{Var}[X]\text{Var}[Y]$$
- Zwei Variablen X, Y heißen **unkorreliert**, falls
$$\text{Cov}[X, Y] = 0$$
Unabhängige Variablen sind immer unkorreliert, aber unkorrelierte Variablen sind nicht immer unabhängig. Informell zeigt die Kovarianz nur "lineare" Abhängigkeit.
- $\text{Cov}[X, Y]^2 = \text{Var}[X]\text{Var}[Y]$ gdw. es Konstanten $a, b \in \mathbb{R}$ gibt, sodass $P(Y = aX + b) = 1$
- Die **Korrelation** zweier Zufallsvariablen X, Y ist definiert als:
$$\text{Cor}[X, Y] = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]}\sqrt{\text{Var}[Y]}}$$

2.7 - Bedingte Verteilungen und Erwartungswerte

- Die **bedingte Verteilung** von Y gegeben $X = x$ ist definiert als
$$A \to P(Y \in A|X = x)$$
- Der **bedingte Erwartungswert** von Y gegeben $X = x$ ist definiert als
$$E[g(Y)|X = x] = \sum_y g(y)P(Y = y|X = x)$$
- Beide dieser Objekte sind Zufallsvariablen, da sie von x abhängen.
- **Turmeigenschaft**: Seien X, Y diskrete Zufallsvariablen und g eine Zufallsvariable. Dann gilt
$$E[E[g(Y)|X]] = E[g(Y)]$$
- Die **Bedingte Dichte** von Y gegeben X ist
$$f^{Y|X}(x, y) = \frac{f^{(X,Y)(x,y)}}{f^X(x)}$$
- Die **Bedingte Erwartung** von $g(Y)$ gegeben X ist
$$E[g(Y)|X](x) = \int g(y)f^{Y|X}(x, y)dy$$
- X und Y sind unabhängig gdw.
$$P(Y \in A|X) = P(Y \in A)\forall A \in \mathcal{B}(\mathbb{R}^d)$$
-

4 - Grenzwertsätze

4.1 - Das Gesetz der Großen Zahlen

- **Schwaches Gesetz der Großen Zahlen:** Seien $X_1, X_2, \ldots, X_n$ paarweise unkorrelliert und identisch verteilt. Sei
$$S_n = \sum_{k=1}^{n} X_k$$
das artithmetische Mittel. Dann gilt $\forall \epsilon > 0$:
$$\lim_{n\to\infty} P(|\frac{S_n}{n} - E[X_1]| \geq \epsilon) = 0$$
d.h. der Durchschnitt einer großen Zahl an Zufallsexperimenten konvergiert gegen den Erwartungswert.

- **Chebyshev's Ungleichung**:
$$P(|X - E[X]| \geq \epsilon) \leq \frac{Var[X]}{\epsilon^2}$$

- Wir sagen, eine Folge $Y_1, Y_2, \ldots$ **konvergiert Stochastisch gegen eine Zufallsvariable** Y, falls
$$\forall \epsilon > 0 \lim_{n\to\infty} P(|Y_n - Y| \geq \epsilon) = 0$$

- Wir sagen, die Folge konvergiert **fast sicher** gegen Y, falls:
$$P(\lim_{n\to\infty} Y_n = Y) = 1$$
Dies ist eine stärkere Bedingung als die stochastische Konvergenz von davor.

- $Y_n \to Y$ fast sicher gdw. es eine Menge $N \in \mathcal{F}$ gibt, sodass $P(N) = 0$ und
$$\lim_{n\to\inf} Y_n(\omega) = Y(\omega) \quad \forall \omega \neq N$$

- **Starkes Gesetz der großen Zahlen:**
$$\frac{S_n}{n} \to E[X_1] \text{ konvergiert fast sicher.}$$

- **Monte Carlo Integration**: Sei $f : [0,1] \to \mathbb{R}$ stetig und U auf $[0,1]$ gleichverteilt. Dann gilt:
$$\int_0^1 f(x)dx = E[f(U)]$$
Aus dem starken Gesetz der großen Zahlen folgt
$$\frac{1}{n}\sum_{k=1}^{n} f(U_k) \to E[f(U_1)] = \int_0^1 f(x)dx$$

4.2 - Der Zentrale Grenzwertsatz

- Y_n **konvergiert gegen** Y **in Verteilung**, falls für die Verteilungsfunktionen F_n und F gilt:

$$F_n(x) \to F(x)$$

  Für alle Stetigkeitsstellen von F (alle x, sodass $F(x) = F(x-)$)

- **Zentraler Grenzwertsatz:** Seien $X_1, X_2, \ldots$ IID (unabhängig und identisch verteilt). Dann gilt

$$\frac{(\sum_{k=1}^{n} X_k) - nE[X_1]}{\sqrt{n\text{Var}[X_1]^2}} \to N(0,1) \text{ in Verteilung}$$

5 - Statistik

- Ein **Schätzer** ist eine Zufallsvariable, welche einen Statistischen Prozess beschreiben soll.
- Eine Folge von Schätzern T_n heißt konsistent