

8. Formelsammlung

8.1 Gesetze zum Rechnen mit Ereignissen

Im Folgenden seien A und B , sowie A_1, \dots, A_n Ereignisse. Die Notation $A \uplus B$ steht für $A \cup B$ und zugleich $A \cap B = \emptyset$ (disjunkte Vereinigung). $A_1 \uplus \dots \uplus A_n = \Omega$ bedeutet also, dass die Ereignisse A_1, \dots, A_n eine Partition der Ergebnismenge Ω bilden.

$$\Pr[\emptyset] = 0$$

$$0 \leq \Pr[A] \leq 1$$

$$\Pr[\bar{A}] = 1 - \Pr[A]$$

$$A \subseteq B \implies \Pr[A] \leq \Pr[B]$$

$\forall i \neq j : A_i \cap A_j = \emptyset \implies$ $\Pr \left[\bigcup_{i=1}^n A_i \right] = \sum_{i=1}^n \Pr[A_i]$	Additionssatz
$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$ allgemeine Form: siehe Satz 9	Inklusion/Exklusion, Siebformel
$\Pr \left[\bigcup_{i=1}^n A_i \right] \leq \sum_{i=1}^n \Pr[A_i]$	Boolesche Ungleichung
$\Pr[A B] = \frac{\Pr[A \cap B]}{\Pr[B]}$ für $\Pr[B] > 0$	Def. bedingte Ws.

$$B \subseteq A_1 \uplus \dots \uplus A_n \implies \\ \Pr[B] = \sum_{i=1}^n \Pr[B|A_i] \cdot \Pr[A_i]$$

Satz von der totalen
Wahrscheinlichkeit

$$\Pr[B] > 0, B \subseteq A_1 \uplus \dots \uplus A_n \implies \\ \Pr[A_i|B] = \frac{\Pr[B|A_i] \cdot \Pr[A_i]}{\sum_{i=1}^n \Pr[B|A_i] \cdot \Pr[A_i]}$$

Satz von Bayes

$$\Pr[A_1 \cap \dots \cap A_n] = \Pr[A_1] \cdot \Pr[A_2|A_1] \cdot \\ \dots \cdot \Pr[A_n|A_1 \cap \dots \cap A_{n-1}]$$

Multiplikationssatz

$$A \text{ und } B \text{ unabhängig} \iff \\ \Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$$

Definition
Unabhängigkeit

8.2 Erwartungswert und Varianz diskreter Zufallsvariablen

Sei X eine diskrete Zufallsvariable. Für Erwartungswert und Varianz gelten die folgenden Formeln (sofern $\mathbb{E}[X]$ und $\text{Var}[X]$ existieren).

$$\begin{aligned}\mathbb{E}[X] &= \sum_{x \in W_X} x \cdot \Pr[X = x] \\ &= \sum_{\omega \in \Omega} X(\omega) \cdot \Pr[\omega] && \text{Erwartungswert} \\ \left(&= \sum_{i=1}^{\infty} \Pr[X \geq i], \quad \text{falls } W_X \subseteq \mathbb{N}_0 \right)\end{aligned}$$

$$\begin{aligned}\text{Var}[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \sum_{x \in W_X} \Pr[X = x] \cdot (x - \mathbb{E}[X])^2 && \text{Varianz}\end{aligned}$$

8.3 Gesetze zum Rechnen mit Zufallsvariablen

Seien $a, b, a_1, \dots, a_n \in \mathbb{R}, f_1, \dots, f_n : \mathbb{R} \rightarrow \mathbb{R}$.

$$X_1, \dots, X_n \text{ unabhängig} \iff \text{für alle } (a_1, \dots, a_n): \\ \Pr[X_1 = a_1, \dots, X_n = a_n] \\ = \Pr[X_1 = a_1] \cdot \dots \cdot \Pr[X_n = a_n]$$

$$X_1, \dots, X_n \text{ unabhängig} \implies f_1(X_1), \dots, f_n(X_n) \text{ unabhängig}$$

$$\mathbb{E}[a \cdot X + b] = a \cdot \mathbb{E}[X] + b$$

$$X(\omega) \leq Y(\omega) \text{ für alle } \omega \in \Omega \implies \\ \mathbb{E}[X] \leq \mathbb{E}[Y]$$

Monotonie des
Erwartungswerts

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X|A_i] \cdot \Pr[A_i]$$

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

$$\text{Var}[a \cdot X + b] = a^2 \cdot \text{Var}[X]$$

$$\begin{aligned} &\mathbb{E}[a_1X_1 + \dots + a_nX_n] \\ &= a_1\mathbb{E}[X_1] + \dots + a_n\mathbb{E}[X_n] \end{aligned}$$

Linearität des
Erwartungswerts

$$\begin{aligned} &X_1, \dots, X_n \text{ unabhängig} \implies \\ &\mathbb{E}[X_1 \cdot \dots \cdot X_n] = \mathbb{E}[X_1] \cdot \dots \cdot \mathbb{E}[X_n] \end{aligned}$$

Multiplikatивität des
Erwartungswerts

$$\begin{aligned} &X_1, \dots, X_n \text{ unabhängig} \implies \\ &\text{Var}[X_1 + \dots + X_n] = \text{Var}[X_1] + \dots + \\ &\text{Var}[X_n] \end{aligned}$$

Varianz
einer Summe

$$X \geq 0 \implies$$

$$\Pr[X \geq t] \leq \mathbb{E}[X]/t \text{ für } t > 0$$

Markov

$$\Pr[|X - \mathbb{E}[X]| \geq t] \\ \leq \text{Var}[X]/t^2 \text{ für } t > 0$$

Chebyshev

siehe Satz 63

Gesetz der
großen Zahlen

Kapitel II Kontinuierliche Wahrscheinlichkeitsräume

1. Einführung

1.1 Motivation

Interpretation der Poisson-Verteilung als Grenzwert der Binomialverteilung.

Beispiel 85

Wir betrachten das Szenario: Bei einem Druckerserver kommen Aufträge in einer Warteschlange an, die alle $1/n$ Zeiteinheiten vom Server abgefragt wird. Der Server nimmt also zu den diskreten Zeitpunkte $1/n, 2/n, 3/n, \dots$ neue Aufträge entgegen. Durch den Grenzwert $n \rightarrow \infty$ „verschmelzen“ diese diskreten Zeitpunkte zu einer kontinuierlichen Zeitachse, und für die Zufallsvariable T , welche die Zeitspanne bis zum Eintreffen des nächsten Auftrags misst, reicht eine diskrete Wertemenge W_T nicht mehr aus.

1.2 Kontinuierliche Zufallsvariablen

Definition 86

Eine **kontinuierliche** oder auch **stetige Zufallsvariable** X und ihr zugrunde liegender **kontinuierlicher (reeller)**

Wahrscheinlichkeitsraum sind definiert durch eine integrierbare Dichte(-funktion) $f_X : \mathbb{R} \rightarrow \mathbb{R}_0^+$ mit der Eigenschaft

$$\int_{-\infty}^{+\infty} f_X(x) \, dx = 1.$$

Eine Menge $A \subseteq \mathbb{R}$, die durch Vereinigung $A = \bigcup_k I_k$ abzählbar vieler paarweise disjunkter Intervalle beliebiger Art (offen, geschlossen, halboffen, einseitig unendlich) gebildet werden kann, heißt **Ereignis**. Ein Ereignis A tritt ein, wenn X einen Wert aus A annimmt. Die Wahrscheinlichkeit von A ist bestimmt durch

$$\Pr[A] = \int_A f_X(x) \, dx = \sum_k \int_{I_k} f_X(x) \, dx.$$

Beispiel 87 (Gleichverteilung)

Eine besonders einfache kontinuierliche Dichte stellt die **Gleichverteilung** auf dem Intervall $[a, b]$ dar. Sie ist definiert durch

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } x \in [a, b], \\ 0 & \text{sonst.} \end{cases}$$

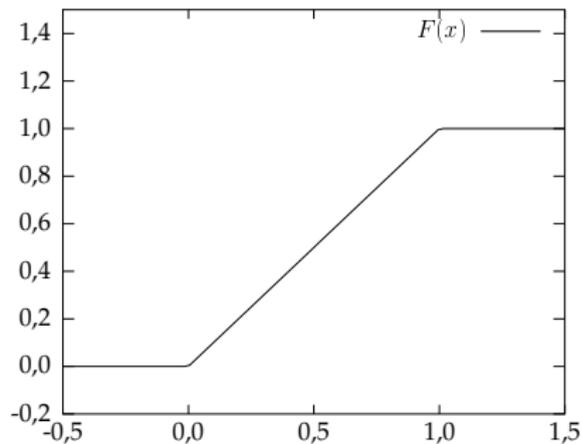
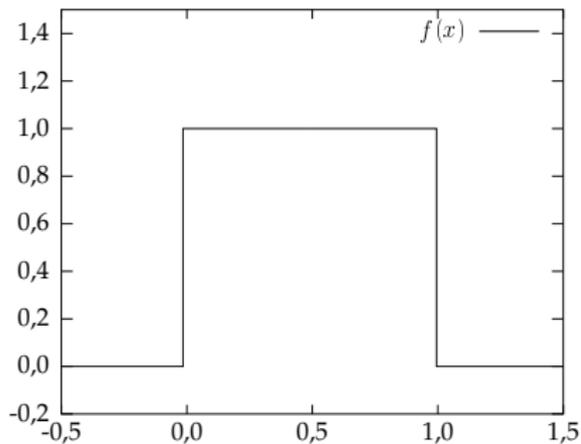
Analog zum diskreten Fall ordnen wir jeder Dichte f_X eine **Verteilung** oder **Verteilungsfunktion** F_X zu:

$$F_X(x) := \Pr[X \leq x] = \Pr[\{t \in \mathbb{R} \mid t \leq x\}] = \int_{-\infty}^x f_X(t) \, dt.$$

Beispiel 88

Die Verteilungsfunktion der Gleichverteilung:

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0 & \text{für } x < a, \\ \frac{x-a}{b-a} & \text{für } a \leq x \leq b, \\ 1 & \text{für } x > b. \end{cases}$$



Gleichverteilung über dem Intervall $[0, 1]$

Beobachtungen: (Eigenschaften der Verteilungsfunktion)

- F_X ist monoton steigend.
- F_X ist stetig. Man spricht daher auch von einer „stetigen Zufallsvariablen“.
- Es gilt: $\lim_{x \rightarrow -\infty} F_X(x) = 0$ und $\lim_{x \rightarrow \infty} F_X(x) = 1$.
- Jeder (außer an endlich vielen Punkten) differenzierbaren Funktion F , welche die zuvor genannten Eigenschaften erfüllt, können wir eine Dichte f durch $f(x) = F'(x)$ zuordnen.

Es gilt

$$\Pr[a < X \leq b] = F_X(b) - F_X(a).$$

Bei den von uns betrachteten Dichten besteht zwischen den Ereignissen „ $a < X \leq b$ “, „ $a \leq X \leq b$ “, „ $a \leq X < b$ “ und „ $a < X < b$ “ kein wesentlicher Unterschied, da

$$\int_{[a,b]} f(t) \, dt = \int_{]a,b]} f(t) \, dt = \int_{[a,b[} f(t) \, dt = \int_{]a,b[} f(t) \, dt.$$

1.3 Kolmogorov-Axiome und σ -Algebren

1.3.1 σ -Algebren

Definition 89

Sei Ω eine Menge. Eine Menge $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ heißt σ -Algebra über Ω , wenn folgende Eigenschaften erfüllt sind:

(E1) $\Omega \in \mathcal{A}$.

(E2) Wenn $A \in \mathcal{A}$, dann folgt $\bar{A} \in \mathcal{A}$.

(E3) Für $n \in \mathbb{N}$ sei $A_n \in \mathcal{A}$. Dann gilt auch $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Für jede (endliche) Menge Ω stellt die Menge $\mathcal{P}(\Omega)$ eine σ -Algebra dar.

Für $\Omega = \mathbb{R}$ ist die Klasse der *Borel'schen Mengen*, die aus allen Mengen $A \subseteq \mathbb{R}$ besteht, welche sich durch (abzählbare) Vereinigungen und Schnitte von Intervallen (offen, halboffen oder geschlossen) darstellen lassen, eine σ -Algebra.

1.3.2 Kolmogorov-Axiome

Definition 90 (Wahrscheinlichkeitsraum, Kolmogorov-Axiome)

Sei Ω eine beliebige Menge und \mathcal{A} eine σ -Algebra über Ω . Eine Abbildung

$$\Pr[\cdot] : \mathcal{A} \rightarrow [0, 1]$$

heißt **Wahrscheinlichkeitsmaß** auf \mathcal{A} , wenn sie folgende Eigenschaften besitzt:

- 1 (W1) $\Pr[\Omega] = 1$.
- 2 (W2) A_1, A_2, \dots seien paarweise disjunkte Ereignisse. Dann gilt

$$\Pr \left[\bigcup_{i=1}^{\infty} A_i \right] = \sum_{i=1}^{\infty} \Pr[A_i].$$

Für ein Ereignis $A \in \mathcal{A}$ heißt $\Pr[A]$ **Wahrscheinlichkeit** von A . Ein **Wahrscheinlichkeitsraum** ist definiert durch das Tupel $(\Omega, \mathcal{A}, \Pr)$.

Die in obiger Definition aufgelisteten Eigenschaften eines Wahrscheinlichkeitsmaßes wurden von dem russischen Mathematiker **Andrei Nikolaevich Kolmogorov** (1903–1987) formuliert. Kolmogorov gilt als einer der Pioniere der modernen Wahrscheinlichkeitstheorie, leistete jedoch auch bedeutende Beiträge zu zahlreichen anderen Teilgebieten der Mathematik. Informatikern begegnet sein Name auch im Zusammenhang mit der so genannten Kolmogorov-Komplexität, einem relativ jungen Zweig der Komplexitätstheorie.

Die Eigenschaften in obiger Definition nennt man auch **Kolmogorov-Axiome**.

Lemma 91

Sei $(\Omega, \mathcal{A}, \Pr)$ ein Wahrscheinlichkeitsraum. Für Ereignisse A, B, A_1, A_2, \dots gilt

- 1 $\Pr[\emptyset] = 0, \Pr[\Omega] = 1.$
- 2 $0 \leq \Pr[A] \leq 1.$
- 3 $\Pr[\bar{A}] = 1 - \Pr[A].$
- 4 *Wenn $A \subseteq B$, so folgt $\Pr[A] \leq \Pr[B].$*

Lemma 91

- 5 (Additionssatz) Wenn die Ereignisse A_1, \dots, A_n paarweise disjunkt sind, so folgt

$$\Pr \left[\bigcup_{i=1}^n A_i \right] = \sum_{i=1}^n \Pr[A_i].$$

Für disjunkte Ereignisse A, B erhalten wir insbesondere

$$\Pr[A \cup B] = \Pr[A] + \Pr[B].$$

Für eine unendliche Menge von paarweise disjunkten Ereignissen A_1, A_2, \dots gilt analog

$$\Pr \left[\bigcup_{i=1}^{\infty} A_i \right] = \sum_{i=1}^{\infty} \Pr[A_i].$$

Beweis:

Wenn wir in Eigenschaft (W2) $A_1 = \Omega$ und $A_2, A_3, \dots = \emptyset$ setzen, so ergibt die Eigenschaft, dass $\Pr[\Omega] + \sum_{i=2}^{\infty} \Pr[\emptyset] = \Pr[\Omega]$.

Daraus folgt $\Pr[\emptyset] = 0$.

Regel 2 und Regel 5 gelten direkt nach Definition der Kolmogorov-Axiome und Regel 1.

Regel 3 erhalten wir mit Regel 5 wegen

$$1 = \Pr[\Omega] = \Pr[A] + \Pr[\bar{A}].$$

Für Regel 4 betrachten wir die disjunkten Ereignisse A und $C := B \setminus A$, für die gilt, dass $A \cup B = A \cup C$. Mit Regel 5 folgt die Behauptung. □

1.3.3 Lebesgue-Integrale

Eine Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ heißt **messbar**, falls das Urbild jeder Borel'schen Menge ebenfalls eine Borel'sche Menge ist.

Z.B. ist für jede Borel'sche Menge A die Indikatorfunktion

$$I_A : x \mapsto \begin{cases} 1 & \text{falls } x \in A, \\ 0 & \text{sonst} \end{cases}$$

messbar. Jede stetige Funktion ist messbar. Auch Summen und Produkte von messbaren Funktionen sind wiederum messbar. Jeder messbaren Funktion kann man ein Integral, das so genannte **Lebesgue-Integral**, geschrieben $\int f \, d\lambda$, zuordnen.

Ist $f : \mathbb{R} \rightarrow \mathbb{R}_0^+$ eine messbare Funktion, so definiert

$$\Pr : A \mapsto \int f \cdot I_A \, d\lambda$$

eine Abbildung auf den Borel'schen Mengen, die die Eigenschaft (W2) der Kolmogorov-Axiome erfüllt. Gilt daher zusätzlich noch $\Pr[\mathbb{R}] = 1$, so definiert f auf natürliche Weise einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \Pr)$, wobei $\Omega = \mathbb{R}$ und \mathcal{A} die Menge der Borel'schen Mengen ist.

1.4 Rechnen mit kontinuierlichen Zufallsvariablen

1.4.1 Funktionen kontinuierlicher Zufallsvariablen

Sei $Y := g(X)$ mit einer Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$.

Die Verteilung von Y erhalten wir durch

$$F_Y(y) = \Pr[Y \leq y] = \Pr[g(X) \leq y] = \int_C f_X(t) \, dt.$$

Hierbei bezeichnet $C := \{t \in \mathbb{R} \mid g(t) \leq y\}$ alle reellen Zahlen $t \in \mathbb{R}$, für welche die Bedingung „ $Y \leq y$ “ zutrifft. Das Integral über C ist nur dann sinnvoll definiert, wenn C ein zulässiges Ereignis darstellt. Aus der Verteilung F_Y können wir durch Differenzieren die Dichte f_Y ermitteln.

Beispiel 92

Sei X gleichverteilt auf dem Intervall $]0, 1[$. Für eine Konstante $\lambda > 0$ definieren wir die Zufallsvariable $Y := -(1/\lambda) \ln X$.

$$\begin{aligned} F_Y(y) &= \Pr[-(1/\lambda) \ln X \leq y] = \Pr[\ln X \geq -\lambda y] \\ &= \Pr[X \geq e^{-\lambda y}] \\ &= 1 - F_X(e^{-\lambda y}) \\ &= \begin{cases} 1 - e^{-\lambda y} & \text{für } y \geq 0, \\ 0 & \text{sonst.} \end{cases} \end{aligned}$$

Beispiel (Forts.)

Damit folgt mit $f_Y(y) = F'_Y(y)$ sofort

$$f_Y(y) = \begin{cases} \lambda e^{-\lambda y} & \text{für } y \geq 0, \\ 0 & \text{sonst.} \end{cases}$$

Eine Zufallsvariable mit einer solchen Dichte f_Y nennt man *exponentialverteilt*.

Beispiel 93

Sei X eine beliebige Zufallsvariable. Für $a, b \in \mathbb{R}$ mit $a > 0$ definieren wir die Zufallsvariable $Y := a \cdot X + b$.

Es gilt

$$F_Y(y) = \Pr[aX + b \leq y] = \Pr\left[X \leq \frac{y-b}{a}\right] = F_X\left(\frac{y-b}{a}\right),$$

und somit

$$f_Y(y) = \frac{d F_Y(y)}{d y} = \frac{d F_X((y-b)/a)}{d y} = f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a}.$$

Simulation von Zufallsvariablen

Unter der **Simulation** einer Zufallsvariablen X mit Dichte f_X versteht man die algorithmische Erzeugung von Zufallswerten, deren Verteilung der Verteilung von X entspricht.

Dazu nehmen wir an, dass die zu simulierende Zufallsvariable X eine stetige, im **Bildbereich** $]0, 1[$ streng monoton wachsende Verteilungsfunktion F_X besitzt. Weiter nehmen wir an, dass U eine auf $]0, 1[$ gleichverteilte Zufallsvariable ist, die wir simulieren können.

Aus unserer Annahme über F_X folgt, dass es zu F_X eine (eindeutige) inverse Funktion F_X^{-1} gibt mit $F_X(F_X^{-1}(x)) = x$ für alle $x \in]0, 1[$.

Sei nun

$$\tilde{X} := F_X^{-1}(U),$$

dann gilt

$$\begin{aligned}\Pr[\tilde{X} \leq t] &= \Pr[F_X^{-1}(U) \leq t] \\ &= \Pr[U \leq F_X(t)] \\ &= F_U(F_X(t)) \\ &= F_X(t).\end{aligned}$$

Beispiel 94

Im obigen Beispiel der Exponentialverteilung gilt $F_X(t) = 1 - e^{-t}$ für $t \geq 0$, und wir erhalten auf $]0, 1[$ die Umkehrfunktion $F_X^{-1}(t) = -\ln(1 - t)$. Also gilt $\tilde{X} = F_X^{-1}(U) = -\ln(1 - U)$.

Statt \tilde{X} haben wir im Beispiel die Zufallsvariable $-\ln U$ betrachtet, die aber offensichtlich dieselbe Verteilung besitzt.