

Mathematik 2
Dr. Thomas Zehrt

Zufallsvariablen 1:
Grundlagen

Benötigtes Vorwissen: Der Stoff der Vorlesung „Statistik“ wird als bekannt vorausgesetzt, insbesondere das **Kapitel „Zufallsvariablen“**. Als Zusammenfassung kann dafür der erste Teil dieses Skriptes dienen. Weiterhin wird das Kapitel „Integration“ aus der Vorlesung „Mathematik 1“ als bekannt vorausgesetzt.

Inhaltsverzeichnis

1	Zufallsvariablen(Wiederholung)	2
2	Stetige Zufallsvariablen	5
3	Erwartungswert, Varianz und Quantil	8
4	Die Ungleichung von Tschebyshev	13
5	Standardisierung einer Zufallsvariablen	16
6	Übungsaufgaben	18

1 Zufallsvariablen(Wiederholung)

Definition 1.1 Sei (Ω, P) ein diskreter Wahrscheinlichkeitsraum. Dann heisst eine Abbildung

$$X : \Omega \longrightarrow \mathbb{R}$$

eine (reellwertige) Zufallsvariable, falls alle Ereignisse der Form

- $\{\omega \in \Omega : X(\omega) = x\} \subset \Omega$ (alle $\omega \in \Omega$ die von X auf x abgebildet werden) für alle reellen Zahlen x und
- $A_I = \{\omega \in \Omega : X(\omega) \in I\} \subset \Omega$ (alle $\omega \in \Omega$ die von X in das Intervall I abgebildet werden) für alle Intervalle I

Wahrscheinlichkeiten besitzen, die dem **Axiomsystem von Kolmogorov** genügen.

Oft wird man sich aber für die Wahrscheinlichkeit interessieren, dass $X(\omega)$ in einem bestimmten Intervall $I = [a, b]$ oder auch $I = (-\infty, b]$ liegt, also dass $X(\omega) \in I$ gilt. Dazu wählen wir die folgenden Schreibweisen:

$$\begin{aligned} P(X \in I) &= P(\{\omega \in \Omega : X(\omega) \in I\}) \\ P(a \leq X \leq b) &= P(\{\omega \in \Omega : a \leq X(\omega) \leq b\}) \\ P(X \leq b) &= P(\{\omega \in \Omega : X(\omega) \leq b\}) \end{aligned}$$

Definition 1.2 Sei X eine Zufallsvariable auf dem Wahrscheinlichkeitsraum (Ω, P) . Dann heisst die Abbildung $F : \mathbb{R} \longrightarrow [0, 1]$ mit

$$F(x) = P(X \leq x) = P(-\infty \leq X \leq x)$$

Verteilungsfunktion der Zufallsvariablen.

Satz 1 (Rechenregeln für Verteilungsfunktionen) Für alle $a, b \in \mathbb{R}$ mit $a < b$ gilt:

$$\begin{aligned} P(X < a) &= P(X \leq a) - P(X = a) = F(a) - P(X = a) \\ P(X > a) &= 1 - F(a) \\ P(X \geq a) &= 1 - F(a) + P(X = a) \\ P(a < X \leq b) &= F(b) - F(a) \\ P(a < X < b) &= F(b) - F(a) - P(X = b) \\ P(a \leq X < b) &= F(b) - F(a) - P(X = b) + P(X = a) \end{aligned}$$

Satz 2 (Eigenschaften jeder Verteilungsfunktion $F = F_X$)

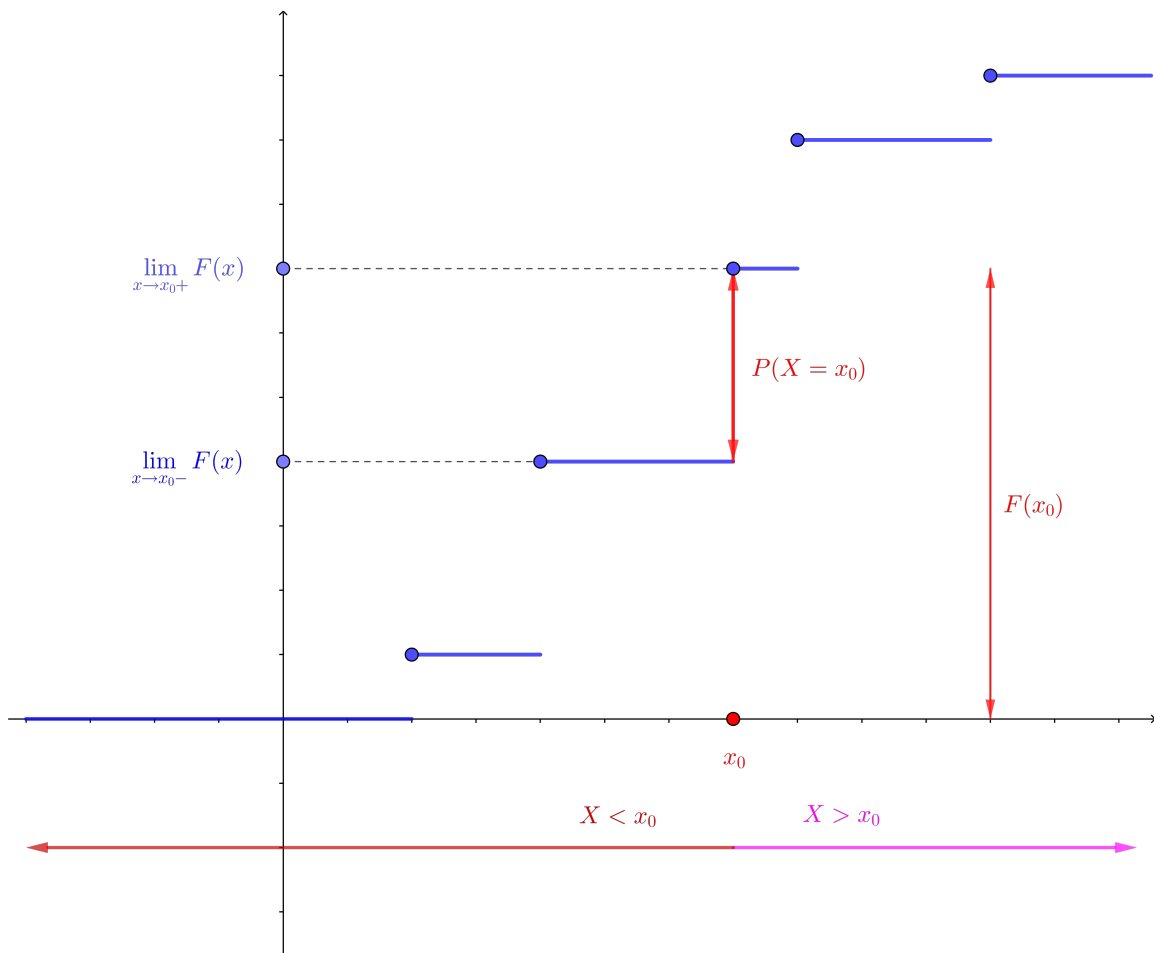
- F ist monoton wachsend, nicht notwendigerweise streng monoton wachsend (diskrete Verteilung)
- In jedem Punkt x_0 existiert der rechts- und linksseitige Grenzwert

$$\lim_{x \rightarrow x_0^-} F(x) = P(X < x_0)$$

$$\lim_{x \rightarrow x_0^+} F(x) = P(X \leq x_0) = F(x_0)$$

- $\lim_{x \rightarrow -\infty} F(x) = 0$
- $\lim_{x \rightarrow \infty} F(x) = 1$
- Wenn F in x_0 eine Sprungstelle hat, so ist die Sprunghöhe (also die Differenz aus rechts- und linksseitigem Grenzwert) genau die Wahrscheinlichkeit dass das Ereignis $X = x_0$ eintritt:

$$\lim_{x \rightarrow x_0^+} F(x) - \lim_{x \rightarrow x_0^-} F(x) = P(X \leq x_0) - P(X < x_0) = P(X = x_0).$$



Beispiel 1.1 Wir betrachten den zweifachen Wurf eines fairen Würfels und wissen:

$$\Omega = \{ (\omega_1, \omega_2) : \omega_i \in \{1, \dots, 6\} \}$$

und P ist die Gleichverteilung auf Ω . Wir definieren eine Zufallsvariable

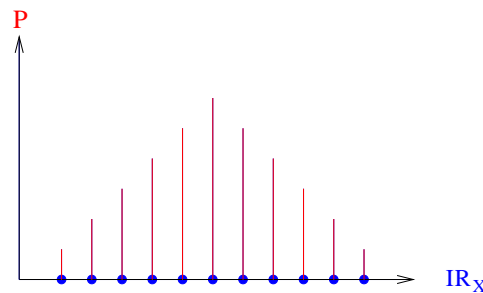
$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \\ (\omega_1, \omega_2) &\longmapsto \omega_1 + \omega_2 \end{aligned}$$

für alle $(\omega_1, \omega_2) \in \Omega$, d.h. X ist eine Funktion die jedem Ergebnis des Experimentes die Augensumme zuordnet. Es gilt zunächst $\mathbb{R}_X = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ und z.B.

$$P(X = 4) = P((3, 1) \text{ oder } (2, 2) \text{ oder } (1, 3)) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{3}{36}$$

Damit ergibt sich letztendlich die folgenden Verteilung für X :

x_i	2	3	4	5	6	7	8	9	10	11	12
$P(X = x_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$



$$F(x) = \begin{cases} 0 & , x < 2 \\ 1/36 & , 2 \leq x < 3 \\ 3/36 & , 3 \leq x < 4 \\ 6/36 & , 4 \leq x < 5 \\ 10/36 & , 5 \leq x < 6 \\ 15/36 & , 6 \leq x < 7 \\ 21/36 & , 7 \leq x < 8 \\ 26/36 & , 8 \leq x < 9 \\ 30/36 & , 9 \leq x < 10 \\ 33/36 & , 10 \leq x < 11 \\ 35/36 & , 11 \leq x < 12 \\ 36/36 & , 12 \leq x \end{cases}$$

Im folgenden unterscheiden wir zwei Typen von Zufallsvariablen

- diskrete Zufallsvariablen:

$$\mathbb{R}_X = X(\Omega) \text{ ist eine abzählbare Menge, z.B. } 0, 1, 2, \dots, 100$$

(siehe Vorlesung „Statistik“, Kapitel 8)

- stetige Zufallsvariablen:

$$\mathbb{R}_X = X(\Omega) \text{ ist eine überabzählbare Menge, z.B. } [0, 200].$$

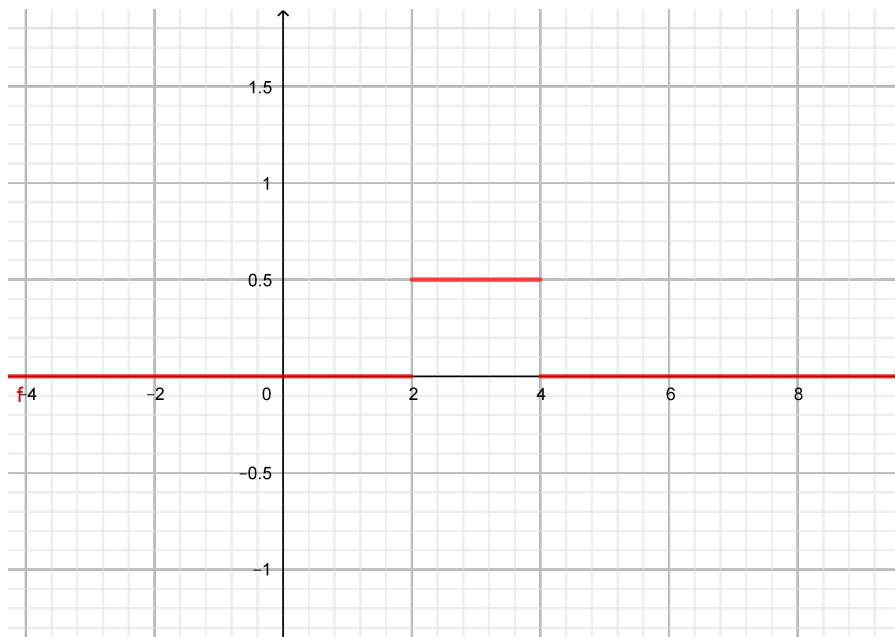
2 Stetige Zufallsvariablen

Definition 2.1 Eine Funktion f heisst Dichte oder Wahrscheinlichkeitsdichte falls sie die folgenden Eigenschaften hat:

1. $f(t) \geq 0$ für alle $t \in \mathbb{R}$,
2. $f(t)$ ist stetig bis auf abzählbar viele Punkte,
3. $\int_{-\infty}^{\infty} f(t) dt = 1$.

Beispiel 2.1 Die Funktion

$$f(t) = \begin{cases} \frac{1}{2} & t \in [2, 4] \\ 0 & \text{sonst} \end{cases}$$

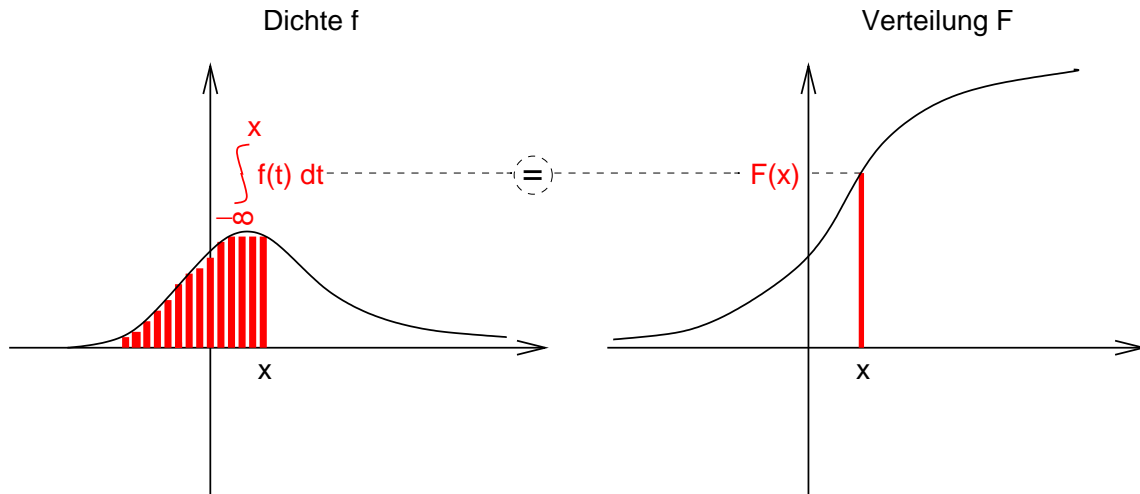


ist eine Wahrscheinlichkeitsdichte, denn

- sie nimmt keine negativen Werte an,
- sie ist fast überall stetig (ausser in den abzählbar vielen Punkten 2 und 4)
- der Flächeninhalt unter der Kurve ist genau 1 (da muss man nicht integrieren, aber man kann).

Definition 2.2 Eine Zufallsvariable X heisst stetig mit der Dichte f falls sich die Verteilungsfunktion $F : \mathbb{R} \rightarrow [0, 1]$ in der folgenden Weise schreiben lässt:

$$F(x) = P(-\infty < X \leq x) = \int_{-\infty}^x f(t) dt.$$



Merke: Der Flächeninhalt unter der Dichtefunktion (links von x) entspricht dem Funktionswert der Verteilungsfunktion an der Stelle x und dieser Flächeninhalt ist als die Wahrscheinlichkeit deutbar, mit der die Zufallsvariable X Werte nicht rechts von x realisiert.

F ist eine (spezielle) Stammfunktion von f und wie wir aus der Vorlesung Mathematik 1 bzw. aus der Schule wissen, gilt natürlich:

$$f(x) = F'(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h} = \lim_{h \rightarrow 0} \frac{1}{h} P(x \leq X \leq x+h)$$

Satz 3 Die Wahrscheinlichkeit, dass eine stetige Zufallsvariable einen beliebigen Wert x_0 annimmt, ist gleich Null:

$$P(X = x_0) = 0$$

Beweis:

Sei $x_0 \in \mathbb{R}$ und wir betrachten das Intervall $(x_0 - \delta, x_0]$. Dann gilt zunächst allgemein: $P(x_0 - \delta < X \leq x_0) = F(x_0) - F(x_0 - \delta)$ also

$$\begin{aligned} P(X = x_0) &= \lim_{\delta \rightarrow 0} P(x_0 - \delta < X \leq x_0) \\ &= \lim_{\delta \rightarrow 0} [F(x_0) - F(x_0 - \delta)] \\ &= F(x_0) - F(x_0) = 0. \end{aligned}$$

Bei stetigen Zufallsvariablen sind Punktereignisse $X = x_i$ nicht von Interesse!!

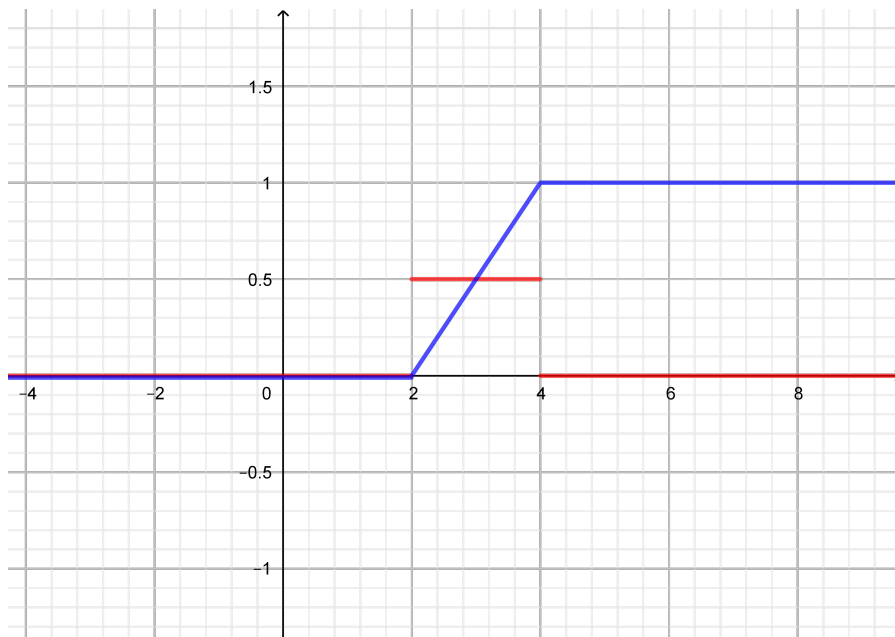
□

Beispiel 2.2 Sei wieder

$$f(t) = \begin{cases} \frac{1}{2} & t \in [2, 4] \\ 0 & \text{sonst} \end{cases}$$

Die zugehörige Verteilungsfunktion ist dann

$$F(x) = \begin{cases} 0 & x < 2 \\ \frac{1}{2}x - 1 & x \in [2, 4] \\ 1 & x > 4 \end{cases}$$



Zusammenfassung:

Wahrscheinlichkeit, dass X einen Wert zwischen a und b annimmt	$P(a \leq X \leq b)$ $= P(a < X < b)$ $= P(a \leq X < b)$
Ausgedrückt durch die Verteilungsfunktion	$= F(b) - F(a)$
Ausgedrückt durch die Dichte:	$= \int_a^b f(t) dt$

3 Erwartungswert, Varianz und Quantil

Definition 3.1 (Zur Erinnerung) X sei eine diskrete Zufallsvariable mit $\mathbb{R}_X = \{x_1, x_2, x_3, \dots, x_n\}$. Dann ist der Erwartungswert $E(X) = \mu$ von X definiert durch

$$E(X) = \sum_{i=1}^n x_i \underbrace{P(X = x_i)}_{p_i}$$

Die Varianz $Var(X) = \sigma^2$ von X (mit $\mu = E(X)$) ist:

$$Var(X) = \sum_{i=1}^n (x_i - \mu)^2 \underbrace{P(X = x_i)}_{p_i}$$

Definition 3.2 X sei eine stetige Zufallsvariable mit zugehöriger Dichte f . Dann ist der Erwartungswert $E(X) = \mu$ von X definiert durch

$$E(X) = \int_{-\infty}^{\infty} t f(t) dt$$

Die Varianz $Var(X) = \sigma^2$ von X (mit $\mu = E(X)$) ist:

$$Var(X) = \int_{-\infty}^{\infty} (t - \mu)^2 f(t) dt$$

Satz 4 (Rechenregeln für Erwartungswert und Varianz)

Seien X, Y Zufallsvariablen und a, b, c reelle Zahlen. Dann gelten die folgenden Rechenregeln:

1. $E(aX + bY + c) = a E(X) + b E(Y) + c$
2. $Var(aX + b) = a^2 Var(X)$
3. Verschiebungssatz der Varianz: $Var(X) = E(X^2) - [E(X)]^2$

Für manche Probleme ist es praktisch, den Begriff des Erwartungswertes weiter zu verallgemeinern.

Definition 3.3 X sei eine stetige Zufallsvariable mit zugehöriger Dichte f . Dann sind für $r = 1, 2, \dots$ die r -ten Momente definiert durch

$$m_r = E(X^r) = \int_{-\infty}^{\infty} t^r f(t) dt$$

Aufgabe 3.1 Die folgende Funktion sei die Wahrscheinlichkeitsdichte einer Zufallsvariablen X .

$$f(t) = \begin{cases} \frac{1}{2} & t \in [2, 4] \\ 0 & \text{sonst} \end{cases}$$

Bestimmen Sie Erwartungswert, die Varianz und die r -ten Momente von X .

Lösung:

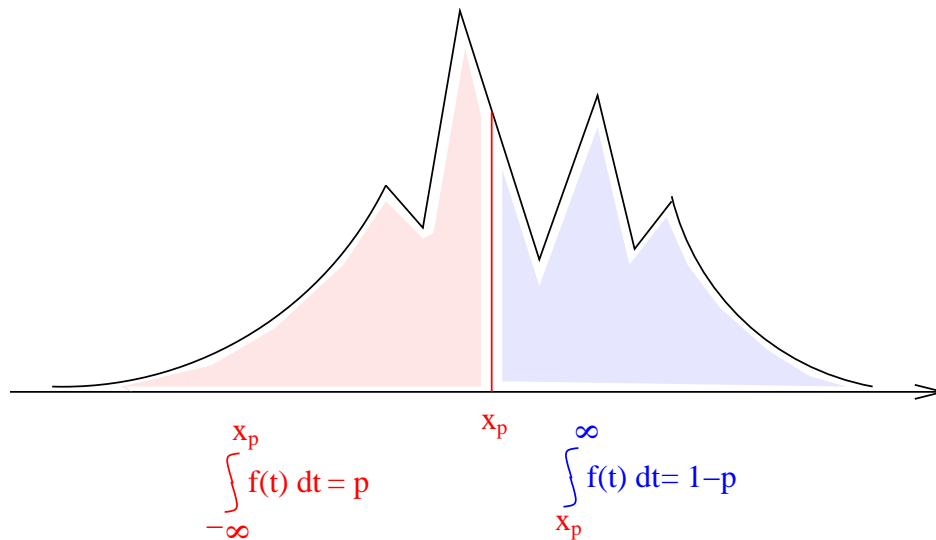
$$E(X) = \int_{-\infty}^{\infty} t f(t) dt = \int_2^4 t \frac{1}{2} dt = 3$$

$$\text{Var}(X) = \int_{-\infty}^{\infty} (t-3)^2 f(t) dt = \int_2^4 (t-3)^2 \frac{1}{2} dt = \frac{1}{3}$$

$$m_r = \int_{-\infty}^{\infty} t^r f(t) dt = \frac{2^r}{r+1} (2^{r+1} - 1)$$

Definition 3.4 Sei X eine stetige Zufallsvariable mit Verteilungsfunktion F . Ein p -Quantil x_p ($0 < p < 1$) ist der Wert der Verteilungsfunktion, für den $F(x_p) = p$ gilt.

Man kann sich also x_p als den Wert einer Verteilung vorstellen, der die Wahrscheinlichkeitsmasse so teilt, dass links von x_p genau die Masse p und rechts von x_p genau die Masse $1 - p$ liegt.



Spezialfälle: unteres Quartil: $x_{0,25}$, Median, Zentralwert: $x_{0,5}$, oberes Quartil: $x_{0,75}$

Aufgabe 3.2 Die folgende Funktion sei die Wahrscheinlichkeitsdichte einer Zufallsvariablen X .

$$f(t) = \begin{cases} \frac{1}{2} & t \in [2, 4] \\ 0 & \text{sonst} \end{cases}$$

Bestimmen Sie $x_{0,25}$, $x_{0,5}$ und $x_{0,75}$.

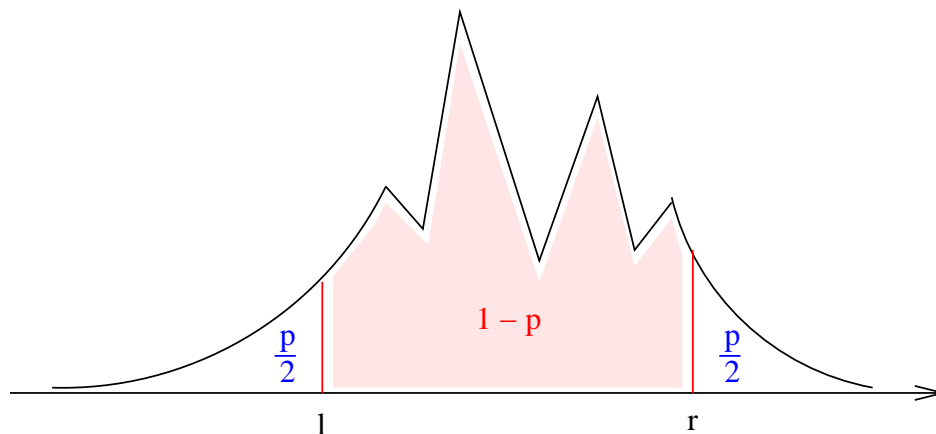
Lösung: Allgemein: $F(x_p) = \frac{1}{2}x_p - 1 = p$ oder $x_p = 2(p + 1)$

In der induktiven Statistik spielen p - und $(1 - p)$ -Quantile (zu $p = 0.025$, $p = 0.05$, ...) insbesondere in der Testtheorie eine wichtige Rolle.

Dort stellen diese Quantile Grenzen der Irrtumswahrscheinlichkeit dar, die die Grundlage für eine Testentscheidung bilden.

Aufgabe 3.3 Sei X eine stetige Zufallsvariable mit Dichtefunktion f . Weiterhin sei eine Zahl p mit $0 < p < 1$ gegeben. Bestimmen Sie zwei Zahlen l und r , so dass

$$P(l \leq X \leq r) = 1 - p$$



In dieser Form ist die Frage nicht eindeutig zu beantworten. Deshalb verlangen wir noch, dass die restliche Wahrscheinlichkeit gleichmässig auf die beiden Aussenbereiche verteilt wird.

Lösung: Die Forderung $P(l \leq X \leq r) = 1 - p$ ist mit der obigen Annahme gleichwertig zur Forderung:

$$p = P(-\infty \leq X \leq l) + P(r \leq X \leq \infty)$$

Verlangen wir nun eine gleichmässige Aufteilung der Wahrscheinlichkeit p auf diese beiden Aussenbereiche folgt:

•

$$\frac{p}{2} = P(-\infty \leq X \leq l) = F(l) \quad \text{oder} \quad l = x_{\frac{p}{2}} \quad \text{das } \frac{p}{2} - \text{Quantil}$$

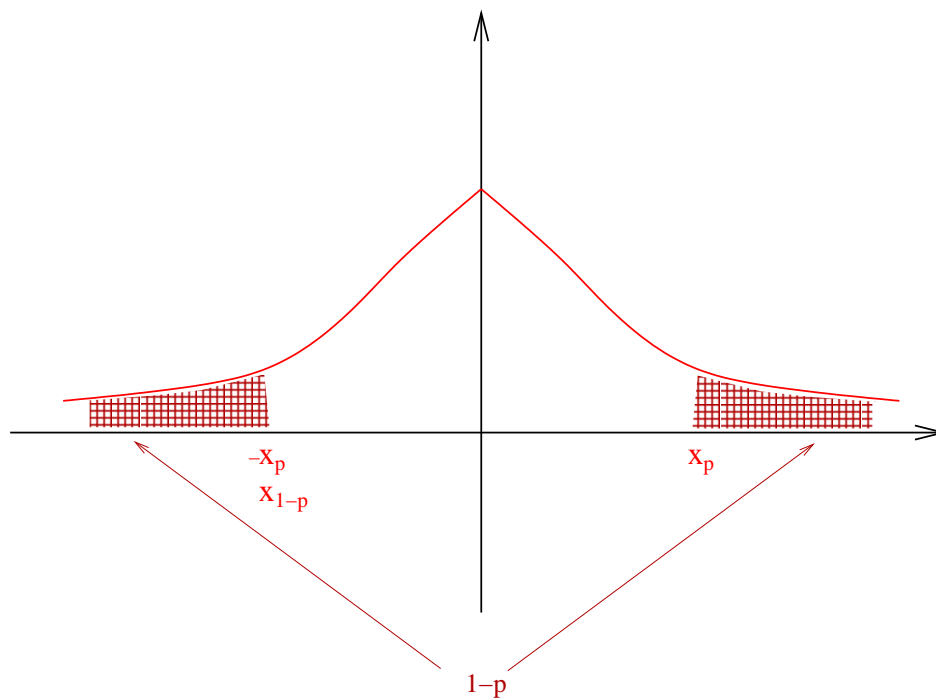
•

$$\frac{p}{2} = P(r \leq X \leq \infty) = 1 - F(r) \quad \text{oder} \quad r = x_{1-\frac{p}{2}} \quad \text{das } (1 - \frac{p}{2}) - \text{Quantil}$$

Falls X eine Zufallsvariable mit gerader Dichtefunktion f ist (d.h. es gilt $f(-t) = f(t)$ für alle t), so gilt stets

$$x_{1-p} = -x_p.$$

Das kann man (hoffentlich) leicht an folgender Zeichnung einsehen.



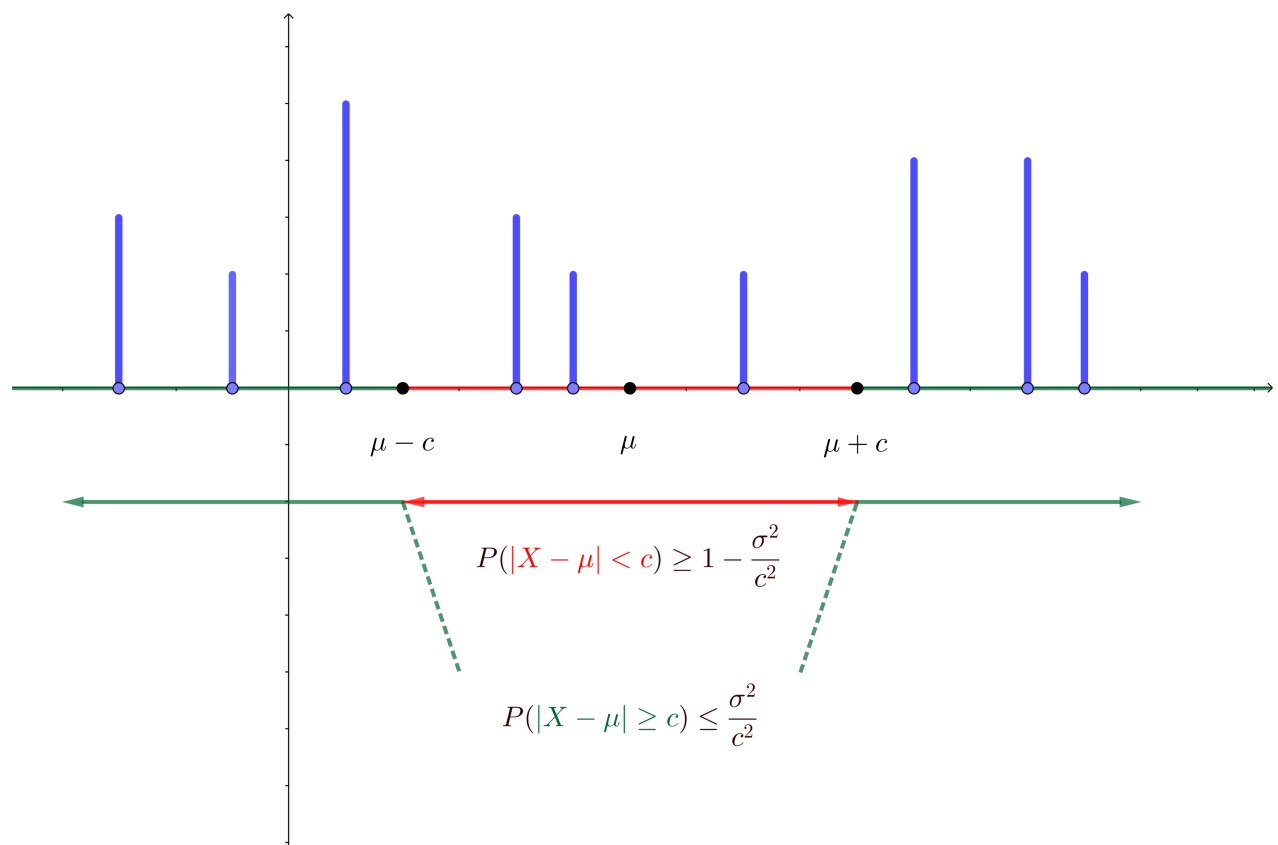
4 Die Ungleichung von Tschebyschev

Satz 5 (Die Ungleichung von Tschebyschev) Sei X eine beliebige Zufallsvariable mit $\mu = E(X)$ und $\sigma^2 = \text{Var}(X)$. Dann gelten für jede positive Zahl c die folgenden beiden (äquivalenten) Ungleichungen:

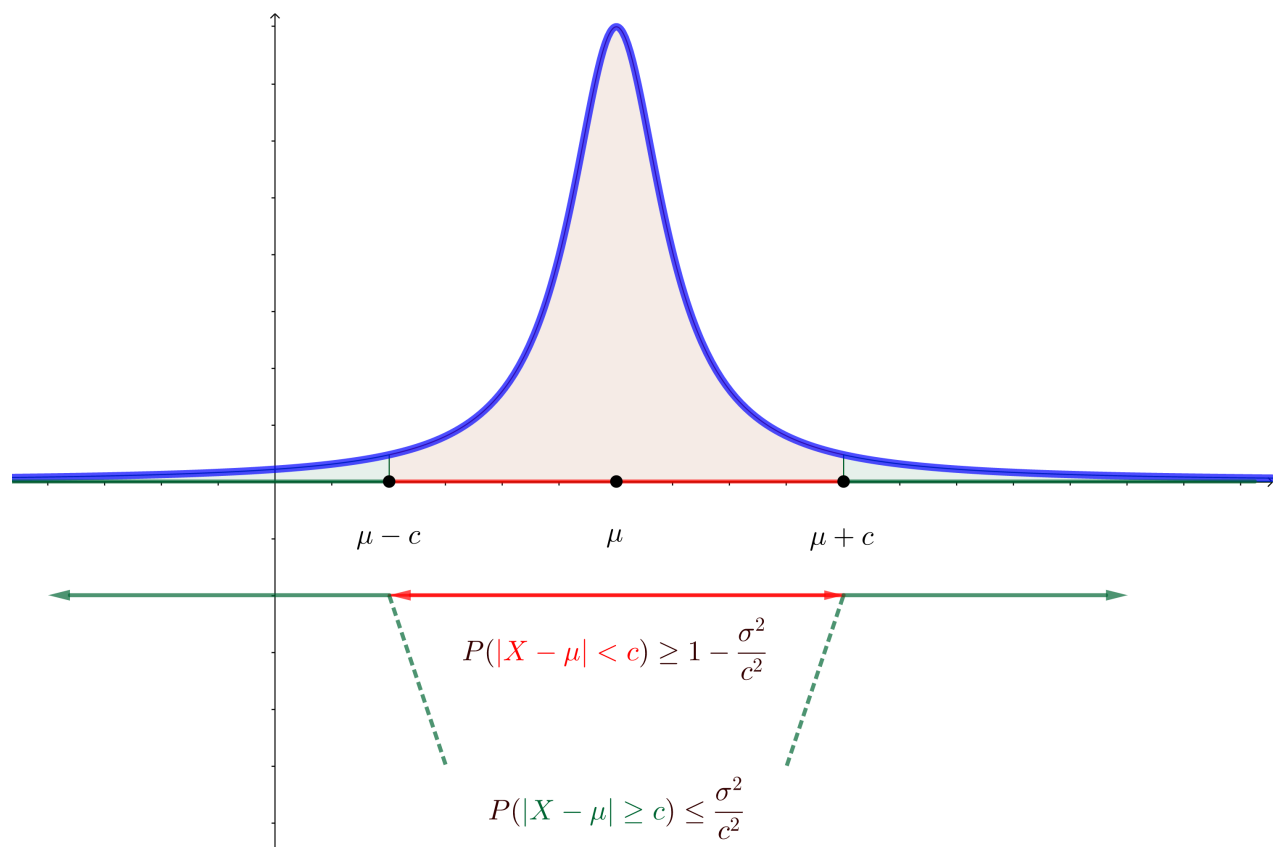
$$P(|X - \mu| < c) \geq 1 - \frac{\sigma^2}{c^2} \quad \text{oder} \quad P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

d.h. man kann relativ leicht die Wahrscheinlichkeit abschätzen, mit der X einen Wert innerherhalb (oder ausserhalb) des Intervalls $[\mu - c, \mu + c]$ annimmt.

Die Ungleichung von Tschebyschev für diskrete Zufallsvariablen



Die Ungleichung von Tschebyschev für stetige Zufallsvariablen

**Beweis:**

Wir beweisen die Ungleichung für stetige Zufallsvariablen. Der Beweis für diskrete Zufallsvariablen verläuft analog.

$$\begin{aligned}
 \sigma^2 &= \int_{-\infty}^{\infty} (t - \mu)^2 f(t) dt \\
 &\geq \int_{\substack{t \text{ s.d.} \\ |t - \mu| \geq c}} (t - \mu)^2 f(t) dt \\
 &\geq \int_{\substack{t \text{ s.d.} \\ |t - \mu| \geq c}} c^2 f(t) dt \\
 &= c^2 P(|X - \mu| \geq c)
 \end{aligned}$$

und auflösen nach $P(|X - \mu| \geq c)$ ergibt die Ungleichung von Tschebyschev für den Aussenbereich. \square

Aufgabe 4.1 Sei X eine stetige Zufallsvariable mit der Dichtefunktion

$$f(t) = \begin{cases} \frac{1}{2} & t \in [2, 4] \\ 0 & \text{sonst} \end{cases}$$

Wir wissen bereits, dass $E(X) = \mu = 3$ und $\text{Var}(X) = \sigma^2 = 1/3$ gilt. Schätzen Sie die Wahrscheinlichkeit $P(|X - \mu| < 1)$ mit Hilfe der Tschebyschev-Ungleichung. Können Sie die Wahrscheinlichkeit auch exakt bestimmen? (Selbstverständlich!)

Lösung: Es gilt nach einfacher Rechnung

$$P(|X - 3| < 1) \geq 1 - \frac{\sigma^2}{c^2} = 1 - \frac{1/3}{1} = \frac{2}{3},$$

was nicht falsch ist. Natürlich kennen wir das exakte Ergebnis

$$P(|X - 3| < 1) = P(2 < X < 4) = 1.$$

5 Standardisierung einer Zufallsvariablen

Definition 5.1 Eine Zufallsvariable heisst *standardisiert*, wenn sie den Erwartungswert 0 und die Varianz 1 hat. Sei X eine Zufallsvariable. Dann heisst die Zufallsvariable

$$Z = \frac{X - E(X)}{\sqrt{\text{Var}(X)}}$$

Standardisierung von X .

Ich erinnere hier an das folgende (diskrete) Beispiel aus dem Kurs Statistik.

Beispiel 5.1 Sei $X \sim B(2, 1/3)$, d.h. die Verteilung von X ist gegeben durch:

0	1	2
$\frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}$	$2 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{4}{9}$	$\frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$

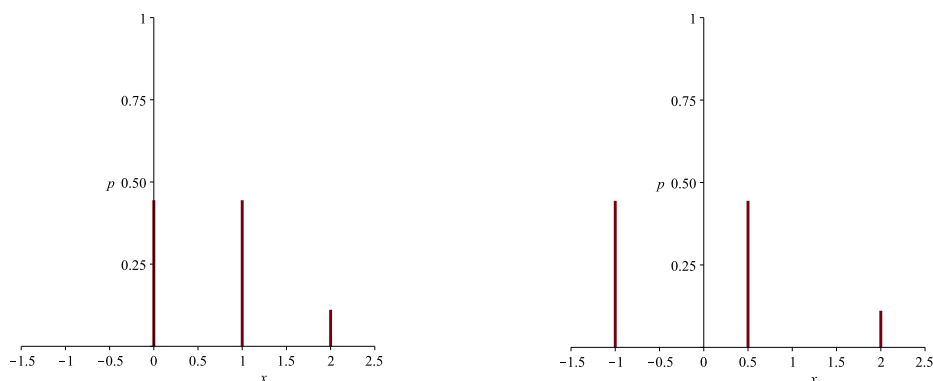
Es gilt: $E(X) = n \cdot p = \frac{2}{3}$, $\text{Var}(X) = n \cdot p \cdot (1 - p) = \frac{4}{9}$ und $\sqrt{\text{Var}(X)} = \frac{2}{3}$. Die Standardisierung von X ist somit

$$Z = \frac{X - 2/3}{2/3}$$

und hat die Verteilung:

$\frac{0-2/3}{2/3} = -1$	$\frac{1-2/3}{2/3} = \frac{1}{2}$	$\frac{2-2/3}{2/3} = 2$
$\frac{4}{9}$	$\frac{4}{9}$	$\frac{1}{9}$

Links sehen Sie die Verteilung von X und rechts die Verteilung der Standardisierung von X . Die drei Wahrscheinlichkeiten sind unverändert. Aber die Positionen der Wahrscheinlichkeiten haben sich so verändert, dass einerseits der Erwartungswert (Schwerpunkt) die 0 ist und andererseits die Varianz (Streuung) gleich 1. Man könnte sagen, dass sich die Verteilung **qualitativ nicht** verändert hat.



Die Verteilung der Standardisierung einer **diskreten** Zufallsvariablen ist einfach zu konstruieren. Etwas komplizierter sieht das bei **stetigen** Zufallsvariablen aus, die über ihre Dichte- oder Verteilungsfunktion beschrieben sind. Der folgende Satz zeigt, wie man in diesem Fall die Dichtefunktion der Standardisierung konstruieren muss.

Satz 6 *Hat die stetige Zufallsvariable X mit $\mu = E(X)$ und $\sigma = \sqrt{\text{Var}(X)}$ die Dichtefunktion f_X und die Verteilungsfunktion F_X , so hat $Z = \frac{X-\mu}{\sigma}$ die Dichtefunktion $f_Z(x) = \sigma \cdot f_X(\mu + \sigma \cdot x)$.*

Beweis: Es gilt

$$F_Z(x) = P\left(Z = \frac{X - \mu}{\sigma} \leq x\right) = P(X \leq \mu + \sigma \cdot x) = F_X(\mu + \sigma \cdot x).$$

Durch Ableiten (Kettenregel) erhält man die Dichtefunktion.

$$f_Z(x) = \frac{d}{dx}F_Z(x) = \frac{d}{dx}F_X(\mu + \sigma \cdot x) = \sigma \cdot f_X(\mu + \sigma \cdot x).$$

□

6 Übungsaufgaben

1. Eine stetige Zufallsvariable X besitze folgende Verteilungsfunktion

$$F(x) = \begin{cases} 0 & \text{für } x < 2 \\ -\frac{1}{4}x^2 + 2x - 3 & \text{für } 2 \leq x \leq 4 \\ 1 & \text{für } x > 4 \end{cases} .$$

- (a) Bestimmen Sie die Dichtefunktion der Zufallsvariablen X .
 (b) Bestimmen Sie den Erwartungswert $E(X)$ und die Varianz $Var(X)$.

2. Gegeben sei die Funktion

$$f(t) = \begin{cases} 2t - 4 & \text{für } 2 \leq t \leq 3 \\ 0 & \text{sonst} \end{cases} .$$

- (a) Zeigen Sie, dass f eine Dichtefunktion ist.
 (b) Ermitteln Sie die zugehörige Verteilungsfunktion.
 (c) Bestimmen Sie $x_{0.25}$, $x_{0.5}$ und $x_{0.75}$.

3. (a) X sei eine stetige Zufallsvariable mit folgender Dichtefunktion

$$f(t) = \begin{cases} ct & \text{für } 1 \leq t \leq 3 \\ 0 & \text{sonst} \end{cases} .$$

Für welchen Wert der Konstanten c ist die oben genannte Funktion tatsächlich eine korrekte Dichtefunktion?

- (b) Setzen Sie den entsprechenden Wert für c ein und bestimmen Sie $P(X > 2)$.

4. Von einer stetigen Zufallsvariablen X sei nur bekannt, dass sie den Erwartungswert 15 und die Varianz 4 besitzt.

- (a) Wie gross ist $P(10 \leq X \leq 20)$ mindestens?
 (b) Bestimmen Sie das kleinste, symmetrisch um 15 gelegene Intervall der Form $[15 - c, 15 + c]$, in welches mit einer Wahrscheinlichkeit von mindestens 0.9 die Werte von X fallen.