Algorithmen und Wahrscheinlichkeit

Verteilungen

Bernoulli-Verteilung

$$X \sim \mathrm{Bernoulli}(p)$$

Dichtefunktion:

$$f_X(x) = \begin{cases} p, & x = 1, \\ 1-p, & x = 0, \\ 0, & \mathrm{sonst} \end{cases} \qquad \mathbb{E}[X] = p$$

Beispiele: Werfen einer Münze, Indikator für Kopf
Werfen eines Würfels, Indikator für “Augenzahl gerade”
Werfen eines Würfels, Indikator für “1”

Binomialverteilung

$$X \sim \mathrm{Bin}(n, p).$$

$$f_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x \in \{0, 1, \ldots, n\} \\ 0, & \text{sonst.} \end{cases}$$

$$\mathbb{E}[X] = np$$

Beispiel: Werfen einer Münze n mal, X = Anzahl Kopf

$$X \sim \mathrm{Po}(\lambda)$$

$$f_X(i) = \begin{cases} \frac{e^{-\lambda}\lambda^i}{i!} & \text{für } i \in \mathbb{N}_0 \\ 0 & \text{sonst.} \end{cases}$$

$$\mathbb{E}[X] = \lambda$$

$\boxed{\text{Bin}(n,\boldsymbol{\lambda}/n) \quad \text{konvergiert für } n \to \infty \quad \text{gegen} \quad \text{Po}(\boldsymbol{\lambda})}$

Beispiel: Modellierung seltener Ereignisse, zum Beispiel
X:= Anzahl Herzinfarkte in der Schweiz in der nächsten Stunde

$$X \sim \mathrm{Geo}(p)$$

$$f_X(i) = \begin{cases} p(1-p)^{i-1} & \text{für } i \in \mathbb{N}, \\ 0 & \text{sonst.} \end{cases}$$

$$F_X(n) = 1 - (1-p)^n \qquad \text{für alle } n=1,2,\ldots$$

$$\mathbb{E}[X] = 1/p$$

Beispiel: Wiederholtes Werfen einer Münze,
X = Anzahl Würfe bis zum ersten Mal Kopf

Geometrische Verteilung

Wichtige Eigenschaft: Gedächtnislosigkeit

Satz Ist $X \sim \mathrm{Geo}(p)$, so gilt für alle $s, t \in \mathbb{N}$:

$$\Pr[X \overset{=}{\cancel{\geq}} s + t \mid X > s] = \Pr[X \overset{=}{\cancel{\geq}} t].$$

Beweis: nachrechnen

Beispiel: Wahrscheinlichkeit im ersten Wurf Kopf zu bekommen ist identisch zur Wahrscheinlichkeit nach 1000 Fehlversuchen im 1001ten Wurf Kopf zu bekommen

Kapitel 2.6

Mehrere Zufallsvariablen

Bedingte Zufallsvariable

Ω Wahrscheinlichkeitsraum, A ⊆ Ω Ereignis mit Pr[A] > 0

Die bedingte Zufallsvariable X|A ist *dieselbe Funktion* wie X, aber der Definitionsbereich ist auf die Menge A eingeschränkt:

Zufallsvariable: X|A : A → ℝ

Dichtefunktion

$$f_{X|A} : \mathbb{R} \to [0,1], \quad x \mapsto \Pr[X = x \mid A]$$

Verteilungsfunktion

$$F_{X|A} : \mathbb{R} \to [0,1], \quad x \mapsto \Pr[X \leq x \mid A]$$

X ist **unabhängig** von A, falls $f_{X|A} = f_X$.

Erwartungswert

$$\mathbb{E}[X \mid A] := \sum_{x \in W_X} x \cdot \Pr[X = x \mid A] = \frac{1}{\Pr[A]} \sum_{\omega \in A} X(\omega) \cdot \Pr[\omega]$$

X, Y Zufallsvariablen

gemeinsame Dichte

$$f_{X,Y}(x, y) := \Pr[X = x, Y = y]$$

"Randdichte von X" (= Dichte von X)

$$f_X(x) = \sum_{y \in W_Y} f_{X,Y}(x, y)$$

(nach dem Additionssatz)

Definition Zufallsvariablen $X_1, \ldots, X_n$ heissen unabhängig genau dann, wenn für alle $(x_1, \ldots, x_n) \in W_{X_1} \times \ldots \times W_{X_n}$ gilt

$$\Pr[X_1 = x_1, \ldots, X_n = x_n] = \Pr[X_1 = x_1] \cdot \ldots \cdot \Pr[X_n = x_n].$$

Lemma Sind $X_1, \ldots, X_n$ unabhängige Zufallsvariablen und sind $S_1, \ldots, S_n \subseteq \mathbb{R}$ beliebige Mengen, dann gilt

$$\Pr[X_1 \in S_1, \ldots, X_n \in S_n] = \Pr[X_1 \in S_1] \cdot \ldots \cdot \Pr[X_n \in S_n].$$

Unabhängigkeit von Zufallsvariablen

Definition (Version 2) Zufallsvariablen $X_1, \ldots, X_n$ heissen unabhängig genau dann, wenn für alle $(x_1, \ldots, x_n) \in W_{X_1} \times \ldots \times W_{X_n}$ gilt

$$\Pr[X_1 = x_1, \ldots, X_n = x_n] = \Pr[X_1 = x_1] \cdot \ldots \cdot \Pr[X_n = x_n].$$

Alternativ:

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n)$$

für alle $(x_1, \ldots, x_n) \in W_{X_1} \times \ldots \times W_{X_n}$

Summe von Zufallsvariablen

Satz Für zwei unabhängige Zufallsvariablen X und Y sei $Z := X + Y$. Es gilt

$$f_Z(z) = \sum_{x \in W_X} f_X(x) \cdot f_Y(z - x).$$

Beweis: Additionssatz und Definition der Unabhängigkeit, siehe Skript

Summe von Zufallsvariablen

Satz Für zwei unabhängige Zufallsvariablen X und Y sei $Z := X + Y$. Es gilt

$$f_Z(z) = \sum_{x \in W_X} f_X(x) \cdot f_Y(z - x).$$

Folgerungen:

Poisson(λ_1) + Poisson(λ_2) = Poisson($\lambda_1+\lambda_2$)

Bin(n, p) + Bin(m, p) = Bin(n+m, p)

Satz 2.64 (Waldsche Identität). N und X seien zwei unabhängige Zufallsvariable, wobei für den Wertebereich von N gelte: $W_N \subseteq \mathbb{N}$. Weiter sei

$$Z := \sum_{i=1}^{N} X_i,$$

wobei $X_1, X_2, \ldots$ unabhängige Kopien von X seien. Dann gilt:

$$\mathbb{E}[Z] = \mathbb{E}[N] \cdot \mathbb{E}[X].$$

Beweis:

$$\mathbb{E}[Z] = \sum_{n \in W_N} \mathbb{E}[Z \mid N = n] \cdot \Pr[N = n]$$

$$\mathbb{E}[Z \mid N = n] = \mathbb{E}[X_1 + \ldots + X_n] = n \cdot \mathbb{E}[X],$$

$$\mathbb{E}[Z] = \sum_{n \in W_N} n \cdot \mathbb{E}[X] \cdot \Pr[N = n] = \mathbb{E}[X] \sum_{n \in W_N} n \cdot \Pr[N = n] = \mathbb{E}[X] \cdot \mathbb{E}[N].$$

Waldsche Identität

- werfe zunächst einen Würfel

  $X :=$ Augenzahl des Würfels

- werfe danach eine Münze und zwar so oft wie die Augenzahl des Würfels angibt

  $Y :=$ Anzahl Kopf

- aus Waldscher Identität folgt: $$E[Y] = E[X]*1/2 = 7/4$$

Kapitel 2.4.2 + 2.7

Varianz und Konzentration

Zufallsvariable - Erwartungswert

Definition Zu einer Zufallsvariablen X definieren wir den *Erwartungswert* $\mathbb{E}[X]$ durch

$$\mathbb{E}[X] := \sum_{x \in W_X} x \cdot \Pr[X = x]$$

Welche Schlüsse können wir ziehen,
wenn wir den Erwartungswert einer Zufallsvariablen kennen?

Zufällige Folgen

Zwei Folgen von K und Z der Länge 200,

- eine Folge entspricht dem Ergebnis von 200 Münzwürfen
- eine ist von mir erzeugt.

Welche Folge ist die zufällige?

Z K Z K Z K K K Z K K Z Z K K K Z Z Z K Z K Z Z K Z K Z Z Z K Z K K Z K K Z Z Z
K Z K K Z Z K K Z K Z K Z K Z K Z K Z Z K Z Z K Z K K Z Z Z K K K Z K Z Z Z K Z
K K Z K Z K Z K K Z K Z K Z K K K Z K Z Z Z Z Z Z Z Z Z K Z Z K K Z Z Z [illegible] K Z K
K K K Z K K K Z Z Z K K K K Z K K Z Z K K Z Z Z K K Z K Z K Z K Z K Z [illegible] K K K K
K Z Z Z Z K Z K K Z K K Z Z Z K K K K K K K Z Z Z K K Z Z K K Z K Z Z Z Z K K Z

Verhältnis Kopf : Zahl = 0. 980198 # Änderungen = 112

K Z Z Z K Z K K Z K Z K Z K Z K Z K Z Z K K Z Z K K Z K Z K K Z Z K Z K Z Z K Z
K Z K Z K Z K Z Z Z Z Z Z K Z K K K K K Z K Z Z K Z K Z Z K Z K K K K Z Z K K K
K K K Z K Z K K Z Z K Z K Z K Z Z Z K Z K Z K K Z K K K Z Z K Z Z Z K Z Z K K Z
Z K Z Z K Z K Z Z K Z K Z K Z K K Z K Z K Z Z K Z K K Z Z K K Z K Z Z Z K Z K Z
K K K Z K Z K K K Z Z K Z Z K Z K Z K K K K Z Z K Z K Z K Z K Z K K Z Z Z K K Z

Verhältnis Kopf : Zahl = 1.00000 # Änderungen = 133

Zufallsvariable - Erwartungswert

Definition Zu einer Zufallsvariablen X definieren wir den *Erwartungswert* $\mathbb{E}[X]$ durch

$$\mathbb{E}[X] := \sum_{x \in W_X} x \cdot \Pr[X = x]$$

Welche Schlüsse können wir ziehen,
wenn wir den Erwartungswert einer Zufallsvariablen kennen?

Was wir gerne hätten:

$$\Pr[\ |X - \mathbb{E}[X]|\ \text{"gross"}\ ] = \text{"klein"}$$

Abweichung vom Erwartungswert

Was wir gerne hätten:

$$\Pr[\ |X - \mathbb{E}[X]|\ \text{"gross"}\ ] = \text{"klein"}$$

Beobachtung: stimmt sicher nicht immer … ☹

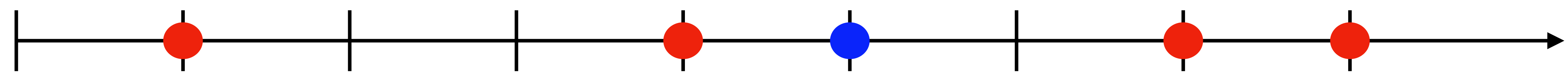
Beispiel:

$$\Pr[\ X = -10^{10}\ ] = \frac{1}{2} = \Pr[\ X = 10^{10}\ ]$$

Abweichung vom Erwartungswert

Was wir gerne hätten:

$$\Pr[\ |X - \mathbb{E}[X]|\ \text{"gross"}\ ] = \text{"klein"}$$



Laplaceraum mit $\omega_1 = 1$, $\omega_2 = 4$, $\omega_3 = 7$, $\omega_4 = 8$ Erwartungswert $\mu = 5$

Was wir messen wollen ist die durchschnittliche Abweichung vom Erwartungswert

zum Beispiel $\frac{1}{4}\sum_i |\omega_i - \mu|$ oder $\frac{1}{4}\sum_i (\omega_i - \mu)^2$

Varianz $\mathrm{Var}[X] = \mathbb{E}[(X - \mu)^2]$, wobei $\mu := \mathbb{E}[X]$

Definition Für eine Zufallsvariable X mit $\mu = \mathbb{E}[X]$ definieren wir die *Varianz* $\mathrm{Var}[X]$ durch

$$\mathrm{Var}[X] := \mathbb{E}[(X-\mu)^2] = \sum_{x \in W_X} (x-\mu)^2 \cdot \Pr[X = x].$$

Die Grösse $\sigma := \sqrt{\mathrm{Var}[X]}$ heisst *Standardabweichung* von X.

Satz Für eine beliebige Zufallsvariable X gilt

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

Beweis: $\mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2 - 2\mu X + \mu^2]$

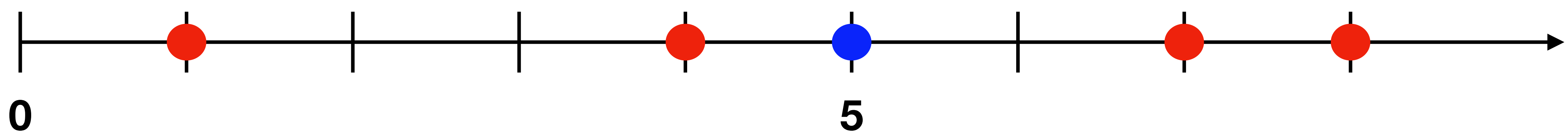
$= \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 = \mathbb{E}[X^2] - \mu^2$

Satz Für eine beliebige Zufallsvariable X und $a, b \in \mathbb{R}$ gilt

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

Beweis: Definition der Varianz und Linearität des Erwartungswertes, siehe Skript

Warum spielt das b keine Rolle?

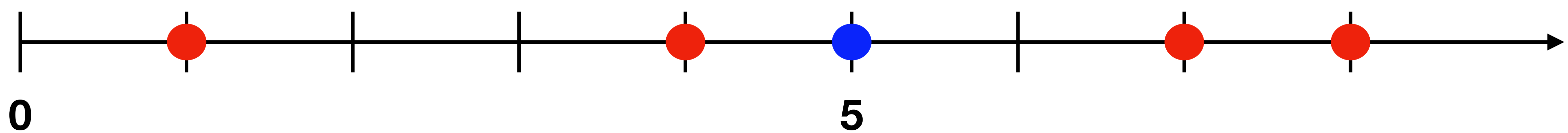


Satz Für eine beliebige Zufallsvariable X und $a, b \in \mathbb{R}$ gilt

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

Beweis: Definition der Varianz und Linearität des Erwartungswertes, siehe Skript

Warum spielt das b keine Rolle?



Abweichung vom Erwartungswert bleibt gleich!

Rechenregeln für Momente

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$? ✓

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$?

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$$?

$$\mathrm{Var}[X \cdot Y] = \mathrm{Var}[\mathrm{X}] \cdot \mathrm{Var}[Y]$$?

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y] \qquad \forall X, Y \text{ unabhängig}$$

$$\begin{aligned}
\mathbb{E}[X \cdot Y] &= \sum_{x \in W_X} \sum_{y \in W_Y} xy \cdot \Pr[X = x, Y = y] \\
&\overset{\text{Unabh.}}{=} \sum_{x \in W_X} \sum_{y \in W_Y} xy \cdot \Pr[X = x] \cdot \Pr[Y = y] \\
&= \sum_{x \in W_X} x \cdot \Pr[X = x] \sum_{y \in W_Y} y \cdot \Pr[Y = y] = \mathbb{E}[X] \cdot \mathbb{E}[Y].
\end{aligned}$$

Rechenregeln

Satz *(Multiplikativität des Erwartungswerts)* Für unabhängige Zufallsvariablen $X_1, \dots, X_n$ gilt

$$\mathbb{E}[X_1 \cdot \dots \cdot X_n] = \mathbb{E}[X_1] \cdot \dots \cdot \mathbb{E}[X_n].$$

Satz Für unabhängige Zufallsvariablen $X_1, \dots, X_n$ und $X := X_1 + \dots + X_n$ gilt

$$\mathrm{Var}[X] = \mathrm{Var}[X_1] + \dots + \mathrm{Var}[X_n].$$

Beweise: siehe Skript

Rechenregeln für Momente

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y] \qquad \forall X, Y$$

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y] \qquad \forall X, Y \text{ unabhängig}$$

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] \qquad \forall X, Y \text{ unabhängig}$$

$$\mathrm{Var}[X \cdot Y] \neq \mathrm{Var}[\mathrm{X}] \cdot \mathrm{Var}[Y] \qquad \text{i.A.}$$

(auch für unabhängige ZV)

Definition Zu einer Zufallsvariablen X definieren wir den *Erwartungswert* $\mathbb{E}[X]$ durch

$$\mathbb{E}[X] := \sum_{x \in W_X} x \cdot \Pr[X = x]$$

Welche Schlüsse können wir ziehen,

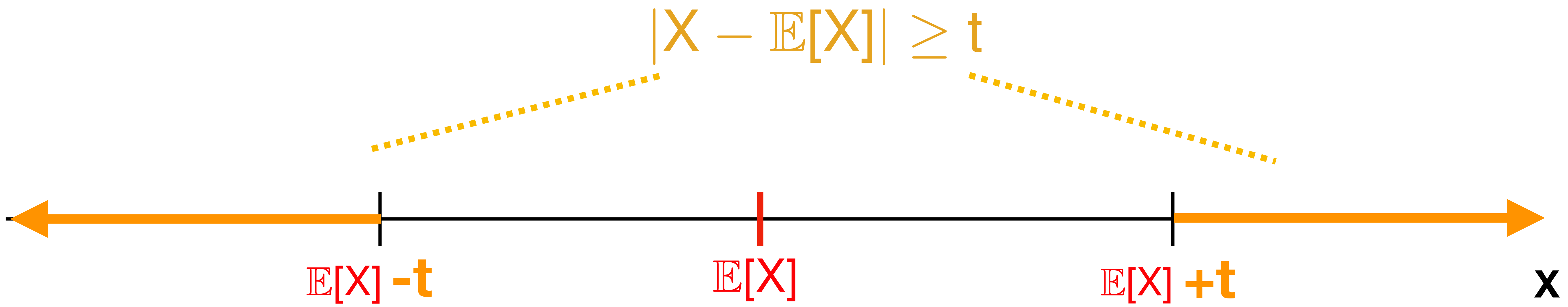
wenn wir den Erwartungswert einer Zufallsvariablen kennen?

Was wir gerne hätten:

$$\Pr[\ |X - \mathbb{E}[X]|\ \text{"gross"}\ ] = \text{"klein"}$$

…gilt wenn Varianz „klein“

X Zufallsvariable

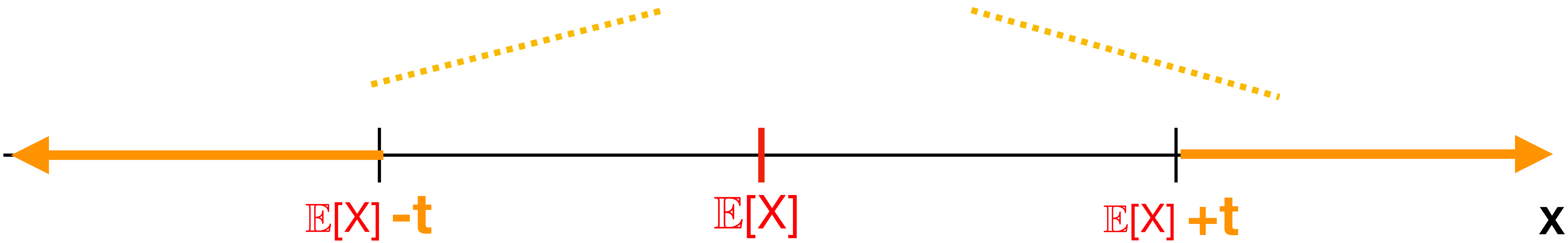


$$\Pr[|X - \mathbb{E}[X]| \geq t] \;\leq\; ???$$

Was wir zeigen werden …

X Zufallsvariable

$$\Pr[|X - \mathbb{E}[X]| \geq t]$$



Chebyshev:

$$\Pr[|X - \mathbb{E}[X]| \geq t] \leq \frac{\mathrm{Var}[X]}{t^2} \qquad \forall X, \quad \forall t > 0$$

Markov Ungleichung

Satz *(Ungleichung von Markov)* Sei X eine Zufallsvariable, die nur nicht-negative Werte annimmt. Dann gilt für alle $t \in \mathbb{R}$ mit $t > 0$, dass

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}.$$

Beweis:

$$\mathbb{E}[X] \overset{\text{Def}}{=} \sum_{x \in W_X} x \cdot \Pr[X = x]$$

$$\geq \cancel{\sum_{x \in W_X,\, x<t} x \cdot \Pr[X = x]} + \sum_{x \in W_X,\, x \geq t} \overset{t}{\cancel{x}} \cdot \Pr[X = x]$$

$$\mathbb{E}[X] \geq t \cdot \sum_{x \in W_X,\, x \geq t} \Pr[X = x] = t \cdot \Pr[X \geq t]$$

Satz *(Ungleichung von Chebyshev)* Sei X eine Zufallsvariable und $t \in \mathbb{R}$ mit $t > 0$. Dann gilt

$$\Pr[|X - \mathbb{E}[X]| \geq t] \;\leq\; \frac{\mathrm{Var}[X]}{t^2}.$$

Beweis:

$$|X - \mathbb{E}[X]| \geq t \iff \underbrace{(X - \mathbb{E}[X])^2}_{Y :=} \geq t^2$$

dann gilt $Y \geq 0$ (da Y ein Quadrat ist)

daher können wir die Markov-Ungleichung anwenden:

$$\Pr[|X - \mathbb{E}[X]| \geq t] = \Pr[Y \geq t^2] \leq \frac{\mathbb{E}[Y]}{t^2} = \frac{\mathrm{Var}[X]}{t^2}$$

Markov:

$$\Pr[X \geq t] \;\leq\; \frac{\mathbb{E}[X]}{t} \qquad \forall\, X \geq 0, \quad \forall\, t > 0$$

Chebyshev:

$$\Pr[|X - \mathbb{E}[X]| \geq t] \;\leq\; \frac{\mathrm{Var}[X]}{t^2} \qquad \forall\, X, \quad \forall\, t > 0$$

Insbesondere: ($\sigma := \sqrt{\mathrm{Var}[X]}$ Standardabweichung)

Für $t := \overset{C}{\cancel{10}}\sigma$

$$\Pr[|X - \mathbb{E}[X]| \geq \overset{C}{\cancel{10}}\sigma] \;\leq\; \frac{\mathrm{Var}[X]}{(\underset{C}{\cancel{10}}\sigma)^2} = \frac{1}{\underset{C^2}{\cancel{100}}} \qquad \forall\, X, \quad \forall\, C > 0$$

Zufällige Folgen

Z K Z K Z K K K Z K K Z Z K K K Z Z Z K Z K Z Z K Z K Z Z Z K Z K K Z K K Z Z Z
K Z K K Z Z K K Z K Z K Z K Z K Z K Z Z K Z Z K Z K K Z Z Z K K K Z K Z Z [illegible] K Z
K K Z K Z K Z K K Z K Z K Z K K K Z K Z Z Z Z Z Z Z Z Z K Z Z K K Z Z Z [illegible] K Z K
K K K Z K K K Z Z Z K K K K Z K K Z Z K K Z Z Z K K Z K Z K Z K Z K Z [illegible] K K K K
K Z Z Z Z K Z K K Z K K Z Z Z K K K K K K K Z Z Z K K Z Z K K Z K Z Z Z Z K K Z

Verhältnis Kopf : Zahl = 0. 980198 # Änderungen = 112

$$C\sigma \stackrel{!}{=} 12 \quad \Longleftrightarrow \quad C = \frac{12}{\sigma} \approx 1.7, \qquad \frac{1}{C^2} \approx 0.34$$

K Z Z Z K Z K K Z K Z K Z K Z K Z K Z Z K K Z Z K K Z K Z K K Z Z K Z K Z Z K Z
K Z K Z K Z K Z Z Z Z Z Z K Z K K K K K Z K Z Z K Z K Z Z K Z K K K K Z Z K K K
K K K Z K Z K K Z Z K Z K Z K Z Z Z K Z K Z K K Z K K K Z Z K Z Z Z K Z Z K K Z
Z K Z Z K Z K Z Z K Z K Z K Z K K Z K Z K Z Z K Z K K Z Z K K Z K Z Z Z K Z K Z
K K K Z K Z K K K Z Z K Z Z K Z K Z K K K K Z Z K Z K Z K Z K Z K K Z Z Z K K Z

Verhältnis Kopf : Zahl = 1.00000 # Änderungen = 133

$$C\sigma \stackrel{!}{=} 33 \quad \Longleftrightarrow \quad C = \frac{33}{\sigma} \approx 4.7, \qquad \frac{1}{C^2} \approx 0.044$$

$$\Pr[|X - \mathbb{E}[X]| \geq C\sigma] \;\leq\; \frac{1}{C^2}$$

100

$\forall X, \ \forall C > 0$

$\sigma := \sqrt{\mathrm{Var}[X]} \approx 7$