Einführung in die Wahrscheinlichkeitstheorie und Statistik
Wintersemenster 2022/23
Dozent: Prof. Dr. Rainer Dahlhaus
Assistenz: Marilena Müller

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

Lösungsskizen zu Übungsblatt 12

Aufgabe 1

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen.

(a) Zeigen Sie, dass das schwache Gesetz der großen Zahlen aus dem zentralen Grenzwertsatz folgt.

(b) X_n besitze die Wahrscheinlichkeitsdichte $f_n(x) = \frac{n+1}{2}|x|^n \mathbb{1}_{(-1,1)}(x)$, wobei $x \in \mathbb{R}$. Gibt es eine Zufallsvariable Z mit $X_n \xrightarrow{D} Z$? Geben Sie Z bei Existenz an.

(c) Entscheiden Sie jeweils, ob eine Zufallsvariable Z existiert, sodasst $X_n \xrightarrow{D} Z$, wobei

(i) $X_n \sim \mathcal{R}[0, 1 + \frac{1}{n}]$,

(ii) $X_n \sim \mathcal{E}(n)$,

(iii) $X_n \sim \mathcal{E}(1/n)$.

(1+2+2) Punkte)

Beweis. (a) Sind die $(X_n)_{n\in\mathbb{N}}$ eine iid Folge von Zufallsvariablen mit $\operatorname{Var}(X_1) < \infty$, so gilt nach dem zentralen Grenzwertsatz

$$\sqrt{n}\frac{\overline{X_n} - \mathbb{E}[X_1]}{\sqrt{\operatorname{Var}(X_1)}} \xrightarrow{D} \mathcal{N}(0,1)$$

bzw.

$$\sqrt{n}(\overline{X_n} - \mathbb{E}[X_1]) \xrightarrow{D} \mathcal{N}(0, \operatorname{Var}(X_1)).$$

Es gilt $\frac{1}{\sqrt{n}} \xrightarrow{\mathbb{P}} 0$. Nach Proposition 14.9(ii) folgt

$$\overline{X_n} - \mathbb{E}[X_1] = \frac{1}{\sqrt{n}} \cdot \sqrt{n}(\overline{X_n} - \mathbb{E}[X_1]) \xrightarrow{\mathbb{P}} 0.$$

Damit folgt $\overline{X_n} \xrightarrow{\mathbb{P}} \mathbb{E}[X_1]$, d.h. das schwache Gesetz der großen Zahlen.

(b) Um die schwache Konvergenz zu untersuchen, müssen wir zunächst die Verteilungsfunktion von X_n bestimmen. Für $x \in (-1,1)$ gilt

$$F_n(x) = \int_{-\infty}^x f_n(y)\mathrm{d}y = \frac{n+1}{2}\int_{-1}^x |y|^n \mathrm{d}y.$$

Im Falle $x \in (-1, 0]$ folgt nun

$$\begin{aligned} F_n(x) &= \frac{n+1}{2}(-1)^n \int_{-1}^x y^n \mathrm{d}y = \frac{(-1)^n}{2}\Big[y^{n+1}\Big]_{-1}^x \\ &= \frac{(-1)^n}{2}\Big(x^{n+1} - (-1)^{n+1}\Big) = \frac{1}{2}(1 - (-x)^{n+1}). \end{aligned}$$

Insbesondere folgt daher $F_n(0) = \frac{1}{2} = \frac{n+1}{2}\int_{-1}^{0} |y|^n \mathrm{d}y$.
Für $x \in (0,1)$ gilt

$$\begin{aligned} F_n(x) &= \frac{n+1}{2}\int_{-1}^{x} |y|^n \mathrm{d}y = \frac{n+1}{2}\int_{-1}^{0} |y|^n \mathrm{d}y + \frac{n+1}{2}\int_{0}^{x} y^n \mathrm{d}y \\ &= \frac{1}{2} + \frac{1}{2}\Big[y^{n+1}\Big]_0^x = \frac{1}{2}(1 + x^{n+1}). \end{aligned}$$

Alternativ können wir für $x \in (0,1)$ folgendes Argument verwenden:
Da f_n achsensymmetrisch zum Ursprung ist, wissen wir, dass $F_n(x) = 1 - F_n(-x)$ gelten muss. Daraus kann das Ergebnis für $x \in (0,1)$ direkt aus dem Ergebnis für $x \in (-1,0)$ abgelesen werden.

Wir weisen die schwache Konvergenz nach. Um zu ermitteln, ob X_n schwach konvergiert, untersuchen wir den punktweisen Limes von F_n für $n \to \infty$. Hier ist

$$F_n(x) = \begin{cases} 0, & x \le -1, \\ \frac{1}{2}(1 - (-x)^{n+1}), & x \in (-1,0], \\ \frac{1}{2}(1 + x^{n+1}), & x \in (0,1), \\ 1, & x \ge 1 \end{cases} \to \begin{cases} 0, & x \le -1, \\ \frac{1}{2}, & x \in (-1,1), \\ 1, & x \ge 1 \end{cases} =: \quad \tilde{F}(x).$$

$\tilde{F}$ ist noch keine Verteilungsfunktion, da sie nicht rechtsseitig stetig ist. Wir können $\tilde{F}$ aber an der Unstetigkeitsstellen $x = -1$ so modifizieren, dass sie rechtsseitig stetig und damit zur Verteilungsfunktion wird. Definiere

$$F(x) := \begin{cases} 0, & x < -1, \\ \frac{1}{2}, & x \in [-1,1), \\ 1, & x \ge 1. \end{cases}$$

F ist eine Verteilungsfunktion und es gilt $F_n(x) \to F(x)$ für alle Stetigkeitspunkte x von F (d.h. für alle $x \in \mathbb{R}\backslash\{-1,1\}$). Damit ist gezeigt, dass $X_n \xrightarrow{D} Z$, wobei Z eine Zufallsvariable mit Verteilungsfunktion F ist.
Genauer ist Z eine diskrete Zufallsvariable mit $\mathbb{P}(Z = -1) = \mathbb{P}(Z = 1) = \frac{1}{2}$.

(c) Wir ermitteln nun wie in (b) jeweils die Verteilungsfunktion F_n von X_n und deren Limes $\tilde{F}$, und nehmen eventuell Modifikationen vor, damit $\tilde{F}$ zur Verteilungsfunktion wird.

(i) Hier ist die Dichte $f_n(x) = \frac{1}{1+\frac{1}{n}}\mathbb{1}_{[0,1+\frac{1}{n}]}(x)$ und

$$F_n(x) = \begin{cases} 0, & x < 0, \\ \frac{x}{1+\frac{1}{n}}, & 0 \le x \le 1 + \frac{1}{n}, \\ 1, & x > 1 + \frac{1}{n} \end{cases} \to \begin{cases} 0, & x < 0, \\ x, & 0 \le x \le 1, \\ 1, & x > 1 \end{cases} =: \tilde{F}(x).$$

$\tilde{F}$ ist an der Stelle $x = 1$ nicht rechsseitig stetig. Wir modifizieren daher zu

$$F(x) := \begin{cases} 0, & x < 0, \\ x, & 0 \le x < 1, . \\ 1, & x \ge 1 \end{cases}$$

Dann ist F rechtsseitig stetig (es ist die Verteilungsfunktion einer $U[0,1]$-Verteilung) und es gilt $F_n(x) \to F(x)$ für alle Stetigkeitspunkte x von F. Das bedeutet, es gibt $Z \sim U[0,1]$ mit $X_n \xrightarrow{D} Z$.

(ii) Hier ist $F_n(x) = \left(1 - e^{-nx}\right)\mathbb{1}_{[0,\infty)}(x)$. Es gilt

$$F_n(x) = \left(1 - e^{-nx}\right)\mathbb{1}_{[0,\infty)}(x) \to \mathbb{1}_{(0,\infty)}(x) =: \tilde{F}(x).$$

$\tilde{F}$ ist an der Stelle $x = 0$ nicht rechtsseitig stetig. Wir modifizieren zu $F(x) := \mathbb{1}_{[0,\infty)}(x)$. Dann ist F rechtsseitig stetig (es ist die Verteilungsfunktion einer konstante Zufallsvariable $Z = 0$). Das bedeutet $X_n \xrightarrow{D} 0$.

(iii) Hier ist $F_n(x) = \left(1 - e^{-x/n}\right)\mathbb{1}_{[0,\infty)}(x)$. Es gilt

$$F_n(x) = \left(1 - e^{-x/n}\right)\mathbb{1}_{[0,\infty)}(x) \to 0 =: \tilde{F}(x).$$

Angenommen, es gäbe eine Zufallsvariable Z mit $X_n \xrightarrow{D} Z$. Dann müsste auch $F_n(x) \to 0 = F_Z(x)$ für alle Stetigkeitspunkte x von F_Z gelten.
Da F_Z nur abzählbar viele Unstetigkeitsstellen hat, gilt sicher $F_Z(x) = 0$ für beliebig große $x \in \mathbb{R}$. Aufgrund der Monotonie von F_Z folgt dann aber schon $F_Z \equiv 0$. Das ist ein Widerspruch zu der Eigenschaft $\lim_{x\to\infty} F_Z(x) = 1$ einer Verteilungsfunktion. Daher kann X_n nicht schwach konvergieren.

□

Aufgabe 2

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen und X eine weitere Zufallsvariable. Weiter sei $c \in R$ eine deterministische Konstante.

(a) Sei $X \sim \mathcal{B}(1, \frac{1}{2})$ Bernoulli-verteilt mit Erfolgswahrscheinlichkeit $\frac{1}{2}$, und $X_n := 1 - X$ für alle $n \in \mathbb{N}$. Zeigen Sie, dass $X_n \xrightarrow{D} X$ aber nicht $X_n \xrightarrow{\mathbb{P}} X$.

(b) Nun gelte $X_n \xrightarrow{D} c$. Zeigen Sie, dass dann sogar $X_n \xrightarrow{\mathbb{P}} c$ folgt.
Hinweis: Drücken Sie $\mathbb{P}(|X_n - c| > \varepsilon)$ mit der Verteilungsfunktion F_{X_n} von X_n aus.

(c) *Die Delta-Methode:* Sei $(a_n)_{n\in\mathbb{N}}$ eine Folge von positiven reellen Zahlen mit $a_n \to \infty$. Sei Z eine Zufallsvariable, sodass $a_n(X_n - c) \xrightarrow{D} Z$. Sei $g : \mathbb{R} \to \mathbb{R}$ eine in einer Umgebung von c stetig differenzierbare Funktion mit $g'(c) \neq 0$. Zeigen Sie, dass dann folgt:

$$a_n(g(X_n) - g(c)) \xrightarrow{D} g'(c) \cdot Z.$$

Hinweis: Nutzen Sie eine Taylor-Entwicklung von g um c: Für $x \in \mathbb{R}$ und ein $\tilde{x} \in \mathbb{R}$ mit $|\tilde{x} - c| \leq |x - c|$ gilt $g(x) = g(c) + (x - c)g'(c) + (x - c) \cdot (g'(\tilde{x}) - g'(c))$. Nutzen Sie Proposition 14.9 der Vorlesung.

(1+2+2) Punkte)

Beweis. (i) Die Zufallsvariablen $X_n = 1 - X$ besitzen dieselbe Verteilung wie die Zufallsvariable X, denn

$$\mathbb{P}((1 - X) = 0) = \mathbb{P}(X = 1) = \frac{1}{2} = \mathbb{P}(X = 0),$$
$$\mathbb{P}((1 - X) = 1) = \mathbb{P}(X = 0) = \frac{1}{2} = \mathbb{P}(X = 1).$$

Daher gilt $F_{X_n}(x) = F_X(x)$ für alle $x \in \mathbb{R}$ und insbesondere

$$F_{X_n}(x) \overset{n\to\infty}{\to} F_X(x) \quad \text{für alle } x \in \mathbb{R}$$

Damit gilt $X_n \xrightarrow{D} X$. Aber für $\varepsilon = \frac{1}{2}$ gilt

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}\left(|1 - 2X| > \frac{1}{2}\right) = \mathbb{P}(X = 0, X = 1) = 1 \not\to 0$$

für $n \to \infty$, weil $\mathrm{Bild}(X) = \{0, 1\}$.

(ii) Seien F_{X_n} die Verteilungsfunktionen der X_n und F_c die Verteilungsfunktion der konstanten Zufallsvariable c. Dann gilt

$$F_c(x) = \mathbb{P}(c \leq x) = \begin{cases} 0, & x < c \\ 1, & x \geq c \end{cases}.$$

Wegen $X_n \xrightarrow{D} c$ folgt dann $F_{X_n}(x) \xrightarrow{n\to\infty} F_c(x)$ für alle Stetigkeitspunkte x von F_c, d.h. für alle $x \in \mathbb{R} \setminus \{c\}$.

Sei nun $\varepsilon > 0$. Dann gilt

$$\begin{aligned}
\mathbb{P}(|X_n - c| > \varepsilon) &= \mathbb{P}(X_n - c > \varepsilon \text{ oder } X_n - c < -\varepsilon) \\
&= \mathbb{P}(X_n > c + \varepsilon \text{ oder } X_n < c - \varepsilon) \\
&= \mathbb{P}(X_n > c + \varepsilon) + \mathbb{P}(X_n < c - \varepsilon) \\
&\leq 1 - \mathbb{P}(X_n \leq c + \varepsilon) + \mathbb{P}(X_n \leq c - \varepsilon) \\
&= 1 - F_{X_n}(c + \varepsilon) + F_{X_n}(c - \varepsilon) \\
&\xrightarrow{n\to\infty} 1 - F_c(c + \varepsilon) + F_c(c - \varepsilon) = 1 - 1 + 0 = 0.
\end{aligned}$$

Alternativ haben wir auch

$$\begin{aligned}
\mathbb{P}(|X_n - c| \geq \varepsilon) &= 1 - \mathbb{P}(|X_n - c| < \varepsilon) \\
&= 1 - \mathbb{P}(c - \varepsilon < X_n < c + \varepsilon) \\
&\leq 1 - \mathbb{P}(c - \frac{\varepsilon}{2} < X_n \leq c + \frac{\varepsilon}{2}) \\
&= 1 - \mathbb{P}(X_n \leq c + \frac{\varepsilon}{2}) + \mathbb{P}(X_n \leq c - \frac{\varepsilon}{2}) \\
&= 1 - F_{X_n}(c + \frac{\varepsilon}{2}) + F_{X_n}(c - \frac{\varepsilon}{2}) \\
&\xrightarrow{n\to\infty} 1 - F_c(c + \frac{\varepsilon}{2}) + F_c(c - \frac{\varepsilon}{2}) = 1 - 1 + 0 = 0.
\end{aligned}$$

Also gilt $X_n \xrightarrow{\mathbb{P}} c$.

(iii) Mit der Taylor-Entwicklung von g aus dem Hinweis folgt durch Einsetzen von $x = X_n$ mit einer reellwertigen Zufallsvariable $\tilde{X}_n$ mit $|\tilde{X}_n - c| \leq |X_n - c|$, dass

$$g(X_n) = g(c) + g'(c) \cdot (X_n - c) + (X_n - c) \cdot (g'(\tilde{X}_n) - g'(c)),$$

woraus folgt, dass

$$a_n \cdot (g(X_n) - g(c)) = g'(c) \cdot a_n(X_n - c) + a_n(X_n - c) \cdot (g'(\tilde{X}_n) - g'(c)).$$

Nun gilt

- trivialerweise $g'(c) \xrightarrow{\mathbb{P}} g'(c)$,
- $\tilde{X}_n \xrightarrow{\mathbb{P}} c$, da $X_n \xrightarrow{\mathbb{P}} c$ und $|\tilde{X}_n - c| \leq |X_n - c|$,
  Es gilt $\varepsilon > 0$*:* $\mathbb{P}(|\tilde{X}_n - c| > \varepsilon) \leq \mathbb{P}(|X_n - c| > \varepsilon) \to 0$.

- $g'(\tilde{X}_n) - g'(c) \xrightarrow{\mathbb{P}} 0$,
  Nach dem CMT (Proposition 14.9(iv)) gilt aufgrund der Stetigkeit von g' in c, dass $g'(\tilde{X}_n) \xrightarrow{\mathbb{P}} g'(c)$. Das Verschieben von $g'(c)$ auf die linke Seite der stochastischen Konvergenz ist trivial, da $g'(\tilde{X}_n) - g'(c) \xrightarrow{\mathbb{P}} 0$ und $g'(\tilde{X}_n) \xrightarrow{\mathbb{P}} g'(c)$ eingesetzt in die Definition der stochastischen Konvergenz dieselben Aussagen liefern.
- $a_n(X_n - c) \xrightarrow{D} Z$ nach Voraussetzung.

Damit folgt

$$a_n \cdot (g(X_n) - g(c)) = \underbrace{\underbrace{\underbrace{g'(c)}_{\xrightarrow{\mathbb{P}} g'(c)} \cdot \underbrace{a_n(X_n - c)}_{\xrightarrow{D} Z}}_{\xrightarrow{D} g'(c)\cdot Z \text{ (14.9(iii))}} + \underbrace{\underbrace{(g'(\tilde{X}_n) - g'(c))}_{\xrightarrow{\mathbb{P}} 0} \cdot \underbrace{a_n(X_n - c)}_{\xrightarrow{D} Z}}_{\xrightarrow{\mathbb{P}} 0 \text{ (14.9(iii))}}}_{\xrightarrow{D} g'(\mu)\cdot Z \text{ (14.9(iii))}} \xrightarrow{D} g'(\mu) \cdot Z$$

□

Aufgabe 3

Sei (X, Y) ein Zufallsvektor in $\mathbb{R}^2$ mit gemeinsamer Wahrscheinlichkeitsdichte

$$f_{X,Y}(x, y) = \exp(-y) \cdot \mathbb{1}_{\{0 \leq x \leq y\}}(x, y).$$

(a) Berechnen Sie $\mathbb{E}[X|Y = y]$ bzw. $\mathbb{E}[X|Y]$.

(b) Berechnen Sie $\mathbb{E}[Y|X = x]$ bzw. $\mathbb{E}[Y|X]$.

(2+2) Punkte)

Beweis. (a) Es gilt für $y \in \mathbb{R}$

$$\mathbb{E}[X|Y = y] = \int_{\mathbb{R}} x \cdot f_{X|Y=y}(x) \, \mathrm{d}x,$$

wobei

$$f_{X|Y=y}(x) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & , \text{ falls } f_Y(y) > 0, \\ 0 & , \text{ sonst} \end{cases}$$

für $x \in \mathbb{R}$. Wir müssen berechnen

$$\begin{aligned} f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y) \, \mathrm{d}x &= \int_{\mathbb{R}} e^{-y} \mathbb{1}_{0 \leq x \leq y}(x, y) \, \mathrm{d}x \\ &= e^{-y} \int_0^y \mathrm{d}x \; \mathbb{1}_{0 \leq y}(y) = y e^{-y} \mathbb{1}_{0 \leq y}(y). \end{aligned}$$

Wir bemerken, dass $f_Y(y) > 0$ falls $y > 0$. Also

$$f_{X|Y=y}(x) = \begin{cases} \frac{1}{y} & , \text{ falls } 0 \leq x \leq y \text{ und } 0 < y, \\ 0 & , \text{ sonst} \end{cases}$$

Insgesamt erhalten wir

$$\begin{aligned} \mathbb{E}[X|Y = y] &= \int_{\mathbb{R}} x \cdot f_{X|Y=y}(x) \, \mathrm{d}x \\ &= \int_{0+}^y x \cdot \frac{1}{y} \, \mathrm{d}x = \left[\frac{1}{2} x^2 \right]_0^y \cdot \frac{1}{y} = \frac{1}{2} y \cdot \mathbb{1}_{0 \leq y}(y). \end{aligned}$$

(b) Analog zu (a) berechnen wir

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f_{X,Y}(x,y)\ \mathrm{d}y = \int_{\mathbb{R}} e^{-y}\mathbb{1}_{0\le x\le y}\ \mathrm{d}y = \mathbb{1}_{0\le x}(x)\int_x^\infty e^{-y}\mathrm{d}y \\ &= \mathbb{1}_{0\le x}(x)\big[-e^{-y}\big]_x^\infty = e^{-x}\mathbb{1}_{0\le x}(x). \end{aligned}$$

Bemerke, dass $f_X(x) > 0$, falls $0 \le x$. Also

$$f_{Y|X=x}(y) = \frac{e^{-y}\mathbb{1}_{0\le x\le y}(x,y)}{e^{-x}\mathbb{1}_{0\le x}(x)} = e^{x-y}\mathbb{1}_{0\le x\le y}(x,y)$$

und

$$\begin{aligned} \mathbb{E}[Y|X=x] &= \int_{\mathbb{R}} y f_{Y|X=x}(y)\ \mathrm{d}y = \int_{\mathbb{R}} y e^{x-y}\mathbb{1}_{0\le x\le y}(x,y)\ \mathrm{d}y \\ &= e^x\mathbb{1}_{0\le x}(x)\int_x^\infty y e^{-y}\mathrm{d}y = e^x\mathbb{1}_{0\le x}(x)\Big[-(y+1)e^{-y}\Big]_x^\infty \\ &= e^x\mathbb{1}_{0\le x}(x)(x+1)e^{-x} = (x+1)\mathbb{1}_{0\le x}(x). \end{aligned}$$

□

Aufgabe 4

Seien X, Y reelle Zufallsvariablen und sei $\mathrm{Var}(Y) < \infty$. Die sogenannte bedingte Varianz von Y gegeben X ist definiert durch

$$\mathrm{Var}(Y|X) := \mathbb{E}[(Y - \mathbb{E}[Y|X])^2|X] = \mathbb{E}[Y^2|X] - \mathbb{E}[Y|X]^2.$$

(a) Zeigen Sie:
$$\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y|X)] + \mathrm{Var}(\mathbb{E}[Y|X]).$$

(b) Zeigen Sie, dass der bedingte Erwartungswert $h^*(X) = \mathbb{E}[Y|X]$ den Abstand

$$\mathbb{E}[(Y - h(X))^2]$$

unter allen (messbaren) Funktionen $h : \mathbb{R} \to \mathbb{R}$ minimiert.

(1+2) Punkte)

Beweis. (a) Wir berechnen

$$\begin{aligned} &\mathbb{E}\big[\mathrm{Var}(Y|X)\big] + \mathrm{Var}\big(\mathbb{E}[Y|X]\big) \\ &= \mathbb{E}\big[\mathbb{E}[Y^2|X]\big] - \mathbb{E}\big[\mathbb{E}[Y|X]^2\big] + \mathbb{E}\big[\mathbb{E}[Y|X]^2\big] - \mathbb{E}\big[\mathbb{E}[Y|X]\big]^2 \\ &= \mathbb{E}[Y^2] - \mathbb{E}[Y]^2, \end{aligned}$$

wobei sich nach dem ersten Schritt zwei Terme wegkürzen und im letzten Schritt die Eigenschaft 16.7 (i) verwendet wurde.

(b) Sei $\mathcal{H} := \{h : \mathbb{R} \to \mathbb{R} :\ h \text{ messbar }\}$ der Raum aller messaren Funktionen. Wir wollen zeigen, dass

$$\mathrm{argmin}_{h\in\mathcal{H}}\mathbb{E}\big[(Y - h(x))^2\big] = \mathbb{E}\big[(Y - \mathbb{E}[Y|X])^2\big].$$

Wir rechnen

$$\begin{aligned} \mathbb{E}\big[(Y - h(X))^2\big] &= \mathbb{E}\Big[\big((Y - \mathbb{E}[Y|X]) + (\mathbb{E}[Y|X] - h(X))\big)^2\Big] \\ &= \mathbb{E}\big[(Y - \mathbb{E}[Y|X])^2\big] + 2\mathbb{E}\big[(Y - \mathbb{E}[Y|X])(\mathbb{E}[Y|X] - h(X))\big] \\ &\quad + \mathbb{E}\big[(\mathbb{E}[Y|X] - h(X))^2\big]. \end{aligned}$$

Der erste Summand ist konstant in h. Der dritte Summand wird von $h(X) = \mathbb{E}[Y|X]$ minimiert. Für den zweiten berechnen wir

$$\begin{aligned}
&\mathbb{E}\big[(Y - \mathbb{E}[Y|X])(\mathbb{E}[Y|X] - h(X))\big] \\
&= \mathbb{E}\big[Y\mathbb{E}[Y|X] - \mathbb{E}[Y|X]^2 - Yh(X) + h(X)\mathbb{E}[Y|X]\big].
\end{aligned}$$

Wir stellen fest, dass sich mit der Turmeigenschaft die zwei ersten Terme zu Null addieren, da $\mathbb{E}[Y\mathbb{E}[Y|X]] = \mathbb{E}[\mathbb{E}[Y|X]^2]$. Wir verwenden nocheinmal die Turmeigenschaft um zu sehen, dass

$$\begin{aligned}
&\mathbb{E}\big[(Y - \mathbb{E}[Y|X])(\mathbb{E}[Y|X] - h(X))\big] \\
&= -\mathbb{E}\big[Yh(X)\big] + \mathbb{E}\big[h(X)\mathbb{E}[Y|X]\big] \\
&= -\mathbb{E}\big[\mathbb{E}[Yh(X)|X]\big] + \mathbb{E}\big[\mathbb{E}[h(X)Y|X]\big] = 0.
\end{aligned}$$

Insgesamt haben wir also die Aussage gezeigt.

$\square$