

Statistik für Informatiker, SS 2017

## 1.4.4 Ergänzung: Hoeffding- und McDiarmid-Ungleichung

Matthias Birkner

<http://www.staff.uni-mainz.de/birkner/StatInfo17/>

21.6.2017



JOHANNES GUTENBERG  
UNIVERSITÄT MAINZ

Seien  $X_1, X_2, \dots, X_n$  u.a.,  $X_i$  habe Werte in  $[a_i, b_i]$  (für gewisse Konstanten  $-\infty < a_i < b_i < \infty$ ), setze

$$S_n := X_1 + \dots + X_n.$$

Seien  $X_1, X_2, \dots, X_n$  u.a.,  $X_i$  habe Werte in  $[a_i, b_i]$  (für gewisse Konstanten  $-\infty < a_i < b_i < \infty$ ), setze

$$S_n := X_1 + \dots + X_n.$$

Offenbar ist  $\text{Var}[X_i] \leq (b_i - a_i)^2$  (denn  $|X_i - \mathbb{E}[X_i]| \leq b_i - a_i$ )  
und somit  $\text{Var}[S_n] = \text{Var}[X_1] + \dots + \text{Var}[X_n] \leq \sum_{i=1}^n (b_i - a_i)^2$

Seien  $X_1, X_2, \dots, X_n$  u.a.,  $X_i$  habe Werte in  $[a_i, b_i]$  (für gewisse Konstanten  $-\infty < a_i < b_i < \infty$ ), setze

$$S_n := X_1 + \dots + X_n.$$

Offenbar ist  $\text{Var}[X_i] \leq (b_i - a_i)^2$  (denn  $|X_i - \mathbb{E}[X_i]| \leq b_i - a_i$ ) und somit  $\text{Var}[S_n] = \text{Var}[X_1] + \dots + \text{Var}[X_n] \leq \sum_{i=1}^n (b_i - a_i)^2$ , die Chebyshev-Ungleichung (Satz 1.71) liefert daher für  $t > 0$

$$P(|S_n - \mathbb{E}[S_n]| \geq t) \leq \frac{\sum_{i=1}^n \text{Var}[X_i]}{t^2} \leq \frac{\sum_{i=1}^n (b_i - a_i)^2}{t^2}.$$

Seien  $X_1, X_2, \dots, X_n$  u.a.,  $X_i$  habe Werte in  $[a_i, b_i]$  (für gewisse Konstanten  $-\infty < a_i < b_i < \infty$ ), setze

$$S_n := X_1 + \dots + X_n.$$

Offenbar ist  $\text{Var}[X_i] \leq (b_i - a_i)^2$  (denn  $|X_i - \mathbb{E}[X_i]| \leq b_i - a_i$ ) und somit  $\text{Var}[S_n] = \text{Var}[X_1] + \dots + \text{Var}[X_n] \leq \sum_{i=1}^n (b_i - a_i)^2$ , die Chebyshev-Ungleichung (Satz 1.71) liefert daher für  $t > 0$

$$P(|S_n - \mathbb{E}[S_n]| \geq t) \leq \frac{\sum_{i=1}^n \text{Var}[X_i]}{t^2} \leq \frac{\sum_{i=1}^n (b_i - a_i)^2}{t^2}.$$

Das folgende Resultat stellt oft eine deutliche Verschärfung der Chebyshev-Ungleichung dar (zumindest für beschränkte Summanden).

## Bericht 1.89 (Hoeffding-Ungleichung(en))

Seien  $X_1, X_2, \dots, X_n$  u.a.,  $X_i$  habe Werte in  $[a_i, b_i]$  (für gewisse Konstanten  $-\infty < a_i < b_i < \infty$ ), setze  $S_n := X_1 + \dots + X_n$ . Dann gilt für  $t \geq 0$

$$P(S_n - \mathbb{E}[S_n] \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right),$$

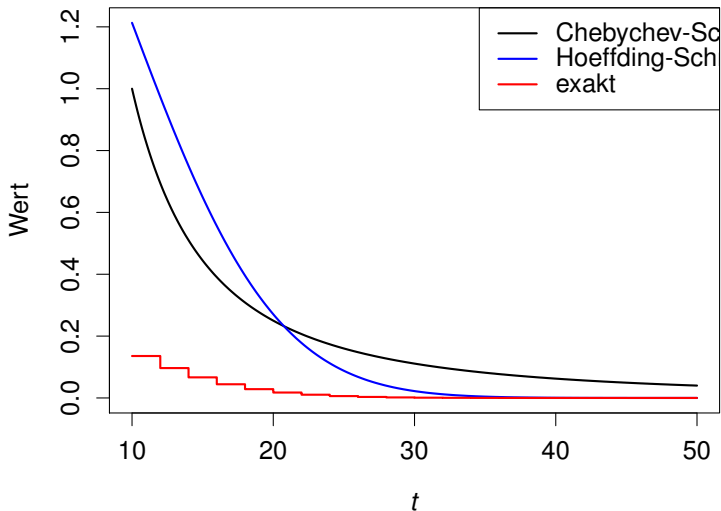
$$P(S_n - \mathbb{E}[S_n] \leq -t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right),$$

insbesondere

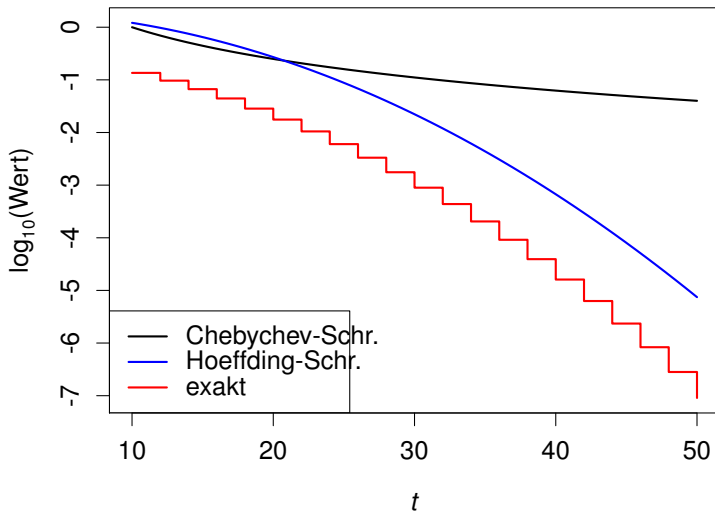
$$P(|S_n - \mathbb{E}[S_n]| \geq t) \leq 2 \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

**Beispiel.**  $n = 100$ ,  $X_i$  u.i.v. mit  $P(X_i = +1) = P(X_i = -1) = \frac{1}{2}$

Schranken für  $P(|S_n - \mathbb{E}[S_n]| \geq t)$



**Beispiel.**  $n = 100$ ,  $X_i$  u.i.v. mit  $P(X_i = +1) = P(X_i = -1) = \frac{1}{2}$   
Schranken für  $P(|S_n - \mathbb{E}[S_n]| \geq t)$  (Werte auf  $\log_{10}$ -Skala)





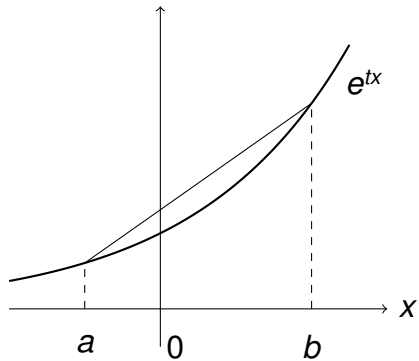
Zum Beweis der Aussagen in Bericht ?? verwendet man folgendes Lemma:

### Lemma 1.90 (Hoeffdings Lemma)

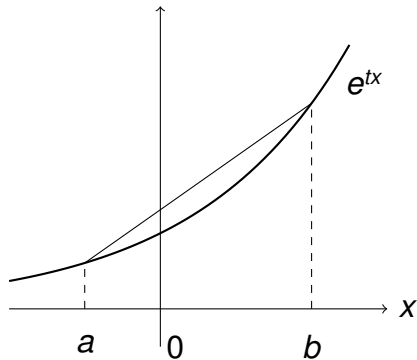
*$a < 0 < b$ ,  $X$  ZV mit Werten in  $[a, b]$  und  $\mathbb{E}[X] = 0$ , dann gilt für  $t \in \mathbb{R}$*

$$\mathbb{E}[e^{tX}] \leq \exp\left(\frac{1}{8}t^2(b-a)^2\right)$$

*Beweis.*  $x \mapsto e^{tx}$  ist konvex,



Beweis.  $x \mapsto e^{tx}$  ist konvex,



daher gilt für jedes  $x \in [a, b]$

$$e^{tx} \leq \frac{b-x}{b-a} e^{ta} + \frac{x-a}{b-a} e^{tb}$$

(beachte  $x = \frac{b-x}{b-a} a + \frac{x-a}{b-a} b$ ).

$$e^{tx} \leq \frac{b-x}{b-a} e^{ta} + \frac{x-a}{b-a} e^{tb}$$

$$e^{tx} \leq \frac{b-x}{b-a} e^{ta} + \frac{x-a}{b-a} e^{tb}$$

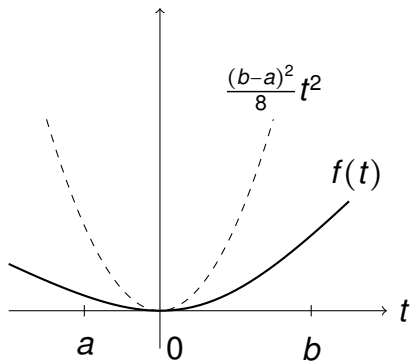
Ersetze  $x \rightsquigarrow X$ , nehme Erwartungswert:

$$\begin{aligned} \mathbb{E}[e^{tX}] &\leq \mathbb{E}\left[\frac{b-X}{b-a} e^{ta} + \frac{X-a}{b-a} e^{tb}\right] = \frac{b-\mathbb{E}[X]}{b-a} e^{ta} + \frac{\mathbb{E}[X]-a}{b-a} e^{tb} \\ &= \frac{b}{b-a} e^{ta} - \frac{a}{b-a} e^{tb}. \end{aligned}$$

$$\mathbb{E}[e^{tX}] \leq \frac{b}{b-a} e^{ta} - \frac{a}{b-a} e^{tb}$$

$$\mathbb{E}[e^{tX}] \leq \frac{b}{b-a} e^{ta} - \frac{a}{b-a} e^{tb}$$

Die Funktion  $f(t) := \log\left(\frac{b}{b-a} e^{ta} - \frac{a}{b-a} e^{tb}\right)$ ,  $t \in \mathbb{R}$



erfüllt  $f(0) = f'(0) = 0$ ,  $f''(t) \leq (b-a)^2/4$

(ggfs. an der Tafel)

$$f(t) := \log\left(\frac{b}{b-a}e^{ta} - \frac{a}{b-a}e^{tb}\right), \quad t \in \mathbb{R} \text{ erfüllt}$$

$$f(0) = f'(0) = 0, \quad f''(t) \leq (b-a)^2/4$$



$$f(t) := \log\left(\frac{b}{b-a}e^{ta} - \frac{a}{b-a}e^{tb}\right), \quad t \in \mathbb{R} \text{ erfüllt}$$

$$f(0) = f'(0) = 0, \quad f''(t) \leq (b-a)^2/4$$

Also ist

$$\begin{aligned} f(t) &= f(0) + \int_0^t f'(s) ds = \int_0^t \left( f'(0) + \int_0^s f''(u) du \right) ds \\ &\leq \frac{(b-a)^2}{4} \int_0^t \int_0^s du ds = \frac{(b-a)^2}{8} t^2 \end{aligned}$$

und somit

$$\mathbb{E}[e^{tX}] \leq \exp(f(t)) \leq \exp\left(\frac{1}{8}t^2(b-a)^2\right)$$

wie behauptet. □

*Beweis der Hoeffding-Ungleichung.* Für  $u > 0$  ist

$$\begin{aligned} P(S_n - \mathbb{E}[S_n] \geq t) &= P\left(\exp(u(S_n - \mathbb{E}[S_n])) \geq e^{ut}\right) \\ &\leq e^{-ut} \mathbb{E}\left[\exp(u(S_n - \mathbb{E}[S_n]))\right] = e^{-ut} \mathbb{E}\left[\prod_{i=1}^n e^{u(X_i - \mathbb{E}[X_i])}\right] \\ &= e^{-ut} \prod_{i=1}^n \mathbb{E}\left[e^{u(X_i - \mathbb{E}[X_i])}\right] \leq e^{-ut} \prod_{i=1}^n \exp\left(\frac{1}{8} u^2 (b_i - a_i)^2\right) \\ &= \exp\left(-ut + u^2 \frac{1}{8} \sum_{i=1}^n (b_i - a_i)^2\right). \end{aligned}$$

(nach Hoeffdings Lemma).

*Beweis der Hoeffding-Ungleichung.* Für  $u > 0$  ist

$$\begin{aligned} P(S_n - \mathbb{E}[S_n] \geq t) &= P\left(\exp(u(S_n - \mathbb{E}[S_n])) \geq e^{ut}\right) \\ &\leq e^{-ut} \mathbb{E}\left[\exp(u(S_n - \mathbb{E}[S_n]))\right] = e^{-ut} \mathbb{E}\left[\prod_{i=1}^n e^{u(X_i - \mathbb{E}[X_i])}\right] \\ &= e^{-ut} \prod_{i=1}^n \mathbb{E}\left[e^{u(X_i - \mathbb{E}[X_i])}\right] \leq e^{-ut} \prod_{i=1}^n \exp\left(\frac{1}{8} u^2 (b_i - a_i)^2\right) \\ &= \exp\left(-ut + u^2 \frac{1}{8} \sum_{i=1}^n (b_i - a_i)^2\right). \end{aligned}$$

(nach Hoeffdings Lemma).

Mit der (optimalen) Wahl  $u = 4t / \sum_{i=1}^n (b_i - a_i)^2$  folgt

$$P(S_n - \mathbb{E}[S_n] \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$



## Bericht 1.91 (McDiarmid-Ungleichung)

$X_1, X_2, \dots, X_n$  unabhängige Zufallsvariablen mit Werten in  $S$ ,  $f: S^n \rightarrow \mathbb{R}$ . Es gebe Konstanten  $c_1, \dots, c_n < \infty$ , so dass für  $i \in \{1, \dots, n\}$ ,  $x_1, \dots, x_n \in S$ ,  $x'_i \in S$  gilt

$$|f(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_n)| \leq c_i.$$

Dann gilt für  $t \geq 0$

$$P\left(f(X_1, \dots, X_n) - \mathbb{E}[f(X_1, \dots, X_n)] \geq t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right)$$

und

$$P\left(|f(X_1, \dots, X_n) - \mathbb{E}[f(X_1, \dots, X_n)]| \geq t\right) \leq 2 \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right).$$

Beachte:

Die Hoeffding-Ungleichung folgt aus der  
McDiarmid-Ungleichung (mit der Wahl  
 $f(x_1, \dots, x_n) = x_1 + \dots + x_n$ ), letztere ist aber in  
allgemeineren Situationen anwendbar.

Wir werden sie hier nicht beweisen.

**Beispiel.** Seien  $X_1, X_2, \dots, X_n$  u.i.v.  $\sim \text{Ber}_p$ ,  $p \in (0, 1)$  (d.h.  $P(X_i = 1) = p = 1 - P(X_i = 0)$ ),

$$W := \sum_{i=2}^n I_{\{X_i \neq X_{i-1}\}}$$

die Anzahl der „Wechsel“ in der Folge  $X_1, X_2, \dots, X_n$   
(z.B. enthält  $(0, 0, 1, 1, 0, 1, 0)$  4 Wechsel).

**Beispiel.** Seien  $X_1, X_2, \dots, X_n$  u.i.v.  $\sim \text{Ber}_p$ ,  $p \in (0, 1)$  (d.h.  $P(X_i = 1) = p = 1 - P(X_i = 0)$ ),

$$W := \sum_{i=2}^n I_{\{X_i \neq X_{i-1}\}}$$

die Anzahl der „Wechsel“ in der Folge  $X_1, X_2, \dots, X_n$   
(z.B. enthält  $(0, 0, 1, 1, 0, 1, 0)$  4 Wechsel).

Beachte: Die Summanden in  $W$  sind nicht unabhängig  
(wir können also nicht die Hoeffding-Ungleichung verwenden).

$$W = \sum_{i=2}^n I_{\{X_i \neq X_{i-1}\}}$$

Es ist  $\mathbb{E}[W] = \sum_{i=2}^n \mathbb{E}[I_{\{X_i \neq X_{i-1}\}}] = (n-1) \cdot 2p(1-p)$   
und wir können schreiben  $W = f(X_1, X_2, \dots, X_n)$  mit  
 $f: \{0, 1\}^n \rightarrow \mathbb{N}_0$ ,

$$f(x_1, \dots, x_n) = \sum_{i=2}^n \mathbf{1}_{\{x_i \neq x_{i-1}\}}.$$



$$W = \sum_{i=2}^n I_{\{X_i \neq X_{i-1}\}}$$

Es ist  $\mathbb{E}[W] = \sum_{i=2}^n \mathbb{E}[I_{\{X_i \neq X_{i-1}\}}] = (n-1) \cdot 2p(1-p)$   
und wir können schreiben  $W = f(X_1, X_2, \dots, X_n)$  mit  
 $f: \{0, 1\}^n \rightarrow \mathbb{N}_0$ ,

$$f(x_1, \dots, x_n) = \sum_{i=2}^n \mathbf{1}_{\{x_i \neq x_{i-1}\}}.$$

$f$  erfüllt die Voraussetzungen der McDiarmid-Ungleichung  
mit  $c_1 = 1, c_2 = c_3 = \dots = c_{n-1} = 2, c_n = 1$ , somit gilt

$$P\left(|W - 2p(1-p)(n-1)| \geq t\right) \leq 2 \exp\left(-\frac{2t^2}{4n-6}\right)$$

$$W = \sum_{i=2}^n I_{\{X_i \neq X_{i-1}\}}$$

Es ist  $\mathbb{E}[W] = \sum_{i=2}^n \mathbb{E}[I_{\{X_i \neq X_{i-1}\}}] = (n-1) \cdot 2p(1-p)$   
und wir können schreiben  $W = f(X_1, X_2, \dots, X_n)$  mit  
 $f: \{0, 1\}^n \rightarrow \mathbb{N}_0$ ,

$$f(x_1, \dots, x_n) = \sum_{i=2}^n \mathbf{1}_{\{x_i \neq x_{i-1}\}}.$$

$f$  erfüllt die Voraussetzungen der McDiarmid-Ungleichung  
mit  $c_1 = 1, c_2 = c_3 = \dots = c_{n-1} = 2, c_n = 1$ , somit gilt

$$P\left(|W - 2p(1-p)(n-1)| \geq t\right) \leq 2 \exp\left(-\frac{2t^2}{4n-6}\right)$$

Wir sehen: Abweichungen um  $t \gg \sqrt{n}$  sind sehr  
unwahrscheinlich.