Statistik für Informatiker, SS 2018

1.5.4 Ergänzung: Hoeffding- und McDiarmid-Ungleichung

Matthias Birkner

http://www.staff.uni-mainz.de/birkner/StatInfo18/

4.6.2018

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

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

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

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

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

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

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

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

Offenbar ist $\mathrm{Var}[X_i] \le (b_i - a_i)^2$ (denn $|X_i - \mathbb{E}[X_i]| \le b_i - a_i$) und somit $\mathrm{Var}[S_n] = \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n] \le \sum_{i=1}^n (b_i - a_i)^2$, die Chebyshev-Ungleichung (Satz 1.77) liefert daher für $t > 0$

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

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

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

Offenbar ist $\mathrm{Var}[X_i] \le (b_i - a_i)^2$ (denn $|X_i - \mathbb{E}[X_i]| \le b_i - a_i$) und somit $\mathrm{Var}[S_n] = \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n] \le \sum_{i=1}^n (b_i - a_i)^2$, die Chebyshev-Ungleichung (Satz 1.77) liefert daher für $t > 0$

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

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

Bericht 1.95 (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 $S_n := X_1 + \cdots + X_n$. Dann gilt für $t \geq 0$

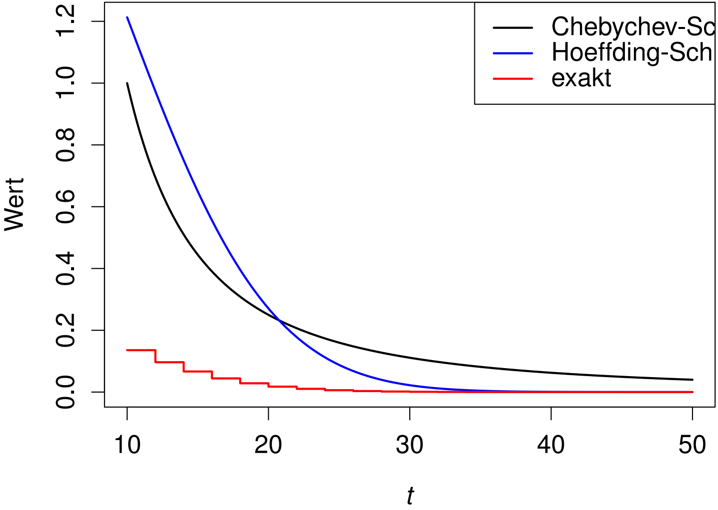
$$P\big(S_n - \mathbb{E}[S_n] \geq t\big) \leq \exp\Big(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\Big),$$
$$P\big(S_n - \mathbb{E}[S_n] \leq -t\big) \leq \exp\Big(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\Big),$$

insbesondere

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

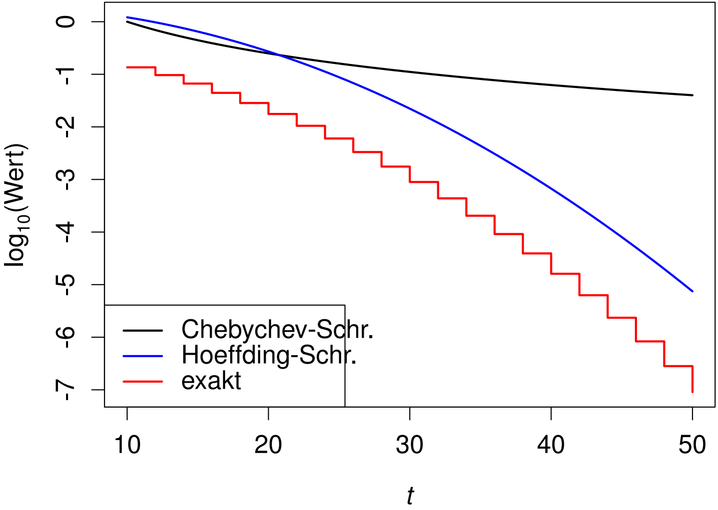
Beispiel. $n = 100$, X_i u.i.v. mit $P(X_i = +1) = P(X_i = -1) = \frac{1}{2}$

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



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

Schranken für $P\big(|S_n - \mathbb{E}[S_n]| \geq t\big)$ (Werte auf $\log_{10}$-Skala)



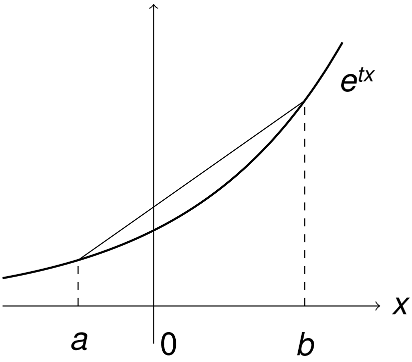
Zum Beweis der Aussagen in Bericht 1.95 verwendet man folgendes Lemma:

Lemma 1.96 (Hoeffdings Lemma)

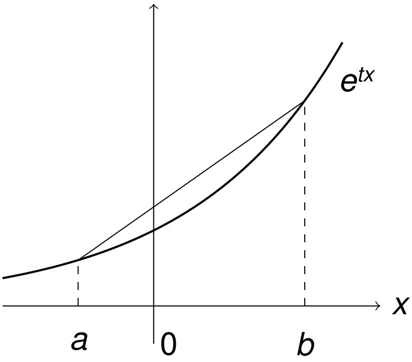
$a < 0 < b$, X ZV mit Werten in $[a, b]$ und $\mathbb{E}[X] = 0$, dann gilt für $t \in \mathbb{R}$

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

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

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

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

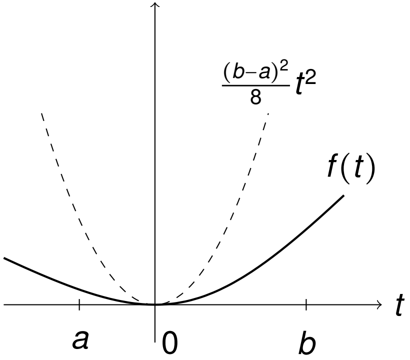
Ersetze $x \rightsquigarrow X$, nehme Erwartungswert:

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

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

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

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



erfüllt $f(0) = f'(0) = 0,\ f''(t) \leq (b-a)^2/4$ (ggfs. an der Tafel)

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

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

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

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

Also ist

$$f(t) = f(0) + \int_0^t f'(s)\,ds = \int_0^t \Big(f'(0) + \int_0^s f''(u)\,du\Big)\,ds$$
$$\le \frac{(b-a)^2}{4}\int_0^t \int_0^s du\,ds = \frac{(b-a)^2}{8}t^2$$

und somit

$$\mathbb{E}\big[e^{tX}\big] \le \exp\big(f(t)\big) \le \exp\Big(\frac{1}{8}t^2(b-a)^2\Big)$$

wie behauptet. $\square$

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

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

(nach Hoeffdings Lemma).

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

$$\begin{aligned}
P\Big(S_n - \mathbb{E}[S_n] \ge t\Big) &= P\Big(\exp\big(u(S_n - \mathbb{E}[S_n])\big) \ge e^{ut}\Big) \\
&\le e^{-ut}\mathbb{E}\Big[\exp\big(u(S_n - \mathbb{E}[S_n])\big)\Big] = e^{-ut}\mathbb{E}\Big[\prod_{i=1}^n e^{u(X_i - \mathbb{E}[X_i])}\Big] \\
&= e^{-ut}\prod_{i=1}^n \mathbb{E}\big[e^{u(X_i - \mathbb{E}[X_i])}\big] \le e^{-ut}\prod_{i=1}^n \exp\big(\tfrac{1}{8}u^2(b_i - a_i)^2\big) \\
&= \exp\Big(-ut + u^2\tfrac{1}{8}\sum_{i=1}^n (b_i - a_i)^2\Big).
\end{aligned}$$

(nach Hoeffdings Lemma).

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

$$P\Big(S_n - \mathbb{E}[S_n] \ge t\Big) \le \exp\Big(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\Big)$$

$\square$

Bericht 1.97 (McDiarmid-Ungleichung)

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

$$\left| f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n) \right| \leq c_i.$$

Dann gilt für $t \geq 0$

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

und

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

Beachte:

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

Wir werden sie hier nicht beweisen.

Beispiel. Seien $X_1, X_2, \ldots, X_n$ u.i.v. $\sim \mathrm{Ber}_p$, $p \in (0,1)$ (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 $X_1, X_2, \ldots, X_n$ (z.B. enthält $(0,0,1,1,0,1,0)$ 4 Wechsel).

Beispiel. Seien $X_1, X_2, \dots, X_n$ u.i.v. $\sim \mathrm{Ber}_p$, $p \in (0, 1)$ (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 $W = f(X_1, X_2, \ldots, X_n)$ mit $f : \{0,1\}^n \to \mathbb{N}_0$,

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

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

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 $W = f(X_1, X_2, \ldots, X_n)$ mit $f : \{0,1\}^n \to \mathbb{N}_0$,

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

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

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

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

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 $W = f(X_1, X_2, \ldots, X_n)$ mit
$f : \{0,1\}^n \to \mathbb{N}_0$,

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

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

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

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