12 Ungleichungen

Wir beginnen mit einer einfachen Ungleichung über die Varianz.

Satz 35 *Es sei X eine zufällige Variable. Dann gilt:*

$$\min_{c \in \mathbb{R}} \mathbf{E}(X-c)^2 = \mathrm{Var}\, X.$$

Beweis: Für alle reellen Zahlen $c \in \mathbb{R}$ gilt:

$$\mathbf{E}(X-c)^2 = \mathbf{E}(X - \mathbf{E}X + \mathbf{E}X - c)^2$$

$$
\begin{aligned}
&= \mathbf{E}(X-\mathbf{E}X)^2 + 2\mathbf{E}((X-\mathbf{E}X)(\mathbf{E}X-c)) + (\mathbf{E}X-c)^2 \\
&= \mathbf{E}(X-\mathbf{E}X)^2 + 2(\mathbf{E}X-c)\underbrace{\mathbf{E}(X-\mathbf{E}X)}_{=0} + (\mathbf{E}X-c)^2 \\
&= \operatorname{Var} X + (\mathbf{E}X-c)^2 \\
&\geq \operatorname{Var} X
\end{aligned}
$$

Setzen wir $c := \mathbf{E}X$ erhalten wir Gleichheit. $\square$

12.1 Jensen-Ungleichung

Satz 36 (Ungleichung von JENSEN**)** *Sei X eine zufällige Variable mit $\mathbf{E}X < \infty$ und g eine differenzierbare und konvexe Funktion. Dann gilt:*

$$\mathbf{E}g(X) \geq g(\mathbf{E}X).$$

Beweis: Sei $T(x)$ die Tangente an die Kurve der Funktion g im Punkt x_0,

$$g(x) \geq T(x) = g(x_0) + \underbrace{g'(x_0)}_{\text{Anstieg der Kurve in } x_0} \cdot (x - x_0).$$

Wir setzen nun $x := X$ und $x_0 := \mathbf{E}X$ und erhalten:

$$g(X) \geq g(\mathbf{E}X) + g'(\mathbf{E}X) \cdot (X - \mathbf{E}X).$$

Daraus folgt:

$$\begin{aligned} \mathbf{E}g(X) &\geq \mathbf{E}(g(\mathbf{E}X) + g'(\mathbf{E}X) \cdot (X - \mathbf{E}X)) \\ &= g(\mathbf{E}X) + g'(\mathbf{E}X) \cdot \underbrace{\mathbf{E}(X - \mathbf{E}X)}_{=0} \\ &= g(\mathbf{E}X) \end{aligned}$$

$\square$

Folg. 6 *Es sei g differenzierbar und konkav. Weiterhin sei X eine zufällige Variable. Dann gilt:*

$$\mathbf{E}g(X) \leq g(\mathbf{E}X).$$

Beweis: Da die Funktion g nach Voraussetzung konkav ist, ist die Funktion $(-g)$ konvex. Dann gilt nach Satz 36:

$$\mathbf{E}((-g)(X)) \geq (-g)(\mathbf{E}X).$$

Daraus folgt:

$$-\mathbf{E}g(X) \geq -g(\mathbf{E}X).$$

$$\mathbf{E}g(X) \leq g(\mathbf{E}X).$$

□

Bsp. 76 *1. Es sei $g(x) = x^2$. Dann gilt nach Satz 36:*

$$\mathbf{E}X^2 \geq (\mathbf{E}X)^2.$$

Daraus aber folgt (die schon bekannte Aussage):

$$\operatorname{Var} X = \mathbf{E}(X - \mathbf{E}X)^2 = \mathbf{E}X^2 - (\mathbf{E}X)^2 \geq 0.$$

2. *Es sei* $g(x) = |x|$. *Dann gilt nach Satz 36:*

$$\mathbf{E}|X| \geq |\mathbf{E}X|.$$

3. *Es sei* $g(x) = \ln x$. *Diese Funktion ist konkav. Also gilt nach Folgerung 6:*

$$\mathbf{E}(\ln X) \leq \ln(\mathbf{E}X).$$

12.2 Markov- und Tschebyschev-Ungleichung

Satz 37 (Ungleichung von MARKOV**)** *Sei* $c > 0$. X *sei eine Zufallsgröße. Dann gilt:*

$$P(|X| > c) \leq \frac{\mathbf{E}|X|}{c}.$$

Beweis: Wir definieren eine Zufallsgröße Y:

$$Y(\omega) := \begin{cases} c & \text{, falls } |X(\omega)| > c \\ 0 & \text{, falls } |X(\omega)| \leq c \end{cases}, \quad \forall \omega \in \Omega.$$

Wir können also schreiben:

$$Y : \quad \begin{pmatrix} 0 & c \\ P(|X| \le c) & P(|X| > c) \end{pmatrix}$$

Offenbar gilt für alle $\omega \in \Omega$:

$$0 \le Y(\omega) \le |X(\omega)|,$$

beziehungsweise:

$$0 \le Y \le |X|.$$

Daraus folgt: $P(|X| - Y \ge 0) = 1$. Nach Satz 11:

$$\begin{aligned} \mathbf{E}(|X| - Y) &\ge 0 \\ \mathbf{E}|X| &\ge EY. \end{aligned}$$

Da die Zufallsgröße Y diskret ist, folgt aus der Definition des Erwartungswertes:

$$\begin{aligned} \mathbf{E}Y &= 0 \cdot P(|X| \leq c) + c \cdot P(|X| > c) \\ &= c \cdot P(|X| > c) \leq \mathbf{E}|X| \end{aligned}$$

Wir stellen um und erhalten:

$$P(|X| > c) \leq \frac{\mathbf{E}|X|}{c}.$$

$\square$

Satz 38 (Ungleichung von Tschebychev**)** *Es sei* $\varepsilon > 0$ *und sei* Y *eine Zufallsgröße. Dann gilt:*

$$P(|Y - \mathbf{E}Y| > \varepsilon) \leq \frac{\operatorname{Var} Y}{\varepsilon^2}.$$

Beweis: Wir verwenden die Ungleichung von MARKOV:

$$P(|X| > c) \leq \frac{\mathbf{E}|X|}{c}.$$

Wir setzen folgendes in diese Ungleichung ein:

$$X := (Y - \mathbf{E}Y)^{2i} \geq 0, \quad c := \varepsilon^{2i} \qquad (i \in \mathbb{N}).$$

Da $\varepsilon > 0$ gilt, ist die Voraussetzung der Ungleichung von MARKOV erfüllt. Wir erhalten also:

$$P\left((Y - \mathbf{E}Y)^{2i} > \varepsilon^{2i}\right) = P(|Y - \mathbf{E}Y| > \varepsilon) \leq \frac{\mathbf{E}(Y - \mathbf{E}Y)^{2i}}{\varepsilon^{2i}}.$$

Wenn wir nun $i := 1$ setzen, so ergibt sich:

$$P(|Y - \mathbf{E}Y| > \varepsilon) \leq \frac{\mathbf{E}(Y - \mathbf{E}Y)^2}{\varepsilon^2} = \frac{\operatorname{Var} Y}{\varepsilon^2}.$$

$\square$

Bem.: Aus der Ungleichung von TSCHEBYCHEV folgt:

$$P(|Y - EY| \le \varepsilon) \ge 1 - \frac{\operatorname{Var} Y}{\varepsilon^2}.$$

Bsp. 77 *Es sei* $X \sim \mathrm{N}(\mu, \sigma^2)$*, also*

$$EX = \mu, \qquad \operatorname{Var} X = \sigma^2.$$

Wir setzen $\varepsilon := k \cdot \sigma$ *(*$k \in \mathbb{N}$*) und erhalten dann mit der Ungleichung von* TSCHEBYCHEV*:*

$$P(|X - \mu| \le k \cdot \sigma) \ge 1 - \frac{\sigma^2}{k^2 \cdot \sigma^2} = 1 - \frac{1}{k^2}.$$

k	*exakt* $\Phi(k\sigma) - \Phi(-k\sigma)$	*Tschebychew-Ungleichung* $1 - \frac{1}{k^2}$
1	0.68629	0
2	0.9545	0.75
3	0.9973	0.89
4	≈ 1	0.93
5	≈ 1	0.96

Bem. 18 *Die Tschebyscheff-Ungleichung gilt für beliebig verteilte Zufallsvariablen, die Erwartungswert und Varianz besitzen, insbesondere liegt die Zufallsvariable X mit Wahrscheinlichkeit ≥ 0.89 im 3σ-Intervall.*

Bsp. 78 *Die Zahl* $med = med(X)$ *heißt Median der Zufallsvariablen* X*, falls*

$$P(X \le med) \ge \frac{1}{2} \quad \textit{und} \quad P(X \ge med) \ge \frac{1}{2}$$

Sei $P(X > 0) = 1$. *Aus der Markov-Ungleichung folgt:*

$$\frac{1}{2} \le P(X \ge med) \le \frac{\mathbf{E}X}{med}, \quad \textit{d.h. } med \le 2 \cdot \mathbf{E}X$$

12.3 Hoeffding-Ungleichung

Satz 39 (Hoeffding-Ungleichung) *Seien* $Y_1, \ldots, Y_n$ *unabhängig und so dass* $\mathrm{E}Y_i = 0$ *und* $a_i \leq Y_i \leq b_i$. *Sei* $\epsilon > 0$. *Dann gilt* $\forall t > 0$:

$$P(\sum_{i=1}^{n} Y_i \geq \epsilon) \leq 2e^{-t\epsilon} \prod_{i=1}^{n} e^{t^2(b_i - a_i)^2/8},$$

Satz 40 (Hoeffding-Ungleichung für Bernoulli ZV) *Seien* $X_1, \ldots, X_n \sim Bi(1, p)$. *Dann gilt* $\forall \epsilon > 0$:

$$P(|\overline{X}_n - p| > \epsilon) \leq 2e^{-2n\epsilon^2},$$

wobei $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

Bsp. 79 *Seien* $X_1, \ldots, X_n \sim Bi(1, p)$*, d.h. Bernoulli:*
$X_i = 1$ *mit Wkt.* p*,* $X_i = 0$ *sonst.*

$n = 100, \epsilon = 0.2.$

Tschebyschev:

$$P(|\overline{X}_n - p|) > \epsilon) \leq \frac{var\overline{X}_n}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2} = 0.0625.$$

Hoeffding:

$$P(|\overline{X}_n - p|) > \epsilon) \leq 2e^{-2\cdot 100\cdot 0.2^2} = 0.00067.$$

 W.Kössler, Humboldt-Universität zu Berlin

Bem.: Es geht sogar noch besser. ZGWS (s. 13):

$$
\begin{aligned}
P(|\overline{X}_n - p|) > \epsilon) &= P\Big(\frac{|\sum_{i=1}^n x_i - np|}{\sqrt{np(1-p)}} > \frac{n\epsilon}{\sqrt{np(1-p)}}\Big) \\
&\approx \Big(1 - \Phi\Big(\frac{n\epsilon}{\sqrt{np(1-p)}}\Big)\Big) + \Phi\Big(-\frac{n\epsilon}{\sqrt{np(1-p)}}\Big) \\
&= 2\Phi\Big(-\frac{n\epsilon}{\sqrt{np(1-p)}}\Big) \\
&\le 2\Phi\Big(\frac{-n\epsilon}{\sqrt{n\frac{1}{4}}}\Big) \\
&= 2\Phi(-2\epsilon\sqrt{n}) \\
&= 2\Phi(-4) \approx 10^{-4}.
\end{aligned}
$$

Bsp. 80 *Sei* $\alpha > 0$ *und*

$$\epsilon_n = \sqrt{\frac{1}{2n}\log(\frac{2}{\alpha})}.$$

Hoeffding:

$$P(|\overline{X}_n - p|) > \epsilon_n) \leq 2e^{-2n\epsilon_n^2} = \alpha.$$

Sei $C = (\overline{X}_n - \epsilon_n, \overline{X}_n + \epsilon_n)$.

$$P(p \notin C) = P(|\overline{X}_n - p| > \epsilon_n) \leq \alpha$$

$$P(p \in C) \geq 1 - \alpha$$

D.h. das zufällige Intervall C *überdeckt den wahren Parameter* p *mit Wkt.* $\geq 1 - \alpha$.

C *heißt* $\underline{1 - \alpha \text{ Konfidenzintervall}}$.

Problem: Schätzung von Binomialwktn.

Vorgabe: ϵ, α.

Gesucht: notwendiger Stichprobenumfang um

$$P(|\hat{p} - p| > \epsilon) < \alpha$$

zu sichern.

Hoeffding: Es genügt:

$$2 \cdot e^{-2n\epsilon^2} < \alpha$$

also

$$n > \frac{-\ln \alpha/2}{2\epsilon^2} = \frac{\ln(2/\alpha)}{2\epsilon^2}.$$

ZGWS:

$$\begin{aligned}
P(|\hat{p}-p|>\epsilon) &= P\Big(\frac{|\hat{p}-p|}{\sqrt{np(1-p)}} > \frac{n\epsilon}{\sqrt{np(1-p)}}\Big) \\
&= 2\Phi\Big(-\frac{n\epsilon}{\sqrt{np(1-p)}}\Big) \qquad < \alpha \\
-\frac{n\epsilon}{\sqrt{np(1-p)}} &< \Phi^{-1}(\frac{\alpha}{2}) \\
\sqrt{n} &> \frac{-\Phi^{-1}(\frac{\alpha}{2})}{\epsilon}\sqrt{p(1-p)} \\
n &> \frac{\big(\Phi^{-1}(1-\frac{\alpha}{2})\big)^2}{4\epsilon^2}
\end{aligned}$$

Vergleich Hoeffding - ZGWS. Sei $\epsilon = 0.01$.

	ZGWS	Hoeffding
δ	$\frac{1}{4\epsilon^2}\Phi^{-1}(1-\frac{\alpha}{2})$	$\frac{1}{2\epsilon^2}\ln\frac{2}{\alpha}$
0.1	6765	15000
0.05	9640	18450
0.01	16641	26490
0.001	27225	38000

Beweis: (Hoeffding-Ungleichung) Sei $t > 0$. Aus der Markov-Ungleichung folgt:

$$\begin{aligned} P(\sum_{i=1}^{n} Y_i \geq \epsilon) &= P(t\sum_{i=1}^{n} Y_i \geq t\epsilon) = P(e^{t\sum_{i=1}^{n} Y_i} \geq e^{t\epsilon}) \\ &\leq e^{-t\epsilon}\mathbf{E}\Big(e^{t\sum_{i=1}^{n} Y_i}\Big) = e^{-t\epsilon}\prod_{i=1}^{n}\mathbf{E}(e^{tY_i}). \end{aligned}$$

Da $a_i \leq Y_i \leq b_i$ läßt sich Y_i als konvexe Kombination von a_i und b_i schreiben,

$$Y_i = \alpha b_i + (1-\alpha)a_i,$$

wobei $\alpha = \frac{Y_i - a_i}{b_i - a_i}$.

NR.: Für konvexe Funktionen $f(x), x \in (a,b)$ gilt:

$$f(x) \leq f(a) + \frac{f(b)-f(a)}{b-a}(x-a) = \alpha f(b) + (1-\alpha)f(a)$$

(Die Kurve g liegt unterhalb der Sekante, $\alpha = \frac{x-a}{b-a}$.).

Da die Exponentialfunktion konvex ist:

$$\begin{aligned} e^{tY_i} &\leq \alpha e^{tb_i} + (1-\alpha)e^{ta_i} \\ &= \frac{Y_i - a_i}{b_i - a_i}e^{tb_i} + \frac{b_i - Y_i}{b_i - a_i}e^{ta_i} \\ \mathbf{E}(e^{tY_i}) &\leq \frac{-a_i}{b_i - a_i}e^{tb_i} + \frac{b_i}{b_i - a_i}e^{ta_i} = e^{g(u)} \end{aligned}$$

wegen $\mathbf{E}Y_i = 0$. Dabei ist

$$\begin{aligned} u &= t(b_i - a_i) \\ g(u) &= -\gamma u + \log(1 - \gamma + \gamma e^u) \\ \gamma &= \frac{-a_i}{b_i - a_i} \\ g(0) &= g'(0) = 0, \quad g''(u) \leq \frac{1}{4} \forall u > 0. \end{aligned}$$

Satz von Taylor: es ex. ein $\xi \in (0, u)$:

$$\begin{aligned} g(u) &= g(0) + ug'(0) + \frac{u^2}{2} g''(\xi) \\ &= \frac{u^2}{2} g''(\xi) \leq \frac{u^2}{8} = \frac{t^2(b_i - a_i)^2}{8} \end{aligned}$$

Daraus folgt:

$$\mathbf{E}(e^{tY_i}) \leq e^{g(u)} \leq e^{t^2(b_i-a_i)^2/8}.$$

Damit:

$$P(\sum_{i=1}^{n} Y_i \geq \epsilon) = e^{-t\epsilon} \prod_{i=1}^{n} \mathbf{E}(e^{tY_i}) \leq e^{-t\epsilon} \prod_{i=1}^{n} e^{t^2(b_i-a_i)^2/8}$$

Beweis: (Hoeffding-Ungleichung für Bernoulli ZV) Sei $Y_i = \frac{1}{n}(X_i - p)$. Dann gilt $\mathbf{E}Y_i = 0$ und $a \le Y_i \le b$, wobei $a = -p/n$ und $b = (1-p)/n$.

Also $(b-a)^2 = 1/n^2$. Aus der Hoeffding-Ungleichung folgt:

$$P(X_n - p > \epsilon) = P(\sum_{i=1}^{n} Y_i > \epsilon) \le e^{-t\epsilon} e^{t^2/(8n)},$$

für jedes $t > 0$. Setze $t = 4n\epsilon$:

$$P(\overline{X}_n - p > \epsilon) \le e^{-2n\epsilon^2}.$$

Analog:

$$P(\overline{X}_n - p < -\epsilon) \le e^{-2n\epsilon^2}.$$

Beides zusammen:

$$P(|\overline{X}_n - p| > \epsilon) \leq 2e^{-2n\epsilon^2}.$$

Satz 41 (Chernov-Ungleichung) *Seien* $X_1, \ldots, X_n \sim Bi(1,p)$. *Dann gilt* $\forall \delta \in (0,1)$:

$$P(\frac{\overline{X}_n - p}{p} > \delta) \leq e^{-pn\frac{\delta^2}{3}}$$

$$P(-\frac{\overline{X}_n - p}{p} > \delta) \leq e^{-pn\frac{\delta^2}{2}}$$

wobei $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

Beweis: s. z.B. in Wikipedia $\square$

Satz 42 (Mill-Ungleichung) *Sei* $Z \sim \mathcal{N}(0,1)$. *Dann gilt:*

$$P(|Z| > t) \leq \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t} = \frac{2\phi(t)}{t}$$