

Peter Zakenbrink - Martingale: Die Ungleichung von Azuma-Hoeffding und McDiarmid

Bachelorseminar zur W-Theorie
Prof. Dr. Gerold Alsmeyer
WWU Münster
Vortrag am 12.12.2013

In Rahmen des heutigen Vortrages möchte ich mich mit der interessanten Frage beschäftigen, inwieweit sich sogenannte Tail-Wahrscheinlichkeiten $(\mathbb{P}(M_n \geq t), \mathbb{P}(M_n \leq -t))$ für Martingale mit beschränkten Zuwächsen abschätzen lassen. Erkenntnis darüber wird die Azuma-Hoeffding-Ungleichung liefern, an die wir uns mithilfe dreier aufschlussreicher Lemmata heranhangeln wollen.

Im Anschluss daran möchten wir die Ungleichung benutzen, um zu einer weiteren Wahrscheinlichkeitsabschätzung zu gelangen, die sich mit der Abweichung einer Funktion angewandt auf stoch. unabh. Zufallsgrößen von ihrem Erwartungswert beschäftigt. Dies wird der Satz von McDiarmid klären.

Damit Sie alle vorab schon eine etwas klarere Vorstellung haben, was wir uns heute erarbeiten wollen, habe ich mich an dieser Stelle entschlossen, das Pferd von hinten aufzuzäumen und Ihnen die beiden großen Ungleichungen gleich einmal offenzulegen.

Für den heutigen Vortrag soll also $(M_n)_{n \geq 0}$ ein Martingal sein, $M_0 = 0$ und die Zuwächse D_n , also gerade $M_n - M_{n-1}$ sollen zunächst durch Konstanten a_n, b_n beschränkt sein, sodass $-a_n \leq D_n \leq b_n$ IP-f.s.

Die Azuma-Hoeffding-Ungleichung besagt nun, dass man die "Tail-Wahrscheinlichkeit" unter diesen Umständen wie folgt abschätzen kann:

Satz 4.55 (Azuma-Hoeffding-Ungleichung)

$$P(M_n^\pm \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{k=1}^n (a_k + b_k)^2}\right)$$

$$P(|M_n| \geq t) \leq 2 \exp\left(-\frac{2t^2}{\sum_{k=1}^n (a_k + b_k)^2}\right) \quad \text{für jedes } t > 0.$$

Wir gelangen also zu einer exponentieller Schranke, bei der zudem die Beschränkungen der Zuwächse eine Rolle spielen.

Ebenfalls eine exponentielle Abschätzung liefert der Satz von McDiarmid, der sich wie bereits erwähnt mit der Funktionsabweichung von seiner Erwartungswert beschäftigt.

Satz 4.60 (McDiarmid)

- X_1, \dots, X_n stoch. unabh. ZG
- $f: \mathbb{R}^n \rightarrow \mathbb{R}$ messbare Fkt. derart, dass $f(X_1, \dots, X_n)$ int'bar und $|f(X_1, \dots, X_n) - f(X_1, \dots, X_{i-1}, X_i', X_{i+1}, \dots, X_n)| \leq c_i$ IP-f.s. mit geeigneten Konstanten c_1, \dots, c_n und alle $i=1, \dots, n$, wobei X_i' eine von X_i unabh. Kopie bezeichnet

In diesem Fall ist die Oszillation von f jeweils im i -ten Element beschränkt. Dann gilt:

$$P(|f(X_1, \dots, X_n) - \mu|^\pm \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{k=1}^n c_k^2}\right)$$

$$P(|f(X_1, \dots, X_n) - \mu| \geq t) \leq 2 \cdot \exp\left(-\frac{2t^2}{\sum_{k=1}^n c_k^2}\right)$$

für alle $t > 0$, wobei $\mu = \mathbb{E}f(X_1, \dots, X_n)$

Bevor wir richtig durchstarten, sei an dieser Stelle der Anschaulichkeit halber noch eine kleine Anwendung gegeben.

Stellen Sie sich hierfür den n -fachen Münzwurf vor.

Wir haben also unabh. und id. vert. Bernoulli-Variablen X_1, \dots, X_n ,

$$\text{wobei } P(X_1=1) = p$$

$$P(X_1=0) = 1-p$$

Betrachten wir nun $f(x_1, \dots, x_n) := \sum_{k=1}^n x_k$ für beliebig fixiertes n , dann können die Konstanten $c_1 = \dots = c_n = 1$ gesetzt werden, da X_i jeweils nur 0 oder 1 annehmen kann.

Will man nun $P(|\bar{X}_n - p| \geq t)$ abschätzen, also die Abweichung des Mittelwerts der X_i vom erwarteten Mittel bei unendlich fachem Münzwurf so gilt:

$$P(|\bar{X}_n - p| \geq t) = P(|S_n - np| \geq n \cdot t) \leq e^{-2nt^2} \left(e^{-\frac{2n^2 t^2}{\sum_{i=1}^n 1}} \right),$$

wobei im letzten Schritt McDiarmid verwendet wurde und sich ein n wegkürzen lässt.

Nun aber genug der Vorausschau, kommen wir zur weiteren Struktur des Vortrags:

- Lemma 4.51 (Balayage)

↳ Lemma 4.53

- Lemma 4.54 (Hoeffding)

↳ Azuma-Hoeffding-Ungleichung (Satz 4.55)

- Azuma-Hoeffding-Maximal-Ungleichung (Satz 4.58)

↳ McDiarmid (Satz 4.60)

- Beispiel

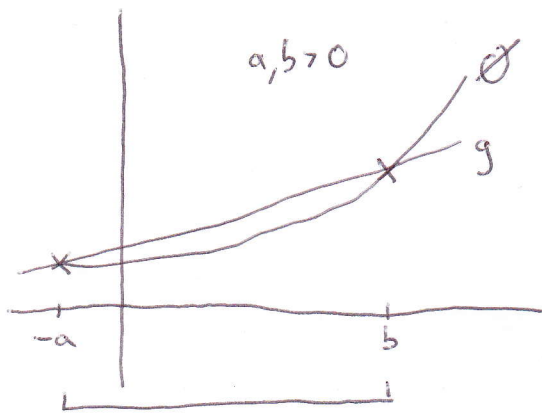
Zu Ihrem Handout:

Sie finden die einzelnen Lemmata auf Ihrem Handout, ich bitte jedoch, diese auch erst nach Abhandlung anzuschauen und für die spätere Benutzung einzubeziehen.

Außerdem haben Sie auf der ersten Seite die beiden Hauptsätze und finden auf Seite 2 ein paar bekannte Formeln, die wir heute benötigen.

Abschließend haben Sie auf der letzten Seite noch ein paar Erklärungen zum Beispiel gegeben.

Nun aber Butter bei die Fische, wir starten durch mit dem Balayage-Lemma: Hierbei geht es anschaulich um eine Extremalitseigenschaft von Zweipunktverteilungen.



Stellen Sie sich vor, Y sei eine $\{-a, b\}$ -wertige, zentrierte Zf, also $P(Y=-a) = \frac{b}{a+b}$, $P(Y=b) = \frac{a}{a+b}$.

Jetzt gilt für jede zentrierte Zf X mit Werten in $[-a, b]$ und jede konvexe Fkt. $\varphi: [-a, b] \rightarrow \mathbb{R}$ folgende Ungleichung:

$$E \varphi(X) \leq E \varphi(Y) = \frac{b}{a+b} \varphi(-a) + \frac{a}{a+b} \varphi(b)$$

Wenn Sie sich meine Zeichnung anschauen wird dies auch sofort klar.

Nimmt X weitere Werte zwischen $[-a, b]$ an, so kann dies der Erwartungswert von $\varphi(X)$ nur runterziehen.

Betrachten Sie für den Beweis die Gerade

$$g(x) = \frac{b}{a+b} \varphi(-a) + \frac{a}{a+b} \varphi(b) + \frac{x}{a+b} (\varphi(b) - \varphi(-a)),$$

welche augenscheinlich durch $(-a, \varphi(-a))$, $(b, \varphi(b))$ verläuft.

Die Gerade verläuft immer oberhalb von ϕ , also gilt $\phi(X) \leq g(X)$ und mit Bildung des Erwartungswertes folgt die Behauptung.

(Konstanten rausziehen, Faktor rausziehen, $\mathbb{E}X=0$). \square

Wir erhalten also das Maximum bei maximaler Varianz.

Übrigens ist "balayage" dem französischen balayer angelehnt, was fegen bedeutet. Man fegt also die Masse in die Eckpunkte $-a$ und b .

Kommen wir direkt zum zweiten Schritt auf dem Weg zur Azuma-Koeffizienten-Ungleichung.

Lemma 4.53

Sei $(M_k)_{0 \leq k \leq n}$ ein Martingal mit $M_0 = 0$ und beschränktem Zuwächsen wie vorher.

Sei ferner $(S_k)_{0 \leq k \leq n}$ die Partialsummenfolge unabh. ZG X_1, \dots, X_n mit $S_0 = 0$.

Für die X_k gilt:

$$X_k \stackrel{d}{=} \frac{b_k}{a_k + b_k} \delta_{-a_k} + \frac{a_k}{a_k + b_k} \delta_{b_k}, \text{ also die Zweipunktverteilungen}$$

wobei δ das Dirac-Maß ist.

Es gilt (der wichtige Teil des Satzes)

$$\mathbb{E}\phi(M_k) \leq \mathbb{E}\phi(S_k)$$

für alle $k=1, \dots, n$ und jede konvexe Fkt. $\phi: \mathbb{R} \rightarrow \mathbb{R}$.

Den Beweis wollen wir per Induktion angehen.

Starten wir mit $k=1$, so folgt

$\mathbb{E}\phi(M_1) \leq \mathbb{E}\phi(S_1)$ direkt aus dem Balayage-Lemma, da der zentrierte

Zuwachs D_1 im Intervall $[-a_1, b_1]$ liegt.

Weitere Erkenntnisse:

- cBdA sind (M_k) und (S_k) unabhängig

↳ (M_k) bildet auch ein Martingal bzgl. $\mathcal{F}_k := \sigma((M_j, S_j), 0 \leq j \leq k)$,
 $0 \leq k \leq n$

- Man beachte außerdem, dass $P(D_{k+1} \in \cdot | \mathcal{F}_k) \in \mathcal{P}_0([-a_{k+1}, b_{k+1}])$ f.s.
↑ zentr. Vert.

Dann folgt: (Ind.schritt)

$$\mathbb{E} \varphi(M_{k+1}) = \mathbb{E} \varphi(M_k + D_{k+1})$$

$$\stackrel{\text{Dalyage}}{\leq} \mathbb{E} \varphi(M_k + X_{k+1})$$

$$= \frac{b_{k+1}}{a_{k+1} + b_{k+1}} \mathbb{E} \varphi(M_k - a_{k+1}) + \frac{a_{k+1}}{a_{k+1} + b_{k+1}} \mathbb{E} \varphi(M_k + b_{k+1})$$

$$\stackrel{\text{ind. vor.}}{\leq} \frac{b_{k+1}}{a_{k+1} + b_{k+1}} \mathbb{E} \varphi(S_k - a_{k+1}) + \frac{a_{k+1}}{a_{k+1} + b_{k+1}} \mathbb{E} \varphi(S_k + b_{k+1})$$

$$= \mathbb{E} \varphi(S_{k+1}) \quad \square$$

Direkt weiter im Text, auf zum letzten Baustein für die Zusammensetzung der Azuma-Hoeffding-Ungleichung.

Wir nähern uns auch namenstechnisch an und kommen zum Lemma von Hoeffding, welches eine Schranke für $\mathbb{E} e^{\theta Y}$, $\theta \in \mathbb{R}$ liefert für eine zentrierte ZG Y mit Zweipunktverteilung.

$$\text{Also } Y \stackrel{d}{=} \frac{b}{a+b} \delta_{-a} + \frac{a}{a+b} \delta_b$$

Dann gilt:

$$\mathbb{E} e^{\theta Y} \leq \exp\left(\frac{\theta^2 (a+b)^2}{8}\right) \quad \text{für alle } \theta \in \mathbb{R}$$

und offensichtlich im Fall $a=b$

$$\mathbb{E} e^{\theta Y} \leq \exp\left(\frac{\theta^2 a^2}{2}\right).$$

Beweis

Wie ich finde ein Klassiker der Gattung absolut verständlich,
aber darauf kommen...?!

$$\text{Wir setzen } p := \frac{a}{a+b} \text{ und } x := \vartheta \cdot (a+b)$$

Dann gilt

$$\mathbb{E} e^{\vartheta Y} = (1-p) \cdot e^{-px} + p \cdot e^{(1-p) \cdot x}$$

denn

$$(1-p) = \left(1 - \frac{a}{a+b}\right) = \frac{b}{a+b}$$

$$-px = \frac{-a}{a+b} \cdot \vartheta \cdot (a+b) = -a \cdot \vartheta$$

$$p = \frac{a}{a+b}$$

$$(1-p)x = \frac{b}{a+b} \cdot \vartheta \cdot (a+b) = b \cdot \vartheta$$

Weiter gilt:

$$\begin{aligned} \mathbb{E} e^{\vartheta Y} &= (1-p) e^{-px} + p \cdot e^{(1-p)x} \\ &= e^{-px} (1-p + p e^x) = e^{g(x)} \end{aligned}$$

$$\text{mit } g(x) := -px + \log(1-p + p e^x) \text{ für } x \in \mathbb{R}.$$

Warum das ganze?

Wir wollen jetzt g per Taylorentwicklung abschätzen, also ran ans Werk.
Entwicklungspunkt soll die 0 sein.

Offensichtlich gilt $g(0) = 0$.

$$g'(x) = -p + \frac{p \cdot e^x}{1-p + p e^x}$$

$$(\ln(x))' = \frac{1}{x}, \text{ multipliziere mit immer Abl.}$$

$$\Rightarrow g'(0) = 0$$

$$g''(x) = \frac{p \cdot e^x (1-p)}{(1-p + p e^x)^2}$$

$$\text{(Quot. regel: } \left(\frac{z}{n}\right)' = \frac{z' \cdot n - n' \cdot z}{n^2}\text{)}$$

$$\begin{aligned} \text{und: } & p \cdot e^x (1-p + p e^x) - p \cdot e^x (p e^x) \\ &= p e^x (1-p) \end{aligned}$$

$$\Rightarrow \left[\underbrace{\frac{1-p}{(1-p + p e^x)}}_a \cdot \left(1 - \underbrace{\frac{1-p}{1-p + p e^x}}_a\right) \right] \leq \frac{1}{4}$$

Warum kleiner als $\frac{1}{4}$?

Offensichtlich sind beide Terme ≤ 1 und aufgrund der Symmetrie ist der Faktor maximal für jeweils $\frac{1}{2}$.

Nun gilt mit Taylorentwicklung und Entwicklungspunkt 0:

$$g(x) = \frac{g''(\xi) \cdot x^2}{2} \leq \frac{x^2}{8} \quad \text{und mit } x := \theta \cdot (a+b),$$

$\xi \in [0, x]$ folgt die Beh. \square

Wunderbar, wir haben alle Zutaten zusammen, um den Beweis der Azuma-Hoeffding-Ungleichung anzutreten!

Wir zeigen nur die erste Ungleichung für M_n^+ , das Pendant für M_n^-

folgt genauso per Übergang zu $(-M_k)_{0 \leq k \leq n}$

(außerdem Betrag: $|M_n| \geq t \Leftrightarrow M_n \leq -t \vee M_n \geq t$)

Legen wir los:

$$\mathbb{P}(M_n \geq t) \stackrel{e, \theta > 0}{=} \mathbb{P}(e^{\theta M_n} \geq e^{\theta t}) \stackrel{\text{Markov}}{\leq} e^{-\theta t} \cdot \mathbb{E} e^{\theta M_n}$$

$$\stackrel{\text{Lemma 4.53}}{\leq} e^{-\theta t} \cdot \mathbb{E} e^{\theta S_n} \stackrel{S_n = X_1 + \dots + X_n}{=} e^{-\theta t} \cdot \prod_{k=1}^n \mathbb{E} e^{\theta X_k}$$

$$\stackrel{X_k \text{ Zweipunkt}}{\leq} e^{-\theta t} \cdot \prod_{k=1}^n e^{\frac{\theta^2 (a_k + b_k)^2}{8}}$$

Hoeffding

$$\stackrel{\text{Produkt}}{=} \exp\left(\frac{\theta^2}{8} \cdot \sum_{k=1}^n (a_k + b_k)^2 - \theta t\right)$$

in Exponent.

$$\stackrel{\text{quadr. Erg.}}{=} \exp\left(\frac{1}{8} \sum_{k=1}^n (a_k + b_k)^2 - \left(\theta - \frac{4t}{\sum_{k=1}^n (a_k + b_k)^2}\right)^2 \frac{zt^2}{\sum_{k=1}^n (a_k + b_k)^2}\right)$$

Der letzte Ausdruck wird gerade minimal

für $\theta = \frac{4t}{\sum_{k=1}^n (a_k + b_k)^2}$ und es folgt die Beh.

\square

Die soeben gezeigte Ungleichung wollen wir nun mithilfe der Doob'schen Ungleichung noch auf Maxima erweitern, was direkt zum folgenden Satz führt.

Satz 4.58 (Azuma-Hoeffding-Maximal-Ungleichung)

Sei die Situation wie im Satz zuvor. Dann gilt weiter:

$$P(\max_{0 \leq k \leq n} M_n^\pm \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{k=1}^n (a_k + b_k)^2}\right) \quad \text{für jedes } t > 0$$

und entsprechendes wieder mit Vorfaktor 2 für den Betrag.

Beweis

Wir zeigen erneut die erste Ungleichung für M_n^+ .

Mit Satz 4.3 ist $(e^{\Theta M_k^+})_{0 \leq k \leq n}$ ein Submartingal, sodass sich die Doob'sche Ungl. anwenden lässt.

$$P(\max_{0 \leq k \leq n} M_k^+ \geq t) = P(\max_{0 \leq k \leq n} e^{\Theta M_k^+} \geq e^{\Theta t}) \stackrel{\text{Doob}}{\leq} e^{-\Theta t} \cdot \mathbb{E} e^{\Theta M_n^+}$$

Nun sind wir gerade in der gleichen Situation wie im Beweis zur Azuma-Hoeffding-Ungl. und es folgt die Beh. \square

Nun wollen wir als Abschluss der Beweise noch den Satz von McDiarmid herleiten.

Dazu noch eine Anmerkung:

Die Ungleichung von Azuma-Hoeffding bleibt richtig, wenn wir die Bedingung der Zuwächse $-a_k \leq D_k \leq b_k$ IP-f.s. durch

$$W_{k-1} \leq D_k \leq W_{k-1} + c_k \quad (W_{k-1} - c_k \leq D_k \leq W_{k-1}) \text{ IP-f.s.}$$

ersetzen, sofern die ZG W_{k-1} \mathcal{F}_{k-1} -messbar sind und man geeignete Konstanten c_k findet.

(Die D_k sind also nicht mehr notwendig beschränkt, haben aber unter \mathcal{F}_{k-1} einen f.s. beschränkten Wertebereich)

Das werden wir uns jetzt zunutze machen.

Beweis

Zunächst wollen wir klären, dass wir f oBdA als beschränkt voraussetzen können.

(sonst setze

$$f_m(x_1, \dots, x_n) := \begin{cases} -m & \text{falls kleiner gleich } -m \\ f(x_1, \dots, x_n) & \text{falls } \in (-m, m) \\ m & \text{falls größer gleich } m \end{cases}$$

und betrachte

$$\lim_{m \rightarrow \infty} \mathbb{P}(|f_m(x_1, \dots, x_n) - \mu_m|^{\pm} \geq t) = \mathbb{P}(|f(x_1, \dots, x_n) - \mu|^{\pm} \geq t),$$

was korrekt bleibt.

Klären wir noch ein paar Fakten:

Wir wählen

- $(\mathcal{F}_k)_{k \in \mathbb{N}}$ als kan. Filtration von $X := (X_1, \dots, X_n)$
- \mathcal{F}_0 triviale σ -Algebra
- Q_k Verteilung der X_k
- (X'_1, \dots, X'_n) unabh. Kopie von X
- $X_{j:k} := (X_j, \dots, X_k)$ für $j \leq k$
- Doob-Martingal $M_k := \mathbb{E}(f(X) | \mathcal{F}_k) - \mu$

$$\Rightarrow M_0 = 0, M_n = f(X) - \mu$$

(Martingal, da

$$\begin{aligned} & \mathbb{E}(M_k | \mathcal{F}_s) \\ &= \mathbb{E}(\mathbb{E}(f(X) | \mathcal{F}_k) - \mu | \mathcal{F}_s) \\ &= \mathbb{E}(f(X) | \mathcal{F}_s) - \mu = M_s \quad (\mathbb{P}\text{-f.s.}) \end{aligned}$$

Wir zeigen, dass M_k beschränkte Zuwächse hat:

$$\begin{aligned} D_k &= M_k - M_{k-1} \\ &= \mathbb{E}(f(X) | \mathcal{F}_k) - \mathbb{E}(f(X) | \mathcal{F}_{k-1}) \\ &= \int \dots \int f(x_{1:k}, x_{k+1:n}) - f(x_{1:k-1}, x_{k:n}) Q_k(dx_k) \dots Q_n(dx_n) \\ &\quad \text{(denn wir müssen entsprechend nach allen Variablen integrieren)} \\ &= \int \dots \int \left(\sup_{u \in \mathbb{R}} f(x_{1:k-1}, u, x_{k+1:n}) - f(x_{1:k-1}, x_{k:n}) \right) Q_k(dx_k) \dots Q_n(dx_n) \\ &\quad - \int \dots \int \left(\sup f(x_{1:k-1}, u, x_{k+1:n}) - f(x_{1:k}, x_{k+1:n}) \right) Q_k(dx_k) \dots Q_n(dx_n) \end{aligned}$$

Wir sehen, dass der untere Teil $\leq C_k$ ist und wir setzen den oberen Teil als W_{k-1} .

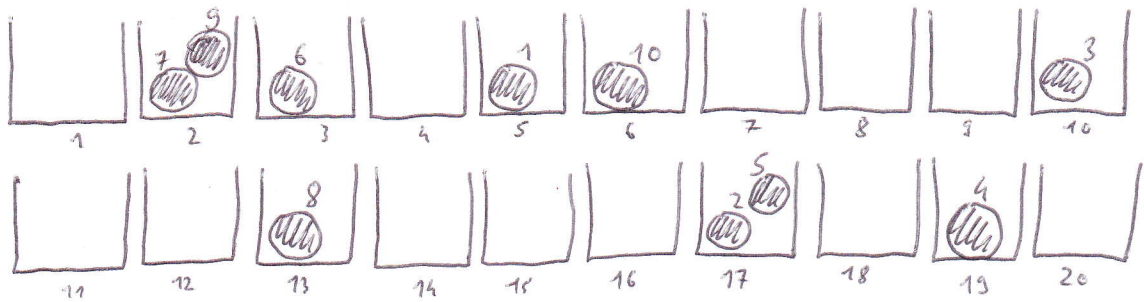
Dann gilt gerade:

$$W_{k-1} - C_k \leq D_k \leq W_{k-1} \quad \text{und wir erhalten unser Resultat. } \square$$

Beispiel: Teilchen - Fächer - Modell

n	Anzahl Kugeln	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8	○ 9	○ 10
m	Anzahl Fächer	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
		[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]

Beispielhafte Verteilung



N_k Anzahl der Fächer mit genau k -Kugeln $k=0, \dots, n$
 $N_0 = 12, N_1 = 6, N_2 = 2, \dots, N_{10} = 0$

X_i Nummer des Fachs mit i -ter Kugel $i=1, \dots, n$
 $X_1 = 5, X_2 = 17, \dots, X_{10} = 6$

U_j Anzahl der Teilchen in Fach j $j=1, \dots, m$
 $U_1 = 0, U_2 = 2, \dots, U_{20} = 0$

Y_i Anzahl verschiedener Werte in $\{X_1, \dots, X_i\}$ $i=2, \dots, n$
 $Y_2 = 2, Y_3 = 3, Y_4 = 4, Y_5 = 4, \dots, Y_{10} = 8$

Bekannt:

U_j besitzt $\text{Bin}(n, \frac{1}{m})$ -Verteilung

$$\text{also } \mathbb{P}(U_j = k) = \binom{n}{k} \cdot \left(\frac{1}{m}\right)^k \cdot \left(1 - \frac{1}{m}\right)^{n-k}$$

Wir können leicht den Erwartungswert von N_k berechnen:

(Um die Anzahl der Fächer mit genau k -Teilchen zu berechnen, schauen wir uns die Anzahl der Teilchen in den einzelnen Fächern an, deren Verteilung bekannt ist)

$$\mathbb{E}N_k = \mathbb{E}\left(\sum_{j=1}^m \mathbb{1}_{\{U_j = k\}}\right) \stackrel{\text{id.wert.}}{=} m \cdot \mathbb{P}(U_1 = k) = m \cdot \binom{n}{k} \cdot \left(\frac{1}{m}\right)^k \cdot \left(1 - \frac{1}{m}\right)^{n-k}$$

Betrachte unser Beispiel:

$$\begin{aligned} \mathbb{E}N_0 &= 20 \cdot \mathbb{P}(U_1 = 0) = 20 \cdot \binom{10}{0} \cdot \left(\frac{1}{20}\right)^0 \cdot \left(\frac{19}{20}\right)^{10} \\ &= 20 \cdot \left(\frac{19}{20}\right)^{10} \approx 11,9747 \dots \end{aligned}$$

Wir erwarten also, dass bei 10 Kugeln und 20 Fächern etwa 12 Fächer unbesetzt bleiben.

Ohne Beweis sei an dieser Stelle angeführt, dass für große $m, n \in \mathbb{N}$ mit $\frac{n}{m} = \theta$ mithilfe des Poissonschen Grenzwertsatzes (WT-Buch S. 131) gilt:

$$\mathbb{E}N_k \approx m \cdot e^{-\theta} \frac{\theta^k}{k!} \quad (\mathbb{E}N_k \text{ approx. linear in } m)$$

(Beweis über simple Umformung bzw. Umschreibung der Binvert.)

Unser Ziel ist es, die Wahrscheinlichkeit dafür abzuschätzen, dass die N_k , also die Anzahl der Fächer mit genau k Teilchen, über einen gewissen Faktor des Erwartungswerts steigt. Also

$$\mathbb{P}(N_k \geq (1+t) \cdot \mathbb{E}N_k) \quad \text{für ein } t > 0$$

Was brauchen wir aus unserem Baukasten um die Azuma-Hoeffding-Ungl. anzuwenden?

⇒ Wir brauchen ein Martingal mit beschränkten Zuwächsen

Hierfür eignen sich besonders gut die Doob-Martingale.

Wir setzen also $M_i := \mathbb{E}(N_k | \mathcal{F}_i)$

Mit \mathcal{F}_0 als trivialer σ -Algebra folgt

$$M_0 = \mathbb{E}(N_k)$$

Mit N_k \mathcal{F}_n -messbar folgt ($N_k = f(X_1, \dots, X_n)$ für geeign. Fkt. f)

$$M_n = N_k$$

Wir untersuchen den einfacheren Fall $k=0$, betrachten also die Anzahl der Fächer mit 0 Kugeln.

Damit gilt:

$$\mathbb{P}(N_0 \geq (1+t) \cdot \mathbb{E}N_0) = \mathbb{P}(N_0 - \mathbb{E}N_0 \geq t \cdot \mathbb{E}N_0) = \mathbb{P}(M_n - M_0 \geq t \cdot \mathbb{E}N_0)$$

Wir zeigen die Beschränktheit, suchen also \mathcal{F}_{i-1} -messbare W_{i-1} und Konstanten c_i , sodass $W_{i-1} \leq D_i := M_i - M_{i-1} \leq W_{i-1} + c_i$ IP-f.s.

Wir sehen:

$$M_1 = \mathbb{E}(N_0 | \mathcal{F}_1) = \mathbb{E}(N_0 | \mathcal{F}_0) = M_0 = \mathbb{E}(N_0)$$

(die Information, wo das erste Teil landet, liefert keine Information über N_k)

also folgt schonmal $W_0 = c_1 = 0$

Nun kommen die X_i ins Spiel:

X_i sind die Anzahl verschiedener Werte in $\{X_1, \dots, X_i\}$

\Rightarrow Anzahl der Fächer mit mindestens einer Kugel nach i Platzierungen

Naja, was passiert, wenn wir die Situation unter F_i bedingen?

Wie viele Kugeln verbleiben? $n-i$

Wie viele leere Fächer? $m-Y_i$

Wie groß ist die W'keit, dass ein Fach bis zum Ende leer bleibt?

$(1 - \frac{1}{m})^{n-i}$ für alle $m-Y_i$ leere Fächer

Also gilt:

$$M_i = \mathbb{E}(N_0 | F_i) = \underbrace{(m - Y_i)}_{\text{Anzahl leerer Fächer}} \underbrace{\left(1 - \frac{1}{m}\right)^{n-i}}_{\text{W'keit, dass diese leer bleiben}}$$

Wenn man sich jetzt noch überlegt, dass

$$Y_i = \begin{cases} Y_{i-1} & \text{falls } X_i \text{ in ein gleiches Fach geht} \\ Y_{i-1} + 1 & \text{falls } X_i \text{ anders als } X_0, \dots, X_{i-1} \end{cases}$$

Dann erhalten wir eine entsprechende Abschätzung für die D_i :

$$D_i = \begin{cases} \frac{m - Y_{i-1}}{m} \left(1 - \frac{1}{m}\right)^{n-i} & \text{falls } X_i \in \{X_1, \dots, X_{i-1}\} \\ -\frac{Y_{i-1}}{m} \left(1 - \frac{1}{m}\right)^{n-i} & \text{falls } X_i \notin \{X_1, \dots, X_{i-1}\} \end{cases}$$

(bsp.w. Rechnung für den ersten Teil:

$$\begin{aligned} D_i &= M_i - M_{i-1} = (m - Y_{i-1}) \left(1 - \frac{1}{m}\right)^{n-i} - (m - Y_{i-1}) \left(1 - \frac{1}{m}\right)^{n-i+1} \\ &= (m - Y_{i-1}) \cdot \left(1 - \frac{1}{m}\right)^{n-i} \left(1 - \left(1 - \frac{1}{m}\right)\right) \\ &= \frac{m - Y_{i-1}}{m} \left(1 - \frac{1}{m}\right)^{n-i} \end{aligned}$$

und wir erhalten abschließend

$$W_{i-1} = -\frac{Y_{i-1}}{m} \left(1 - \frac{1}{m}\right)^{n-i}, \quad c_i = \left(1 - \frac{1}{m}\right)^{n-i}$$

Außerdem ergibt sich:

$$\begin{aligned}\sum_{i=1}^n C_i^2 &= \sum_{i=1}^n \left(1 - \frac{1}{n}\right)^{2(n-i)} = \sum_{i=0}^{n-1} \left(1 - \frac{1}{n}\right)^{2i} = \sum_{i=0}^{n-1} \left(\left(1 - \frac{1}{n}\right)^2\right)^i \\ &= \frac{\left(1 - \frac{1}{n}\right)^{2n} - 1}{\left(1 - \frac{1}{n}\right)^2 - 1} = \frac{\left(1 - \frac{1}{n}\right)^{2n} - 1}{-\frac{2}{n} + \frac{1}{n^2}} = \frac{n^2}{2n-1} \left(1 - \left(1 - \frac{1}{n}\right)^{2n}\right) \\ &\approx \frac{n}{2} \cdot (1 - e^{-2\theta}) \quad \left(\frac{n}{n} = \theta \Rightarrow n = \frac{n}{\theta}\right)\end{aligned}$$

Damit folgt

$$\begin{aligned}P(N_0 - \mathbb{E}N_0 \geq t \cdot \mathbb{E}N_0) &\leq \exp\left(-\frac{2t^2 (\mathbb{E}N_0)^2}{\frac{n^2}{2n-1} \left(1 - \left(1 - \frac{1}{n}\right)^{2n}\right)}\right) \\ &\approx \exp\left(-\frac{4nt^2 e^{-2\theta}}{1 - e^{-2\theta}}\right)\end{aligned}$$

Es ergibt sich damit beispielsweise für $t=0,16$ (mehr als 13 leere Fächer)

$$\Rightarrow \exp\left(-\frac{4 \cdot 20 \cdot 0,16^2 \cdot e^{-1}}{1 - e^{-1}}\right) \approx 0,303647$$