Seminar Randomisierte Algorithmen:

Ungleichungen und Restwahrscheinlichkeiten

Jan Oppor und Patrick Fricker

Betreuung:
Prof. Dr. Rolf Klein, Dr. Elmar Langetepe,
Annette Ebbers-Baumann und Tom Kamphans

April 2003

Inhaltsverzeichnis

1 Chernoff-Ungleichungen

1.1 Einleitung

Durch die Markov- und Chebycheff-Ungleichungen können wir für allgemeine Zufallsvariablen Abschätzungen der Art $\mathbf{Pr}[X \geq t] \leq \epsilon$ angeben. Hier sollen zunächst Summen unabhängiger Zufallsvariablen aus Poisson-Versuchen betrachtet werden. Somit sind schärfere Schranken zu erwarten. Solche Schranken (so genannte Chernoff-Schranken) werden nach einem einführenden Beispiel erklärt und bewiesen.

Am Beispiel des Routing-Problems in Parallel-Computern wird die Anwendung dieser Technik erläutert. Dabei wird eine verblüffende Laufzeit-Abschätzung für einen randomisierten Algorithmus gezeigt.

1.2 Einführendes Beispiel

Als einführendes Beispiel wollen wir die Spiele des 1. FC Köln betrachten. Der 1. FC Köln gewinnt ein Spiel mit Wahrscheinlichkeit $\frac{1}{3}$. Gesucht ist eine obere Schranke für die Wahrscheinlichkeit, dass mehr als die Hälfte aller Spiele in einer Saison gewonnen werden.

Sei X_i die Indikatorvariable für den Gewinn des i-ten Spiels, und sei $Y_n = \sum_{i=1}^{n} X_i$ die Zufallsvariable der Anzahl der Siege bei n Spielen. Wir fragen uns also $\mathbf{Pr}[Y_n > \frac{n}{2}] <?$.

Wir wollen nun die so genannte Chernoff-Ungleichung

$$\mathbf{Pr}[X > (1+\delta)\mu] < \left[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\mu}$$

anwenden. Dazu berechnen wir zunächst den Erwartungswert

$$\mathbf{E}[Y_n] = \mu = \frac{n}{3} \ .$$

Dann bestimmt man δ durch

$$(1+\delta)\frac{n}{3} = \frac{n}{2}, \text{ somit } \delta = \frac{1}{2}.$$

Nun kann man einsetzen:

$$\mathbf{Pr}[Y_n < \frac{n}{2}] < \left[\frac{\exp(\frac{1}{2})}{(1+\frac{1}{2})^{(1+\frac{1}{2})}}\right]^{\frac{n}{3}}$$

und erhält $\approx 0,959^n$. Das heißt, die Restwahrscheinlichkeit fällt mit steigendem n exponentiell ab. Je mehr Spiele der FC absolviert, desto stärker kommt die „wahre Natur“ des 1. FC Köln zum Tragen.

1.3 Chernoff-Ungleichung

Nun wollen wir diese Ungleichung herleiten, die uns so viel Freude bereitet hat. Dazu sind einige Voraussetzungen notwendig. Hier werden Summen von Zufallsvariablen unabhängiger Poisson-Versuche betrachtet. Poisson-Versuche sind Experimente, bei denen die Ergebnisse entweder 1 oder 0 sind. Die Wahrscheinlichkeit, mit der ein Ereignis eintritt, darf bei unterschiedlichen Versuchen verschieden sein. Wir schauen uns also die Zufallsvariable X mit

$$X = \sum_i X_i \text{ für unabhängige } X_i \text{ aus Poisson-Versuchen an.}$$

Um nun die Chernoff-Ungleichung

$$\mathbf{Pr}[X > (1+\delta)\mu] < \left[\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right]^\mu$$

zu beweisen werden einige Tricks benutzt. Zum einen wird e^{tX} anstatt X für ein beliebiges t betrachtet. Durch die Unabhängigkeit ist der Erwartungswert der Produkte das Produkt der Erwartungswerte. Und anschließend wählen wir das t geschickt, um eine gute Schranke zu erhalten.

1.4 Beweis

Durch Substitution der Zufallsvariable X durch e^{tX} erhalten wir

$$\mathbf{Pr}[X > (1+\delta)\mu] = \mathbf{Pr}[e^{tX} > e^{t(1+\delta)\mu}].$$

Mit Hilfe der Markov-Ungleichung ergibt sich die Abschätzung

$$\mathbf{Pr}[e^{tX} > e^{t(1+\delta)\mu}] < \frac{\mathbf{E}[\exp(tX)]}{e^{t(1+\delta)\mu}}.$$

Umformen des Erwartungswertes $\mathbf{E}\left[e^{tX}\right]$ führt zu

$$\begin{aligned} \mathbf{E}[\exp(tX)] &= \mathbf{E}\left[\exp(t\sum_{i=1}^n X_i)\right] \\ &= \mathbf{E}\left[\prod_{i=1}^n \exp(tX_i)\right] \\ &= \prod_{i=1}^n \mathbf{E}[\exp(tX_i)]\,. \end{aligned}$$

Der Erwartungswert der Zufallsvariablen X_i ist

$$\mathbf{E}[\exp(tX_i)] = p_i e^t + (1-p_i).$$

Mit

$$\mathbf{Pr}[X > (1+\delta)\mu] < \frac{\prod_{i=1}^n \mathbf{E}[\exp(tX_i)]}{\exp(t(1+\delta)\mu)}$$

und

$$\mathbf{E}\Big[e^{tX_i}\Big] = p_i e^t + (1-p_i)$$

ergibt sich:

$$\mathbf{Pr}[X > (1+\delta)\mu] < \frac{\prod_{i=1}^{n}(1+p_i(e^t-1))}{\exp(t(1+\delta)\mu)}.$$

Mit $(1+x) < e^x = \sum\limits_{i=0}^{\infty} \frac{x^i}{i!}$ hat man:

$$\begin{aligned}
\mathbf{Pr}[X > (1+\delta)\mu] &< \tfrac{\prod_{i=1}^{n}\exp(p_i(e^t-1))}{\exp(t(1+\delta)\mu)} \\
\mathbf{Pr}[X > (1+\delta)\mu] &< \tfrac{\exp\left(\sum_{i=1}^{n}p_i(e^t-1)\right)}{\exp(t(1+\delta)\mu)} \\
&= \tfrac{\exp((e^t-1)\mu)}{\exp(t(1+\delta)\mu)}.
\end{aligned}$$

Das Minimum des rechten Terms lässt sich durch differenzieren und 0-setzen ermitteln und liegt bei $t = \ln(1+\delta)$.

$$\begin{aligned}
\mathbf{Pr}[X > (1+\delta)\mu] &< \tfrac{\exp((e^{\ln(1+\delta)}-1)\mu)}{\exp(\ln(1+\delta)(1+\delta)\mu)} \\
&= \left[\tfrac{\exp(1+\delta-1)}{\exp(\ln(1+\delta)+\ln(1+\delta)\delta)}\right]^\mu \\
&= \left[\tfrac{\exp(1+\delta-1)}{(1+\delta)(1+\delta)^\delta}\right]^\mu \\
&= \left[\tfrac{e^\delta}{(1+\delta)^{(1+\delta)}}\right]^\mu.
\end{aligned}$$

Damit ist die obige Chernoff-Ungleichung gezeigt.

1.5 Abschätzung für Abweichungen nach unten

Nun sollen Werte *kleiner* einer Schranke betrachtete werden. Dies verläuft analog. Man erhält:

$$\mathbf{Pr}[X < (1-\delta)\mu] < \left[\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right]^\mu$$

Für $\delta \in (0,1]$ findet man

$$\mathbf{Pr}[X < (1-\delta)\mu] < \exp(\frac{-\mu\delta^2}{2})$$

durch die Taylor-Entwicklung von $\ln(1-\delta)$ an der Stelle 0. Wir definieren für die beiden Richtungen:

$$F^+(\mu,\delta) := \left[\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right]^\mu$$

bzw.

$$F^-(\mu,\delta) := \exp(\frac{-\mu\delta^2}{2}).$$

1.6 Beispiel

Der 1. FC Köln bekommt einen neuen Trainer, mit dem die Wahrscheinlichkeit ein Spiel zu gewinnen auf $\frac{3}{4}$ steigt. Sei X_i und Y_n wie oben definiert.

Gesucht ist eine obere Schranke für die Wahrscheinlichkeit, dass *weniger* als die Hälfte aller Spiele in einer Saison gewonnen werden. Zunächst wird der Erwartungswert berechnet:

$$\mathbf{E}[Y_n] = \mu = \frac{3}{4}n$$

Das zugehörige δ berechnet sich durch

$$(1-\delta)\mu = \frac{n}{2};\ \delta = \frac{1}{3}n.$$

Durch Einsetzen in die Chernoff-Ungleichung erhält man die Abschätzung:

$$\mathbf{Pr}[Y_n < \frac{n}{2}] < F^-(\mu, \delta) < 0.96^n.$$

Bei steigender Anzahl von Spielen geht damit die Wahrscheinlichkeit, weniger als die Hälfte der Spiele zu gewinnen gegen 0.

1.7 Die rechte Seite

Für eine bessere Handhabe für den rechten Teil der Gleichung definieren wir:

$$\mathbf{Pr}[X < (1-\delta)\mu] < \exp(\frac{-\mu\delta^2}{2}) = F^-(\mu, \delta).$$

Nun können wir für ein gegebenes ϵ ein δ wählen, d. h. für ϵ ist ein δ gesucht, sodass

$$\mathbf{Pr}[X < (1-\delta)\mu] < \epsilon \text{ gilt.}$$

Wir suchen $\Delta^-(\mu, \epsilon)$ mit

$$F^-(\mu, \Delta^-(\mu, \epsilon)) = \epsilon.$$

Dies ergibt sich durch

$$\Delta^-(\mu, \epsilon) = \sqrt{\frac{2 \ln 1/\epsilon}{\mu}}.$$

Sucht man eine ähnliche Formel für

$$\mathbf{Pr}[X > (1+\delta)\mu] < \left[\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right]^\mu = F^+(\mu, \delta),$$

so muss für prägnante Formeln eine Fallunterscheidung bei den Abschätzungen gemacht werden:
Für kleine δ zwischen 0 und $2e-1$:

$$\Delta^+(\mu, \epsilon) < \sqrt{\frac{4 \ln 1/\epsilon}{\mu}}$$

Für große $\delta > 2e - 1$:

$$\Delta^+(\mu, \epsilon) < \frac{\log_2 1/\epsilon}{\mu} - 1$$

1.8 Beispiel

Wir verteilen n Bälle auf n Eimer. Sei Y_1 die Anzahl der Bälle im ersten Eimer. Wir wollen nun eine Abschätzung mit der Chernoff-Ungleichung erhalten. Gesucht ist ein m mit

$$\mathbf{Pr}[Y_1 > m] \leq 1/n^2.$$

Dazu raten wir, dass

$$\Delta^+ > 2e - 1$$

ist. Mit $\Delta^+(1, 1/n^2)$ findet man

$$m < 2\log_2 n - 1,$$

also

$$\mathbf{Pr}[Y_1 > 2\log_2 n - 1] \leq 1/n^2.$$

2 Routing in einem Parallel-Computer

Nun betrachten wir ein größeres Beispiel. Wir wollen Nachrichten in einem Netzwerk von Prozessoren verschicken. Wir modellieren dies als ein Problem auf Graphen. Gegeben ist ein gerichteter Graph $G = (V, E), V = \{0, \ldots, N-1\}$, wobei jedes $v \in V$ einen Prozessor darstellt und jedes $e \in E$ zu einer (gerichteten) Verbindung korrespondiert. Durch jede Verbindung kann pro Schritt/Takt ein Paket (eine Nachricht) weitergereicht werden.

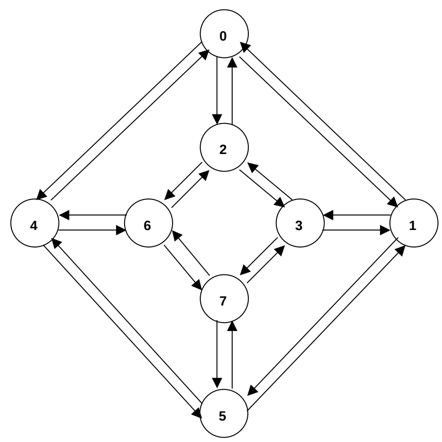

Abbildung 1: Netzwerk mit acht Prozessoren.

2.1 Permutation-Routing-Problem

Wir stellen uns vor, dass jeder Knoten $i, 1 \leq i \leq N$ ein Paket ν_i besitzt. Für jedes dieser Pakete ν_i gibt es einen Ziel-Knoten $d(i)$. Die $d(i)$ sind paarweise verschieden, sodass man diese als Permutationen von $\{1, \ldots, N\}$ auffassen kann.

Das Permutation-Routing-Problem besteht nun darin, die Anzahl der Schritte, die man benötigt, um alle Pakete ans Ziel zu bringen, klein zu halten. Jede ausgehende Kante soll durch eine Queue versehen werden, weil pro Schritt jede Kante nur von höchstens einem Paket genutzt werden darf. Die Größe der Queue ist damit durch N beschränkt.

2.2 Vergessliche Algorithmen

Hier soll nun eine Untermenge der Algorithmen betrachtet werden, die zur Lösung dieses Problems möglich wären, nämlich die so genannten „vergesslichen" Algorithmen. Diese berechnen für jedes Paket eine Route von Knoten i nach Knoten $d(i)$, ohne bei der Berechnung einer Route die anderen Quellen j und Ziele $d(j)$ für $j \neq i$ zu berücksichtigen.

Wir betrachten den Spezialfall eines Netzwerkes, welches die Struktur eines boolschen Hyper-Würfels aufweist. Das heißt, wir haben $N = 2^n$ Knoten, die wir so durchnummerieren, dass jeder Knoten eine Binär-Zahl der Länge n erhält, die sich um jeweils ein Bit von den Nummern der Nachbarknoten unterscheidet (siehe Abbildung 2).

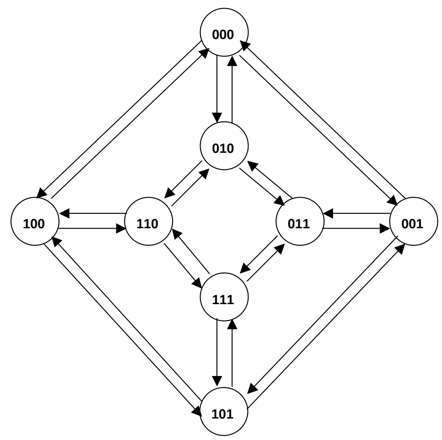

Abbildung 2: Knoten in Binär-Darstellung

2.3 Der Bit-fixing Algorithmus

Eine elegante Lösung ist durch den „Bit-fixing Algorithmus“ gegeben. Dabei wird jedes Bit des Ziels $\sigma(i)$ mit dem Bit von der aktuellen Position des Pakets ν_i von links nach rechts verglichen. Man betrachtet nun das erste Bit, das nicht übereinstimmt. Dieses Bit wird nun angepasst (gefixed), d. h. das Paket wird zu demjenigen Nachbarknoten geschickt, dessen Nummer auch in diesem Bit mit dem Ziels übereinstimmt.

Soll zum Beispiel ein Paket von (1011) nach (0000) gesendet werden, so wäre der Pfad: (1011), (0011), (0001), (0000). Unglücklicherweise lässt sich zeigen, dass dieser deterministische Algorithmus eine schlechte worst-case-Laufzeit hat ($\Omega(\sqrt{2^n/n})$; siehe Borodin und Hopcroft [1]).

2.4 Randomisierter Algorithmus

Der Algorithmus selber ist verblüffend einfach, und deshalb eine weitere Betrachtung wert. Wir wollen versuchen, die erwartete Laufzeit einer randomisierten Version klein zu halten.

Dazu wird jedes Paket ν_i zwei Phasen durchlaufen. Es wird ein zufälliges Zwischenziel $\sigma(i)$ für alle $i \in \{1, \ldots, N\}$ gewählt. Das Paket ν_i wird in der ersten Phase nach $\sigma(i)$ versendet. Sind alle Pakete am Zwischenziel, so beginnt Phase Zwei, bei der jedes Paket von dort aus zum eigentlichen Ziel versendet wird. In jeder Phase wird die Bit-fixing Strategie für die Routen-Festlegung benutzt.

Wie viele Schritte sind notwendig, damit alle ν_i ihre Ziele erreichen?

2.5 Anzahl der benötigten Schritte

Wir betrachten nun zunächst nur die erste Phase. Für einen beliebigen Knoten i sei $\rho_i = (e_1, \ldots, e_k)$ die Route des Pakets ν_i. Die Anzahl der Schritte setzt sich zusammen aus der Länge der Route von ρ_i, die höchstens n betragen kann, und der Verzögerung durch die Queues. Sei S die Menge von Paketen (außer ν_i), die mindestens eine Kante aus ρ_i benutzten, also die Knoten, die eine Verzögerung von ν_i auslösen können. Die Verzögerung für das Paket ν_i kann höchstens $|S|$ sein.

2.6 Abschätzung

Für die Berechnung der Anzahl der Schritte definieren wir nun die Indikatorvariablen H_{ij}, wobei H_{ij} gleich Eins ist, wenn sich die Pfade ρ_i und ρ_j mindestens eine Kante teilen, und sonst den Wert Null annimmt. Mit dem oben Gesagten ist somit die Verzögerung von ν_i höchstens $\sum_{j=1}^{N} H_{ij}$.

Sei für eine Kante e im Hyper-Würfel $T(e)$ eine Zufallsvariable, welche die Anzahl der Routen durch e bestimmt. Für eine feste Route $\rho_i = (e_1, \ldots, e_k)$ mit $k \leq n$ ist die Summe über die schneidenden Pakete H_{ij} kleiner gleich der Summe über die Anzahl der Routen durch die Kanten, also $\sum_{j=1}^{N} H_{ij} \leq \sum_{l=1}^{k} T(e_l)$. Der Erwartungswert lässt sich dementsprechend abschätzen durch

$$E[\sum_{j=1}^{N} H_{ij}] \leq \sum_{l=1}^{k} E[T(e_l)].$$

Offensichtlich ist der Erwartungswert der Länge von ρ_j gleich $n/2$ für jedes j. Somit beträgt der Erwartungswert der Summe über alle Pfade von Paketen $Nn/2$. Da im boolschen Hyper-Würfel jeder Knoten mit n Kanten verbunden ist, gibt es insgesamt Nn Kanten. Der Erwartungswert für $T(e)$ ist somit $1/2$ für alle e. Mit der obigen Abschätzung ist der Erwartungswert von $\sum_{j=1}^{N} H_{ij}$ kleiner gleich $k/2$ und somit kleiner gleich $n/2$.

2.7 Chernoff-Ungleichung

Mit Hilfe der Chernoff-Ungleichung wollen wir die Anzahl der benötigten Verzögerungen abschätzen. Als gutes Limit hat sich $6n$ erwiesen ($\delta = 11$), d.h. wir suchen zunächst eine obere Schranke dafür, dass die Anzahl der Verzögerungen höchstens Sechs mal so groß ist wie die Anzahl der Bits.

Die H_{ij} sind unabhängige Zufallsvariablen, wodurch die Anwendung der Chernoff-Ungleichung möglich ist. Durch Einsetzen in die Chernoff-Ungleichung erhält man als Schranke 2^{-6n}. Da es insgesamt 2^n Pakete gibt, liegt die Wahrscheinlichkeit, dass ein Paket die $6n$ überschreitet damit bei 2^{-5n}.

Die Länge der Route ist durch n beschränkt, womit die Anzahl der Schritte zur Verteilung aller Pakete $7n$ beträgt, mit der Gegen-Wahrscheinlichkeit $1 - 2^{-5n}$. Bei der Betrachtung der Phase Zwei kann analog zur Phase Eins vorgegangen werden. Also lässt sich die Gesamt-Wahrscheinlichkeit für die Anzahl der Verzögerungen beider Phasen folgendermaßen abschätzen:

$$2(\frac{1}{32})^n \leq (\frac{1}{16})^n < (\frac{1}{2})^n = \frac{1}{N}$$

Nun wissen wir, dass der Erwartungswert für die Anzahl von Schritten für alle Pakete weniger als 15n beträgt, denn

$$14n(1-\frac{1}{N})+(nN+14n)\frac{1}{N}=14n-14n\frac{1}{N}+n+14n\frac{1}{N}=15n.$$

Das heißt, mit Wahrscheinlichkeit $(1-1/N)$ werden $14n$ Schritte gebraucht, wie weiter oben berechnet; mit Wahrscheinlichkeit $(1/N)$ schlägt das Verfahren fehl, und stattdessen wird anders vorgegangen: es wird einfach ein Paket nach dem anderen unter Benutzung des Bit-fixing Algorithmus von i nach $d(i)$ verschickt, also nN Schritte. Hinzu kommen noch die $14n$ verschwendeten Schritte.

2.8 Fazit

Dies ist ein überraschendes Ergebnis, da es keinen deterministischer Algorithmus gibt, bei dem die Anzahl der Schritte annähernd so niedrig abgeschätzt werden kann wie der Erwartungswert der randomisierten Version.

3 Martingale

Nun möchten wir uns einer etwas allgemeineren Situation zuwenden. Wir wollen Abschätzungen für Wahrscheinlichkeiten von Zufallsvariablen betrachten, für die nicht die Unabhängigkeit vorausgesetzt wird. Dazu werden wir so genannte Martingale betrachten, was auch einen kurzen Blick auf Filtrationen einschließt. Tatsächlich wird in zwei Schritten vorgegangen. Zunächst soll mit einer einfachen Definition etwas Intuition für Martingale geschaffen werden, um danach mit Hilfe einer allgemeineren Definition zu guten Schranken zu kommen. Abschließend wird eine solche Schranke auf ein Problem mit Bällen und Eimern angewandt.

3.1 Motivation

Wir betrachten als einleitendes Beispiel das Roulette-Spiel. Dazu wird das Vermögen des Spielers nach dem Durchgang i durch die Zufallsvariable X_i beschrieben. Die X_i sind also nicht unabhängig.

Die Strategie des Spielers besteht darin, mit einem Einsatz von 1 beginnend bei Verlust den Einsatz für das nächste Spiel zu verdoppeln und ansonsten den Gewinn einzustreichen. Ursprünglich wurden solche Systeme als Martingale bezeichnet. Hier wird dieser Begriff etwas allgemeiner gefasst werden.

3.2 Grundbegriffe

Zunächst wenden wir uns gemeinsamen Verteilungen und bedingten Wahrscheinlichkeiten zu. Ist die gemeinsame Verteilung von $\mathbf{Pr}[X = x]$ und $\mathbf{Pr}[Y = y]$ gegeben durch

$$p(x, y) = \mathbf{Pr}[\{X = x\} \cap \{Y = y\}],$$

so lässt sich die bedingte Wahrscheinlichkeit schreiben als

$$\mathbf{Pr}[X = x|Y = y] = \frac{p(x, y)}{\mathbf{Pr}[Y = y]} = \frac{p(x, y)}{\sum_x p(x, y)}.$$

Darunter versteht man die Wahrscheinlichkeit von $X = x$ unter Kenntnis von $Y = y$. In Tabelle 1 wird dies an Hand eines Beispiels demonstriert.

Die bedingte Erwartung $\mathbf{E}[X|Y = y]$ ist gegeben durch

$$\frac{\sum_x xp(x, y)}{\sum_x p(x, y)}.$$

Unter dem Erwartungswert $\mathbf{E}[X|Y]$ versteht man die Zufallsvariable, die durch die Funktion $f(y) = \mathbf{E}[X|Y = y]$ veranschaulicht werden kann.

Wir wollen dies mit Hilfe eines Beispiels klarer machen. Ein Würfel wird n mal geworfen. Die Zufallsvariable X_i sei die Anzahl der Würfe mit Augenzahl i. Hier fragen wir uns wie

N \ X_1	0	1	2	3	$\mathbf{Pr}[N = n]$
1	2/27	0	0	1/27	1/9
2	6/27	6/27	6/27	0	2/3
3	0	6/27	0	0	2/9
$\mathbf{Pr}[X_1 = x_1]$	8/27	12/27	6/27	1/27	

Tabelle 1: Beispiel für gemeinsame Verteilungen

der Erwartungswert von X_1 (Anzahl der geworfenen Einsen) unter Kenntnis von X_2 ist. Einfach sieht man, dass gilt:

$$\mathbf{E}[X_1|X_2] = \frac{n - X_2}{5}.$$

Ebenso lässt sich das Beispiel um die Kenntnis der Anzahl der geworfenen Dreien erweitern:

$$\mathbf{E}[X_1|X_2, X_3] = \frac{n - X_2 - X_3}{4}$$

Eine interessante Eigenschaft der bedingten Erwartungswerte lässt sich zeigen durch:

$$\begin{aligned}
\mathbf{E}[\mathbf{E}[X|Y]] &= \sum_y (\mathbf{E}[X|Y = y]\mathbf{Pr}[Y = y]) \\
&= \sum_y (\tfrac{\sum_x xp(x,y)}{\mathbf{Pr}[Y=y]}\mathbf{Pr}[Y = y]) \\
&= \sum_y \sum_x xp(x, y) \\
&= \sum_x \sum_y xp(x, y) \\
&= \sum_x x\mathbf{Pr}[X = x] = \mathbf{E}[X]
\end{aligned}$$

Es gilt also $\mathbf{E}[\mathbf{E}[X|Y]] = \mathbf{E}[X]$.

3.3 Einfache Definition

Nun sind wir gerüstet, um eine erste Definition für Martingale angeben zu können. Dazu sei $X_0, X_1, \ldots$ eine Sequenz von Zufallsvariablen. Ist nun für alle i die bedingte Erwartung von X_i unter den Kenntnis von $X_0, \ldots, X_{i-1}$

- gleich X_{i-1}, so nennt man die Sequenz ein Martingal;
- für den Fall $\mathbf{E}[X_i|X_0, \ldots, X_{i-1}] \leq X_{i-1}$ nennt man die Sequenz ein Supermartingal,
- für $\mathbf{E}[X_i|X_0, \ldots, X_{i-1}] \geq X_{i-1}$ ein Submartingal.

Für Martingale ist für alle $i \geq 0$ $\mathbf{E}[X_i] = \mathbf{E}[X_0]$, da nach Definition

$$\mathbf{E}[E[X_i|X_0, \ldots, X_{i-1}]] = E[X_{i-1}]$$

ist, und durch die oben gezeigte Eigenschaft $\mathbf{E}[\mathbf{E}[X|Y_1, \ldots, Y_n]] = \mathbf{E}[X]$ gilt.

3.4 Generalisierung

Nun möchten wir den Begriff des Martingals so erweitern, dass Ausdrücke wie

$$\mathbf{E}[X_i|Z_0, \ldots, Z_{i-1}] = X_{i-1}$$

möglich sind, d. h. wir wollen beliebiges Wissen über die letzten Spiele betrachten können. Dazu sind allerdings einige Begriffe notwendig.

3.5 Filtration

Sei $\{\mathbb{F}_0, \ldots, \mathbb{F}_n\}$ eine Menge für die $\mathbb{F}_0 \subseteq \mathbb{F}_i \subseteq \ldots \subseteq \mathbb{F}_n$ gilt und in der für jedes $\mathbb{F}_i$ $(\Omega, \mathbb{F}_i)$ eine σ-Algebra ist. Dabei sei $\mathbb{F}_0 = \{\emptyset, \Omega\}$ und $\mathbb{F}_n = 2^\Omega$. Dann ist $\mathbb{F}_0 \subseteq \mathbb{F}_i \subseteq \ldots \subseteq \mathbb{F}_n$ eine Filtration.

Dies lässt sich mit Hilfe einer Partitionierung veranschaulichen. Sei $\Omega = \dot{\bigcup}_j B_j^i$, sodass die B_j^i die σ-Algebra $\mathbb{F}_i$ generieren. Dann muss die Partitionierung von $\mathbb{F}_{i+1}$ eine Verfeinerung der Partitionierung von $\mathbb{F}_i$ sein.

Beispiel.
Wir betrachten einen randomisierten Algorithmus, der n Zufalls-Bits benutzt. Sei Ω die Menge der 2^n Binärwörter der Länge n. Wir partitionieren durch B_w^i mit $w \in \{0,1\}^i$, die Partitionen entsprechen also den Mengen von Zahlen, die den selben Bit-Präfix besitzen. Dadurch werden $\mathbb{F}_i$ generiert. Durch diese Partitionierung ist $\mathbb{F}_0, \ldots, \mathbb{F}_n$ eine Filtration. So wäre $\mathbb{F}_2 = pot(B_{00}^2, B_{01}^2, B_{10}^2, B_{11}^2)$. Die Partitionierung ist in Abbildung 3 dargestellt.

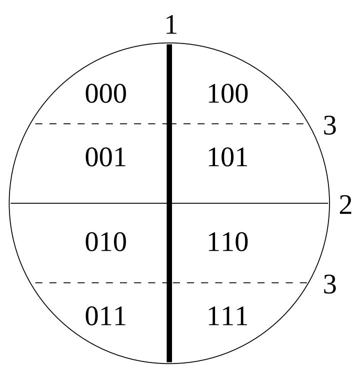

Abbildung 3: Partitionierung von Zahlen aus drei Bits

3.6 $\mathbb{F}$-messbar

Sei X eine Zufallsvariable und für alle k sei $\{X \leq k\} \in \mathbb{F}_i$. Dann heißt X $\mathbb{F}_i$-messbar. Ist X $\mathbb{F}_i$-messbar, so ist X auch $\mathbb{F}_j$-messbar für $j \geq i$.

Wir betrachten zwei Beispiele. X_j sei die Zufallsvariable für die Anzahl der Einsen in den ersten j Bits. X_2 ist $\mathbb{F}_2$-messbar, denn für jedes k ist $\{X_2 \geq k\} \in \mathbb{F}_2$ (siehe Abbildung

3). Anders sieht das bei der Parität aus. Es lässt sich kein $\mathbb{F}_i$ mit $i < 3$ finden, sodass $\{X = 1\} \in \mathbb{F}_i$ gilt. Somit ist die Zufallsvariable für die Parität nur $\mathbb{F}_3$-messbar.

3.7 Bedingter Erwartungswert bezüglich $\mathbb{F}_j$

Ist X $\mathbb{F}_i$-messbar, so ist X nicht unbedingt $\mathbb{F}_{i-1}$-messbar. D. h. es kann ein k mit $\{X \geq k\} \notin \mathbb{F}_{i-1}$ existieren. Es stellt sich nun die Frage, wie sich X auf der Partitionierung von $\mathbb{F}_{i-1}$ verhält.

Dazu soll der bedingte Erwartungswert $\mathbf{E}[X|\mathbb{F}_{i-1}]$ von X auf Partitionen von $\mathbb{F}_{i-1}$ bestimmt werden. Hierbei handelt es sich wieder um eine Zufallsvariable. Dies lässt sich dann auf $\mathbf{E}[X|\mathbb{F}_j]$ für beliebige $j < i$ erweitern.

Für die Definition für einen solchen bedingten Erwartungswert soll die obige Definition von bedingten Erwartungswerten von Zufallsvariablen herangezogen werden: $\mathbf{E}[X|\mathbb{F}] := \mathbf{E}[X|Y]$ für eine Zufallsvariable Y, die verschiedene Werte für verschiedene Elementar-Ereignisse von $\mathbb{F}$ annimmt. Y kann dabei als Indikator für die Elementar-Ereignisse von $\mathbb{F}$ betrachtet werden. Umgekehrt kann man aus $\mathbf{E}[X|Y]$ dann $\mathbf{E}[X|\sigma(Y)]$ ableiten, indem man die Partitionierung so wählt, dass Y auf den Partitionen jeweils den selben Wert annimmt. Dabei steht $\sigma(Y)$ für die kleinste σ-Algebra, für die Y messbar ist.

Beispiel.
Angenommen, der Zufallsraum besteht aus den Menschen in NRW. So könnte man eine Filtration generieren, indem man nach Geschlecht, Größe, Alter und schließlich nach den Individuen partitioniert.

3.8 Allgemeine Definition von Martingalen

Sei $\mathbb{F}_0, \mathbb{F}_1, \ldots$ eine Filtration und seien $X_0, X_1, \ldots$ Zufallsvariablen. Ist X_i $\mathbb{F}_i$-messbar, so ist $X_0, \ldots, X_n$ ist ein Martingal, falls

$$\mathbf{E}[X_{i+1}|\mathbb{F}_i] = X_i$$

gilt. Die erste Definition mit $\mathbf{E}[X_{i+1}|X_0, \ldots, X_i] = X_i$ kann als Spezialfall aufgefasst werden, indem die Partitionierung entsprechend gewählt wird.

3.9 Konstruktion von Martingalen aus Zufallsvariablen

Sei $(\Omega, \mathbb{F}, \mathbf{Pr})$ ein Wahrscheinlichkeitsraum und X eine Zufallsvariable bezüglich demselben. Sei $\mathbb{F}_0, \mathbb{F}_1, \ldots$ eine (beliebige) Filtration. Dann ist mit der Definition $X_i := \mathbf{E}[X|\mathbb{F}_i]$ ein (Doob-) Martingal $X_0, \ldots, X_n$ gegeben.

Intuitiv entspricht eine Filtration $\mathbb{F}_0, \mathbb{F}_1, \ldots$ wachsendem Wissen. Die Zufallsvariablen X_i sind die sich verändernden Erwartungswerte mit zunehmendem Wissen. Falls die Differenzen zwischen den X_i und X_{i-1} beschränkt sind, so verhält sich X_n ähnlich wie X_0. Für das Beispiel des randomisierten Algorithmus bedeutet das, dass die Laufzeit des Algorithmus nahe an dessen Erwartungswert liegen würde. Dies soll im Folgenden etwas formaler ausgedrückt werden.

3.10 Azumas Ungleichung

Sei $X_0, X_1, \ldots$ ein Martingal. Ist für jedes k $|X_k - X_{k-1}| \leq c_k$, wobei c_k von k abhängen kann, so gilt für alle $t \geq 0$ und $\lambda > 0$ die Ungleichung

$$\mathbf{Pr}[|X_t - X_0| \geq \lambda] \leq 2\exp\left(-\frac{\lambda^2}{2\sum_{k=1}^t c_k^2}\right).$$

Beispiel.
Seien $Z_1, \ldots, Z_n$ unabhängige Zufallsvariablen von Poisson-Versuchen und $S = \sum_{i=1}^n Z_i$. Wir wählen $X_i = \mathbf{E}[S|Z_1, \ldots, Z_i]$; dann ist $|X_i - X_{i-1}| \leq 1$. Man berechnet $X_0 = \mathbf{E}[S] = \frac{n}{2}$. Dann erhält man durch die Azuma-Ungleichung

$$\mathbf{Pr}[|S - \mathbf{E}[S]| \geq \lambda] \leq 2\exp(-\frac{\lambda^2}{2n}).$$

Diese Abschätzung ist etwas schlechter als die, die man durch die Chernoff-Ungleichung erhalten würde. Hier wurde die Unabhängigkeit jedoch nicht ausgenutzt.

Eine Vereinfachung ergibt sich für einen Spezialfall. Sei $X_0, X_1, \ldots$ ein Martingal. Gilt für jedes k $|X_k - X_{k-1}| \leq c$, wobei c unabhängig von k ist, so gilt für alle $t \geq 0$ und $\lambda > 0$,

$$\mathbf{Pr}[|X_t - X_0| \geq \lambda c\sqrt{t}] \leq 2e^{-\lambda^2/2}.$$

3.11 Lipschitz Bedingung

Sei $f : \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \rightarrow \mathbb{R}$ eine Abbildung nach $\mathbb{R}$. Gelte weiter für alle $x_1 \in \mathcal{D}_1, \ldots, x_n \in \mathcal{D}_n$ und $y_i \in \mathcal{D}_i$:

$$|f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)| - |f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n)| \leq 1$$

Dann lässt sich für die Zufallsvariablen $X_1, \ldots, X_n$ durch $Y_0 = \mathbf{E}[f(X_1, \ldots, X_n)]$ und $Y_i = \mathbf{E}[f(X_1, \ldots, X_n)|X_1, \ldots, X_i]$ ein Martingal definieren mit $|Y_i - Y_{i-1}| \leq 1$.

3.12 Bälle und Eimer

Zum Abschluss soll ein längeres Beispiel durchgerechnet werden. Wir betrachten dazu das Verteilen von n Bällen auf m Eimer. Die Zufallsvariable Z steht für die Anzahl der leeren Eimer.

Der Erwartungswert lässt sich durch $\mu = \mathbf{E}[Z] = n\left(1 - \frac{1}{n}\right)^m$ berechnen. Im Folgenden soll die Azuma-Ungleichung angewandt werden, um die Abweichung von Z vom Erwartungswert abzuschätzen. Es soll also für $\lambda > 0$

$$\mathbf{Pr}[|Z - \mu| \geq \lambda] \leq 2\exp\left(-\frac{\lambda^2(n - 1/2)}{n^2 - \mu^2}\right)$$

bewiesen werden.

Sei $\mathbb{F}_t$ die durch die ersten t Bälle erzeugte σ-Algebra, und $Z_t = \mathbf{E}[Z|\mathbb{F}_t]$ die bedingte Erwartung zum Zeitpunkt t, dann ist $Z_0, \ldots, Z_n$ ein Martingal; weiter gilt $Z_0 = \mathbf{E}[Z]$ und $Z_m = Z$. Zur einfacheren Handhabe sei weiter

$$z(Y,t) = E[Z|Y \text{ leere Eimer zum Zeitpunkt } t].$$

Die Wahrscheinlichkeit dafür, dass ein leerer Eimer in den nächsten $m-t$ Schritten keinen Ball erhält ist $(1-\frac{1}{n})^{m-t}$. Also ist $z(Y,t)$ gleich $Y(1-\frac{1}{n})^{m-t}$.

Sei Y_t die Anzahl der leeren Eimer zum Zeitpunkt t, dann ist

$$Z_{t-1} = z(Y_{t-1}, t-1) = Y_{t-1}(1-\frac{1}{n})^{m-t+1}.$$

Wir betrachten nun den Schritt von $t-1$ nach t. Mit der Wahrscheinlichkeit $1-\frac{Y_{t-1}}{n}$ geht der t-te Ball in einen nicht leeren Eimer. Dann folgt

$$Z_t = z(Y_t, t) = z(Y_{t-1}, t) = Y_{t-1}(1-\frac{1}{n})^{m-t}.$$

Mit der Wahrscheinlichkeit $\frac{Y_{t-1}}{n}$ geht der t-te Ball in einen leeren Eimer. Dann folgt

$$Z_t = z(Y_t, t) = z(Y_{t-1}-1, t) = (Y_{t-1}-1)(1-\frac{1}{n})^{m-t}.$$

Für das Anwenden der Azuma-Ungleichung müssen wir Schranken für die Differenzen $\Delta_t = Z_t - Z_{t-1}$ zwischen den Z_t finden: Z_{t-1} ist ja gleich $Y_{t-1}(1-\frac{1}{n})^{m-t+1}$. Wir betrachten die zwei Fälle wieder getrennt:

1. mit der Wahrscheinlichkeit $1-\frac{Y_{t-1}}{n}$ gilt $Z_t = Y_{t-1}(1-\frac{1}{n})^{m-t}$.

$$\begin{aligned}\Delta_{t_1} &= Y_{t-1}(1-\tfrac{1}{n})^{m-t} - Y_{t-1}(1-\tfrac{1}{n})^{m-t+1}\\ &= Y_{t-1}(1-\tfrac{1}{n})^{m-t}(1-(1-\tfrac{1}{n}))\\ &= \tfrac{Y_{t-1}}{n}(1-\tfrac{1}{n})^{m-t}\end{aligned}$$

2. mit der Wahrscheinlichkeit $\frac{Y_{t-1}}{n}$ gilt $Z_t = (Y_{t-1}-1)(1-\frac{1}{n})^{m-t}$.

$$\begin{aligned}\Delta_{t_2} &= (Y_{t-1}-1)(1-\tfrac{1}{n})^{m-t} - Y_{t-1}(1-\tfrac{1}{n})^{m-t+1}\\ &= Y_{t-1}(1-\tfrac{1}{n})^{m-t}(1-(1-\tfrac{1}{n})) - (1-\tfrac{1}{n})^{m-t}\\ &= -(1-\tfrac{Y_{t-1}}{n})(1-\tfrac{1}{n})^{m-t}\end{aligned}$$

Offensichtlich gilt $0 \leq Y_{t-1} \leq n$. Durch Einsetzten dieser Werte in die Δ_{t_i} erhält man die Schranken $-(1-\frac{1}{n})^{m-t} \leq \Delta_t \leq (1-\frac{1}{n})^{m-t}$. Dadurch ist für $c_t = (1-\frac{1}{n})^{m-t}$ die Differenz $|Z_t - Z_{t-1}| \leq c_t$.
Durch Rechnen ergibt sich

$$\sum_{t=1}^{m} c_t^2 = \frac{1-(1-\frac{1}{n})^{2m}}{1-(1-\frac{1}{n})^2} = \frac{n^2-\mu^2}{2n-1}.$$

Dies lässt sich nun in Azumas Ungleichung einsetzen, sodass man

$$\mathbf{Pr}[|Z-\mu| \geq \lambda] \leq 2\exp\left(-\frac{\lambda^2(n-1/2)}{n^2-\mu^2}\right)$$

erhält.

3.13 Rückblick

In diesem Abschnitt wurden Zufallsvariable betrachtet, die nicht notwendigerweise unabhängig waren. Dies wurde gehandhabt, indem der bedingter Erwartungswert bezüglich einer σ-Algebra definiert wurde. Damit konnten Filter und durch diese Martingale eingeführt werden. Mit deren Hilfe konnten dann Schranken für den Fall angegeben werden, dass die Differenzen zwischen den Zufallsvariable des Martingals beschränkt sind.

Literatur

[1] A. Borodin und J.E. Hopcroft, »Routing, merging, and sorting on parallel models of computation.«, *Journal of Computer and System Sciences*, 30:130–145, 1985.

[2] R. Motwani und P. Raghavan, *Randomized Algorithms*, Cambridge University Press, 1995.