Die Methode des 2.Moments

Christoph Schmidt

July 13, 2004

1 Einleitung

Die Varianz einer Zufallsvariablen ist ihre mittlere quadratische Abweichung von ihrem Erwartungswert.

$$Var[X] = E[(X - EX)^2] = E[X^2] - E[X]^2$$

Der Term $E[X^2]$ wird auch als 2.Moment bezeichnet.

Ist nun die Varianz sehr klein (viel kleiner als $E[X]^2$), so ist X *fast immer fast gleich* seinem Erwartungswert.

Die Methode des zweiten Momentes ist nun ein Verfahren, welches mithilfe dieses Sachverhaltes die Existenz einer kombinatorischen Struktur mit bestimmten Eigenschaften beweist.

2 Die Chebyshev-Ungleichung

Mathematisch präziser wird obiger Sachverhalt durch die Chebyshev-Ungleichung dargestellt:

$$P(|X - E[X]| \geq \lambda) \leq Var[X]/\lambda^2 \tag{1}$$

Wählt man als einen Spezialfall $\lambda = E[X]$, so erhält man

$$P(X = 0) \leq P(|X - E[X]| \geq E[X]) \leq Var[X]/E[X]^2 \tag{2}$$

Hieraus ergibt sich eine erste Anwendung der Methode des zweiten Momentes: Betrachtet werden soll eine zählbare Eigenschaft, z.B. die Anzahl der Kreise der Länge 5, die ein Graph enthält. Will man nun die Existenz eines Objektes, das mindestens eine solche Struktur enthält, beweisen, so legt man über alle möglichen Objekte einen Wahrscheinlichkeitsraum, modelliert mit einer Zufallsvariablen X die Anzahl der zählbaren Eigenschaft in einem Objekt und beweist mit obiger Gleichung, dass $P(X = 0) < 1$. Folglich muss ein Objekt mit der gewünschten Eigenschaft existieren, da $P(X \geq 1) > 0$.

Allgemein bezeichnet man als ,,Methode des zweiten Moments" die Anwendung der probabilistischen Methode unter Zuhilfenahme der Chebyshev-Ungleichung.

3 Mathematische Grundlagen

3.1 Varianzrechnung

Sei $X = X_1 + ... + X_n$ die Summe von n Zufallsvariablen. Dann ist

$$Var[X] = \sum_{i,j=1}^{n} Cov[X_i, X_j] = \sum_{i=1}^{n} Var[X_i] + \sum_{i \neq j} Cov[X_i, X_j]$$

Hierbei ist die Kovarianz $Cov[X_i, X_j] = E[X_i X_j] - E[X_i]E[X_j]$. Sind X_i und X_j stochastisch unabhängig, so gilt $E[X_i X_j] = E[X_i]E[X_j] \Longrightarrow Cov[X_i, X_j] = 0$. Der Term $Cov[X_i, X_j]$ leistet also keinen Beitrag zur Varianz von X.

3.2 Arithmetisch-geometrische Ungleichung

Im ersten Vortrag wurde die arithmetisch-geometrische Ungleichung bewiesen. Sie wird hier als bekannt vorausgesetzt : Seien $a_1, ..., a_n$ nicht negativ. Dann gilt:

$$\frac{1}{n} \sum_{i=1}^{n} a_i \geq (\prod_{i=1}^{n} a_i)^{\frac{1}{n}} \tag{3}$$

3.3 Markov-Ungleichung

Aus der Stochastik ist die Markov-Ungleichung bekannt:

$$P(|X| > c) \leq \frac{E(|X|)}{c} \qquad \forall c > 0 \tag{4}$$

3.4 Chebyshev-Ungleichung für Folgen

Ebenfalls unter dem Namen Chebyshev-Ungleichung bekannt ist folgende Ungleichung für zwei Zahlenfolgen, wobei $a_1, ..., a_n$ monoton steigend und $b_1, ..., b_n$ monoton fallend ist. Es gilt :

$$\sum_{i=1}^{n} a_i b_i \leq \frac{1}{n} (\sum_{i=1}^{n} a_i)(\sum_{i=1}^{n} b_i) \tag{5}$$

Beweis :

$$\begin{aligned}
\textstyle\sum_i a_i b_i - \frac{1}{n} \sum_i a_i \sum_i b_i \quad & = \sum_i a_i b_i - \frac{1}{n} \sum_i \sum_j a_i b_j \\
= \textstyle\frac{1}{n} \sum_i \sum_j (a_i b_i - a_i b_j) \quad & = \frac{1}{2*n} \sum_i \sum_j (a_i b_i + a_j b_j - a_i b_j - a_j b_i) \\
= \textstyle\frac{1}{2*n} \sum_i \sum_j (a_i (b_i - b_j) - a_j (b_i - b_j)) \quad & = \frac{1}{2*n} \sum_i \sum_j (a_i - a_j)(b_i - b_j) \leq 0
\end{aligned}$$

denn da $a_1, ..., a_n$ monoton steigend und $b_1, ..., b_n$ monoton fallend ist, gilt $(a_i - a_j)(b_i - b_j) \leq 0$ für alle i, j.

4 Anwendung 1 : Mengenseparatoren

Ein Teilbereich der Komplexitätstheorie analysiert die Komplexität der Berechnung von booleschen Funktionen. Um sowohl die Zeit- als auch die Speicherplatzkomplexität der Berechnung zu berücksichtigen, verwendet man als Modell sogenannte ,,Branching programs”:

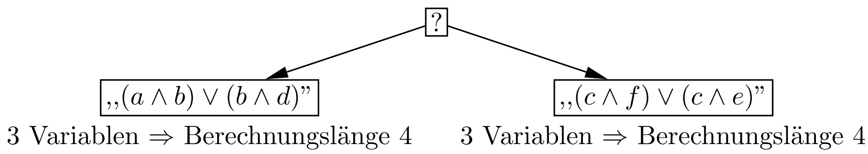
Ein branching program (kurz BP) ist ein gerichteter, azyklischer Graph mit einem bestimmten Startknoten. Jeder Knoten, der keine Senke ist, ist mit einer Variablen beschriftet; die vom Knoten ausgehenden Kanten entsprechen den möglichen Werten der Variablen, 0 oder 1. Die beiden Senken sind mit den Ausgabewerten 0 bzw. 1 beschriftet. Um den Funktionswert einer Variablenbelegung zu bestimmen, folgt man dem Pfad vom Startknoten bis zu einer Senke, wobei die Werte der Eingabevariablen, mit denen die Knoten beschriftet sind,bestimmen, welchen Kanten man folgt. Der Ausgabewert ist dann der Wert der Senke, in die man gelangt. Die Zeitkomplexität einer Funktion ist nun die maximale Länge eines Pfades vom Startknoten zu einer Senke; die Platzkomplexität ist die Anzahl der Knoten des Graphen.

Nichtdeterministische BPs besitzen zusätzlich noch ,,Rate”knoten mit zwei ausgehenden Kanten. Führt eine der aus diesem Knoten ausgehenden Kanten mit der aktuellen Variablenbelegung zu einer 1-Senke, so wird diese Kante in der Berechnung ausgewählt. Ein Rateknoten entspricht also einem logischen ,,$\vee$”. Nichtdeterministische BPs werden zur Abschätzung der Zeitkomplexität einer Funktion nach unten verwendet.

Für eine Funktion f mit n Variablen lässt sich immer ein BP der Tiefe n konstruieren, das f berechnet. In vielen Fällen lässt sich diese Berechnungstiefe mithilfe des Nichtdeterminismus verbessern. Dies wird nun an einem Beispiel demonstriert:

Sei $f = (a \wedge b) \vee (c \wedge f) \vee (b \wedge d) \vee (c \wedge e)$. f enthält 6 Variablen. Daher ist die max. Berechnungslänge 7.

Betrachte nun folgendes BP (der ,,?”-Knoten ist ein Rateknoten, die unteren Knoten stehen für die BPs der entsprechenden Teilformeln):



Wie bereits erwähnt entspricht der ,,Rate”knoten einem logischen ,,$\vee$”. Durch Trennung der Disjunktionsglieder in zwei Gruppen, die jeweils weniger Variablen enthalten, wird eine Reduktion der Berechnungslänge erreicht. In diesem Beispiel hat das nichtdeterministische BP eine Berechnungslänge von $1 + 4 = 5$.

Wie gut ist die Reduktion für eine gegebene Formel? Eine Antwort darauf liefert der Satz von Beame, Saks und Thathatchar(1998). Dazu formulieren wir obige Fragestellung mengentheoretisch um:

Sei $\mathcal{F} = \{F_1, ...F_m\}$ eine Familie von Teilmengen einer Menge X. Ein *Separator* für $\mathcal{F}$ ist ein Paar (S,T) disjunkter Teilmengen von X, wobei jedes Element von $\mathcal{F}$ disjunkt zu S oder zu T ist. Die Größe eines Separators ist das Minimum von $|S|$ und $|T|$.

Der Grad d_x eines Elementes $x \in X$ in $\mathcal{F}$ ist die Anzahl der Elemente von $\mathcal{F}$, die x enthalten.Der durchschnittliche Grad von $\mathcal{F}$ ist

$$d = \frac{1}{|X|} \sum_{x \in X} d_x \tag{6}$$

Auf obiges Problem angewendet ist X die Menge der Variablen, $\mathcal{F}$ ist die Menge der Variablenmengen der Disjunktionsglieder, und die Separatormengen S und T enthalten jeweils die Variablen, die man aus einem Teil-BP eliminieren konnte. Also :
$X = \{a, b, c, d, e, f\}$
$\mathcal{F} = \{\{a, b\}, \{c, f\}, \{b, d\}, \{c, e\}\}$
$S = \{c, e, f\}$, $T = \{a, b, d\}$
Der Grad d_x entspricht der Anzahl des Auftretens der Variablen x.

Mithilfe der Methode des zweiten Momentes zeigten Beame, Saks und Thathachar folgenden Satz über die Existenz eines Separators :

Satz: Sei $\mathcal{F}$ eine Familie von nichtleeren Teilmengen einer n-elementigen Menge, die jeweils höchstens r Elemente enthalten. Sei d der durchschnittliche Grad von $\mathcal{F}$. Dann hat $\mathcal{F}$ einen Separator der Größe mindestens $(1 - \delta)2^{-d}n$, wobei

$$\delta = \sqrt{\frac{dr2^{d+1}}{n}} \tag{7}$$

Beweis: Zum Beweis legen wir einen Wahrscheinlichkeitsraum über alle möglichen Separatoren und beweisen, dass mit positiver Wahrscheinlichkeit ein Separator der Größe mindestens $(1 - \delta)2^{-d}n$ existiert. Dies geschieht wie folgt: Sei $X = \bigcup_{F \in \mathcal{F}} F$ und $n = |X|$. Nun wird jedes $F \in \mathcal{F}$ unabhängig und gleichverteilt mit Wahrscheinlichkeit $\frac{1}{2}$ rot oder blau gefärbt. Sei S die Menge der Elemente x, für die gilt: alle F_i mit $x \in F_i$ sind rot gefärbt. Analog sei T die Menge der Elemente x, für die gilt: alle F_i mit $x \in F_i$ sind blau gefärbt. Da jedes $x \in X$ in mindestens einem F enthalten ist, folgt, dass S und T disjunkt sind. Außerdem gilt für alle $F \in \mathcal{F}$ entweder $F \cap S = \emptyset$ (F ist blau gefärbt) oder $F \cap T = \emptyset$ (F ist rot gefärbt).

Nun muss nur noch mithilfe der Methode des 2. Moments gezeigt werden, dass mit positiver Wahrscheinlichkeit S und T mindestens $(1-\delta)2^{-d}n$ Elemente enthalten.

Sei Z_x die Indikatorvariable des Ereignisses $\{x \in S\}$. Es gilt $P(Z_x = 1) = P(x \in S) = 2^{-d_x}$ und $E[Z_x] = P(Z_x = 1) = 2^{-d_x}$. Nun sei $Z = \sum_x Z_x$. Es gilt $Z = |S|$.

Mit der arithmetisch-geometrischen Mittelwertungleichung (3) folgt:

$$E[Z] = \sum_x E[Z_x] = \sum_x 2^{-d_x} \geq n2^{-\sum_x d_x/n} = n2^{-d} \tag{8}$$

Mithilfe der Methode des 2. Moments soll im Folgenden gezeigt werden, dass Z mit hoher Wahrscheinlichkeit fast gleich seinem Erwartungswert ist.

Dazu beschränken wir die Varianz

$$Var[Z] = \sum_x Var[Z_x] + \sum_{x \neq y} Cov(Z_x, Z_y) \tag{9}$$

nach oben. Da Z_x eine Bernoulli Zufallsvariable (binäre ZV mit Träger {0,1}) ist, gilt:

$$Var[Z_x] = E[Z_x] \underbrace{-E[Z_x]^2}_{\leq 0} \leq E[Z_x]$$

Hieraus folgt:

$$\sum_x Var[Z_x] \leq E[Z] \tag{10}$$

Wenn für zwei Elemente x,y kein $F \in \mathcal{F}$ beide Elemente enthält, so sind Z_x und Z_y stochastisch unabhängig, also unkorreliert $\Rightarrow Cov(Z_x, Z_y) = 0$. Interessant sind also nur x,y, wobei es ein $F \in \mathcal{F}$ gibt mit $x, y \in F$. Für ein festes x ist die Anzahl dieser möglichen Paare maximal $(r-1)d_x$. Für ein solches Paar x,y gilt

$$Cov(Z_x, Z_y) = E[Z_x Z_y] \underbrace{-E[Z_x]E[Z_y]}_{\leq 0} \leq E[Z_x Z_y] \leq E[Z_x] = 2^{-d_x}$$

Es folgt:

$$\sum_{x \neq y} Cov(Z_x, Z_y) \leq (r-1) \sum_x d_x 2^{-d_x} \tag{11}$$

Sortiere $\{d_x\}$ aufsteigend. Dann ist $\{2^{-d_x}\}$ monoton fallend und nicht-negativ. Mit der Chebyshev-Ungleichung für Folgen (5) lässt sich der Term wie folgt abschätzen :

$$\sum_{x \neq y} Cov(Z_x, Z_y) \leq \frac{r-1}{n} (\sum_x 2^{-d_x})(\sum_x d_x) = d(r-1)E[Z] \tag{12}$$

Nun setzt man (10) und (12) in (9) ein und erhält für die Varianz von Z

$$Var[Z] \leq (d(r-1)+1)E[Z] \leq drE[Z],$$

da jedes $x \in X$ in mind. einem F vorkommt und somit $d = \frac{1}{n} \sum_x d_x \geq 1$ gilt.

Schließlich wendet man die Chebyshev-Ungleichung (1) an:

$$P(Z < (1-\delta)E[Z]) \leq P(|Z - E[Z]| > \delta E[Z]) < \frac{Var[Z]}{\delta^2 E[Z]^2} \leq \frac{dr}{\delta^2 E[Z]}$$

Setzt man nun δ aus (7) und $Z = |S|$ ein, so folgt :

$$P(|S| < (1-\delta)E[Z]) < \frac{drn}{dr2^{d+1}E[Z]} \leq \frac{n}{2^{d+1}n2^{-d}} = \frac{1}{2}$$

Analog lässt sich für $|T|$ $P(|S| < (1-\delta)E[Z]) < \frac{1}{2}$ zeigen. Also haben mit positiver Wahrscheinlichkeit S und T die Mächtigkeit von mind. $(1-\delta)E[Z] \geq (1-\delta)2^{-d}n$. Somit existiert ein Separator (S,T), der das Theorem erfüllt.

5 Schwellwertfunktion für eine 4-Clique im Zufallsgraphen

Ein *Zufallsgraph* G(n,p) ist ein Graph mit n Knoten, wobei jede Kante stochastisch unabhängig mit Wahrscheinlichkeit p im Graphen vorhanden ist. Ist p gering, so enthält G(n,p) wahrscheinlich wenige Kanten, ist p hoch, so enthält G(n,p) wahrscheinlich viele Kanten. Wie hoch ist nun die Wahrscheinlichkeit, dass so ein Graph eine 4-Clique enthält, in Abhängigkeit von p? Man könnte annehmen, dass diese Wahrscheinlichkeit linear mit p steigt.

Tatsächlich gibt es aber einen scharf begrenzten Schwellwert p(n), so dass gilt: Ist p etwas kleiner als p(n), so enthält G(n,p) wahrscheinlich keine 4-Clique; ist p etwas größer als p(n), so enthält G(n,p) wahrscheinlich eine 4-Clique. Folgender Satz bestimmt die Schwellwertfunktion für eine 4-Clique:

Satz: Die Schwellwertfunktion für einen Zufallsgraphen G(n,p), eine 4-Clique zu enthalten, ist $p = n^{-2/3}$.

Beweis: Sei S eine Menge von 4 Knoten, A_S das Ereignis, dass S eine 4-Clique in $G(n,p)$ induziert, und X_S die Indikatorvariable des Ereignisses A_S. Da eine 4-Clique 6 Kanten enthält und diese stochastisch unabhängig mit Wahrscheinlichkeit p im Graphen vorkommen, ist

$$P(A_S) = p^6 \Rightarrow P(X_S = 1) = p^6 \Rightarrow E[X_S] = p^6,$$

wobei letztere Folgerung gilt, da X_S eine Indikatorvariable ist.

Sei nun $X = \sum X_S$ die Anzahl der 4-Cliquen in $G(n,p)$. Um zu zeigen, dass $p = n^{-2/3}$ ein Schwellwert des Ereignisses „G enthält eine 4-Clique" ist, beweisen wir:

1. $P(X \geq 1) \to 0$ für $p \ll n^{-2/3}$
2. $P(X \geq 1) \to 1$ für $p \gg n^{-2/3}$

Beweis zu 1): Aus der Markov-Ungleichung (4) folgt mit $X \geq 0$ und c=1 bereits die Behauptung:

$$P(X \geq 1) \leq E[X] = \binom{n}{4}p^6 \leq \frac{n^4p^6}{24} \to 0$$

für $p \ll n^{-2/3}$

Beweis zu 2): Sei $p \gg n^{-2/3}$. $P(X \geq 1) \to 1$ ist äquivalent zu $P(X=0) \to 0$. Der Spezialfall der Chebyshev-Ungleichung(2) liefert

$$P(X=0) \leq \frac{Var[X]}{E[X]^2}$$

Es reicht also, die Varianz Var[X] abzuschätzen. Da X die Summe von Zufallsvariablen X_S ist, gilt

$$Var[X] = \sum_S Var[X_S] + \sum_{S \neq T} Cov(X_S, X_T)$$

mit

$$Var[X_S] = E[X_S] - \underbrace{E[X_S]^2}_{\geq 0} \leq E[X_S] = p^6$$

Es gibt $\binom{n}{4} = O(n^4)$ Mengen S, also ergibt der erste Teil der Summe einen Anteil von $O(n^4p^6)$. Nun werden Paare S,T mit $S \neq T$ betrachtet: Enthalten die von S und T induzierten Graphen keine gemeinsamen Kanten, so sind X_S, X_T s.u. $\Rightarrow Cov(X_S, X_T) = 0$, der Summand leistet keinen Beitrag. Wirksame Anteile entstehen für $|S \cap T| = 2$ (eine gemeinsame Kante) und $|S \cap T| = 3$ (drei gemeinsame Kanten).

Es gibt $\binom{n}{4}\binom{4}{2}\binom{n-4}{2} \in O(n^6)$ Paare S,T mit $|S \cap T| = 2$, für sie ist

$$Cov(X_S, X_T) = E[X_S X_T] - \underbrace{E[X_S]E[X_T]}_{\geq 0} \leq E[X_S X_T] = O(p^{11}),$$

da $S \cup T$ elf Kanten induziert. Dies führt zu einem Beitrag von $O(p^{11}n^6)$. Analog gibt es $\binom{n}{4}\binom{4}{3}(n-4) \in O(n^5)$ Paare S,T mit $|S \cap T| = 3$, für sie ist

$$Cov(X_S, X_T) = E[X_S X_T] - \underbrace{E[X_S]E[X_T]}_{\geq 0} \leq E[X_S X_T] = O(p^9),$$

da $S \cup T$ neun Kanten induziert. Dies führt zu einem Beitrag von $O(p^9n^5)$.

Insgesamt ergibt sich eine Varianz von

$$Var[X] = O(n^4p^6 + n^6p^{11} + n^5p^9) = o(n^8p^{12}) = o(E[X]^2),$$

da $p \gg n^{-2/3}$. Daher gilt P(X=0)=o(1). Somit enthält G(n,p) wahrscheinlich eine 4-Clique.

6 Zusammenfassung

Die Methode des zweiten Moments bezeichnet die Verwendung der Chebyshev-Ungleichung in probalistischen Existenzbeweisen. Sie liefert immer dann sinnvolle Abschätzungen, wenn die Varianz der untersuchten Zufallsvariablen X klein gegenüber $E[X]^2$ ist. Besonders Beweise kombinatorischer Fragestellungen lassen sich mit ihr elegant lösen, wenn man einen geeigneten Zufallsraum modelliert hat.