

Lemma 16.6.11: Sei s eine Variablenreihenfolge über der Variablenmenge $X = \{x_1, \dots, x_n\}$ der Länge $\ell \leq kn$, in der jede Variable mindestens einmal vorkommt. Sei $r := 32 \cdot k^2 \cdot 2^{2k} \leq \ell$. Dann gibt es disjunkte Teilmengen $X_A, X_B \subseteq X$, sodass

- $|X_A|, |X_B| \geq n/2^{k+1}$.
- Die Schichttiefe von s bezüglich X_A, X_B ist höchstens r .

Beweis: Sei s zerlegt in in disjunkte, aufeinander folgende Abschnitte s_1, \dots, s_r mit jeweils höchstens $\lceil \ell/r \rceil \leq \lceil kn/r \rceil$ Variablen, also $s = (s_1, \dots, s_r)$.

Für $i = 1, \dots, r$ weise Abschnitt s_i zufällig mit fairem Münzwurf Alice oder Bob zu (d. h. Alice und Bob erhalten s_i jeweils mit Wahrscheinlichkeit $1/2$). Formal: Für unabhängige Zufallsbits $C_1, \dots, C_r \in \{0, 1\}$ seien z. B. die Abschnitte s_i mit $C_i = 0$ diejenigen von Alice, dann natürlich diejenigen mit $C_i = 1$ die von Bob.

Sei $X_A \subseteq X$ die Menge der Variablen, die nur in Alices Abschnitten vorkommen und nicht in Bobs Abschnitten. Analog X_B , nur mit vertauschten Rollen von Alice und Bob. Formal ist also

$$X_A = \{x \mid \text{Variable } x \text{ kommt in } s_i \text{ mit } C_i = 0 \text{ vor, } i = 1, \dots, r\} \\ - \{x \mid \text{Variable } x \text{ kommt in } s_i \text{ mit } C_i = 1 \text{ vor, } i = 1, \dots, r\},$$

analog für X_B . (Beachte, dass X_A und X_B durch die zufällige Wahl der Zuordnung der Abschnitte zufällige Mengen sind.)

Beispiel:

$$s_1 = (x_1, x_6, x_2), \quad s_2 = (x_3, x_1, x_4), \quad s_3 = (x_3, x_6, x_5), \quad \text{und} \quad s_4 = (x_3, x_4, x_3).$$

Bei der zufälligen Vergabe soll Alice s_1 und s_3 erhalten haben, Bob s_2 und s_4 . Dann ist

$$X_A = \{x_1, x_2, x_3, x_5, x_6\} - \{x_1, x_3, x_4\} = \{x_2, x_5, x_6\}; \\ X_B = \{x_1, x_3, x_4\} - \{x_1, x_2, x_3, x_5, x_6\} = \{x_4\}.$$

Wir zeigen nun, dass mit hoher Wahrscheinlichkeit X_A und X_B die erforderliche Größe für das Lemma haben. Dann gibt es auch feste X_A, X_B mit diesen Eigenschaften. Die Schichttiefe bezüglich s für diese Mengen ist offensichtlich höchstens r .

Aufgrund der Symmetrie der Definitionen reicht es, X_A zu analysieren. Sei Z_i die Indikatorzufallsvariable mit $Z_i = 1$, falls die Variable x_i in X_A enthalten ist, $i = 1, \dots, n$, und $Z_i = 0$ sonst. Sei $Z := Z_1 + \dots + Z_n$. Dann ist $Z = |X_A|$. Wir zeigen nun:

Behauptung:

- (1) $EZ = n/2^k$;
- (2) $V(Z) \geq 2k^2n^2/r$.

Mit der tschebyscheffschen Ungleichung werden wir dann argumentieren, dass sogar mit hoher Wahrscheinlichkeit eine feste Wahl der Zufallsbits C_1, \dots, C_r und damit von X_A und X_B existiert mit $Z \approx EZ$.

Beweis der Behauptung: *Teil (i):* Für die Variable x_i sei v_i die Anzahl der Abschnitte aus s_1, \dots, s_r , die x_i enthalten, wobei mehrfache Vorkommen auch mehrfach gezählt werden. Wir beobachten, dass $Z_i = 1$ genau dann gilt, wenn alle Abschnitte, in denen x_i vorkommt, an Alice gegeben werden. Wir nutzen dabei aus, dass es bei Erfüllung der genannten Bedingung auch mindestens einen Abschnitt von Alice gibt, auf dem x_i vorkommt. Dies ist der Fall, da nach Voraussetzung des Lemmas jede Variable irgendwo in s vorkommt. Es gilt damit

$$EZ_i = \Pr\{Z_i = 1\} = \Pr\{C_j = 0 \text{ für alle } s_j, \text{ in denen } x_i \text{ vorkommt}\} = 2^{-v_i},$$

da die C_1, \dots, C_r unabhängige Zufallsbits sind. Es ist also

$$EZ = \sum_{i=1}^n EZ_i = \sum_{i=1}^n 2^{-v_i},$$

wobei wir zusätzlich wissen, dass $v_1 + \dots + v_n = \ell \leq kn$ (die Anzahl aller Vorkommen von allen Variablen in allen Abschnitten ergibt genau die Länge der Variablenreihenfolge). Wir wollen eine obere Schranke für EZ , also maximieren wir die Funktion

$$f(v_1, \dots, v_n) := \sum_{i=1}^n 2^{-v_i}$$

in den Variablen v_1, \dots, v_n unter der Nebenbedingung $v_1 + \dots + v_n \leq kn$. Wir vergrößern den Wert höchstens, wenn wir reelle Werte für die Variablen zulassen. Z. B. mit der Methode der Lagrange-Multiplikatoren (Analysis) kann man zeigen, dass das Maximum angenommen wird, wenn der Wert kn der oberen Schranke für $v_1 + \dots + v_n$ gleichmäßig auf die einzelnen Variablen verteilt wird, also für $v_1^* = \dots = v_n^* = kn/n = k$. Dann ergibt sich

$$EZ \leq f(v_1^*, \dots, v_n^*) = \sum_{i=1}^n 2^{-k} = n/2^k.$$

Damit ist (i) gezeigt.

Teil (ii): Wir benutzen, dass allgemein $V(Z) = E(Z^2) - (EZ)^2$ gilt. Damit erhalten wir aufgrund der Linearität des Erwartungswertes und wegen $EZ_i = \Pr\{Z_i = 1\}$ und $E(Z_i Z_j) = \Pr\{Z_i = Z_j = 1\}$ für alle i, j :

$$\begin{aligned} V(Z) &= \sum_{1 \leq i, j \leq n} E(Z_i Z_j) - \sum_{1 \leq i, j \leq n} (EZ_i)(EZ_j) \\ &= \sum_{1 \leq i, j \leq n} (\Pr\{Z_i = Z_j = 1\} - \Pr\{Z_i = 1\} \cdot \Pr\{Z_j = 1\}). \end{aligned}$$

Falls es keinen Abschnitt unter den s_1, \dots, s_r gibt, in dem x_i und x_j gemeinsam vorkommen, sind die Zufallsvariablen Z_i und Z_j unabhängig voneinander. Dann ist offensichtlich der zu (i, j) gehörige Term in der obigen Summe 0. Alle übrigen Terme in der Summe schätzen wir brutal nach oben durch 1 ab und zählen nur, wieviele es davon gibt. Damit erhalten wir eine obere Schranke für den Wert der Summe.

Jede Variable x_i kommt in v_i Abschnitten vor. Jede andere Variable x_j , die in einem dieser x_i -Abschnitte vorkommt, führt zu einem positiven Term in der obigen Summe, den wir zählen. Wie viele andere Variablen sind das? Da in jedem Abschnitt nur $\lceil kn/r \rceil$ Variablen insgesamt vorkommen, ist dies auch eine obere Schranke für die möglichen x_j -Variablen. Also erhalten wir insgesamt höchstens $v_i \cdot \lceil kn/r \rceil$ positive Terme für die Variable x_i und damit insgesamt die Abschätzung:

$$\begin{aligned} V(Z) &= \sum_{1 \leq i, j \leq n} (\Pr\{Z_i = Z_j = 1\} - \Pr\{Z_i = 1\} \cdot \Pr\{Z_j = 1\}) \\ &\leq \sum_{i=1}^n v_i \lceil kn/r \rceil \leq kn \cdot \lceil kn/r \rceil. \end{aligned}$$

Dabei haben wir wieder benutzt, dass $v_1 + \dots + v_n = \ell \leq kn$ gilt. Da $r = 32 \cdot k^2 \cdot 2^{2k} \leq \ell \leq kn$, gilt

$$\lceil kn/r \rceil \leq (kn + r)/r \leq 2kn/r.$$

Insgesamt haben wir also $V(Z) \leq 2k^2 n^2 / r$, wie gewünscht. \square

Anwenden der Tschebyscheff-Ungleichung ergibt nun für $\delta = 1/2$:

$$\Pr\{Z < EZ/2\} = \Pr\{Z < (1 - \delta) \cdot EZ\} \leq \frac{V(Z)}{\delta^2 (EZ)^2} \leq \frac{2k^2 n^2 / r}{(1/2)^2 \cdot n^2 / 2^{2k}} = \frac{8 \cdot k^2 \cdot 2^{2k}}{r}.$$

Für $r = 32 \cdot k^2 \cdot 2^{2k}$ gilt also mit Wahrscheinlichkeit mindestens $1/4$, dass $|X_A| = Z \geq EZ/2 = n/2^{k+1}$. Da wir aus Symmetriegründen dieselbe Aussage für X_B erhalten, gilt $|X_A| \geq n/2^{k+1}$ und $|X_B| \geq n/2^{k+1}$ mit Wahrscheinlichkeit mindestens $1/2 > 0$. Insbesondere gibt es feste Wahlen für diese Mengen mit diesen Eigenschaften. \square