Extremal Combinatorics

Linearität des Erwartungswertes

Mark Walecki-Mingers

Lehr- und Forschungsgebiet Informatik I

RWTH Aachen

Inhaltsübersicht

- Definition des Erwartungswert
- Eigenschaften des Erwartungswertes
  - Linearität
  - Schubfacheigenschaft
- 3 Anwendungen aus der Graphentheorie
  - Anzahl von Hamiltonpfade in Turnieren
  - Größe von dominierenden Mengen
  - Unabhängigkeitszahl
- eine Anwendung aus der Kombinatorik
  - Summenfreie Mengen
- eine Anwendung aus der Erfüllbarkeitslehre
  - 3-SAT

Erwartungswert

Definition: Sei X eine diskrete Zufallsvariable mit Träger $T = \{x_1, \ldots\}$. Falls $\sum_{i=1}^{\infty} |x_i| \, P(X = x_i) < \infty$ existiert, so heißt

$$E[X] := \sum_{i=1}^{\infty} x_i P(X = x_i)$$

der *Erwartungswert* der Zufallsvarible X.

Erwartungswert

Beispiel 1: Erwartungswert der Indikatorfunktion

$$\mathbb{1}_A = \begin{cases} 1 & ,\omega \in A \\ 0 & ,\omega \notin A \end{cases}$$

$$\begin{aligned} E[X] &\overset{\textit{Def.}}{=} \sum_{\omega\in\Omega} \mathbb{1}_A(\omega) Prob(\{\omega\}) \\ &= \sum_{\omega\in A} 1 \cdot Prob(\{\omega\}) + \sum_{\omega\in A^C} 0 \cdot Prob(\{\omega\}) \\ &= \sum_{\omega\in A} 1 \cdot Prob(\{\omega\}) \\ &= Prob(A) \end{aligned}$$

Erwartungswert

Beispiel 2: Erwartungswert der Binomialverteilung

$$
\begin{aligned}
E[X] &\overset{\text{Def.}}{=} \sum_{k=0}^{n} k Prob(X = k) \\
&= \sum_{k=1}^{n} k Prob(X = k) \\
&= \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n(n-1)!}{(k-1)!(n-1-(k-1))!} p p^{k-1} (1-p)^{n-1-(k-1)} \\
&= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-1-(k-1)} \\
&\overset{\text{Indexshift}}{=} np \sum_{k'=0}^{n'} \binom{n'}{k'} p^{k'} (1-p)^{n'-k'} \\
&= np \, (p + (1-p))^{n'} \\
&= np
\end{aligned}
$$

Linearität

Lemma: Seien X, Y zwei Zufallsvariablen und $a, b \in \mathbb{R}$. Falls die einzelnen Erwartungswerte existieren, dann gilt:

$$E[aX + bY] = aE[X] + bE[Y]$$

$$\begin{aligned}
E[aX + bY] &= \sum_{x \in T_x} \sum_{y \in T_y} (ax + by) P(X = x, Y = y) \\
&= \sum_{x \in T_x} \sum_{y \in T_y} ax P(X = x, Y = y) \\
&\quad + \sum_{x \in T_x} \sum_{y \in T_y} by P(X = x, Y = y) \\
&\overset{\textit{abs. konv.}}{=} \sum_{x \in T_x} ax \sum_{y \in T_y} P(X = x, Y = y) \\
&\quad + \sum_{y \in T_y} by \sum_{x \in T_x} P(X = x, Y = y) \\
&= a \sum_{x \in T_x} x P(X = x) + b \sum_{y \in T_y} y P(Y = y) \\
&= aE[X] + bE[Y]
\end{aligned}$$

Der Erwartungswert ist linear.

Schubfacheigenschaft

Lemma: Sei X eine Zufallsvariable und ihr Erwartungswert $E[X]$ existiere. Dann existiert ein $t^* \in \Omega$ mit $X(t^*) \geq E[X]$.

Angenommen die Aussage gilt nicht, d.h. $\forall t : X(t) < E[X] =: \mu$

$$\begin{aligned}
E[X] &\overset{\textit{Def.}}{=} \textstyle\sum_{t\in\Omega} P(X=t)\cdot X(t) \\
&\overset{\textit{Vor.}}{<} \textstyle\sum_{t\in\Omega} P(X=t)\cdot \mu \\
&= \mu \cdot \underbrace{\sum_{t\in\Omega} P(X=t)}_{=1} \\
&= \mu \\
&= E[X]
\end{aligned}$$

Hamiltonpfade/Turniere (1)

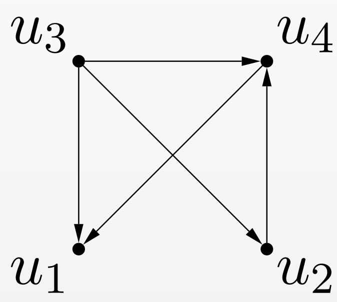
Definition: Sei $G = (V, E)$ ein Graph. Ein Pfad H heißt Hamilton-Pfad genau dann, wenn H jeden Knoten $v \in V$ genau einmal enthält.

Hamiltonpfade/Turniere (1)

Definition: Sei $G = (V, E)$ ein Graph. Ein Pfad H heißt Hamilton-Pfad genau dann, wenn H jeden Knoten $v \in V$ genau einmal enthält.

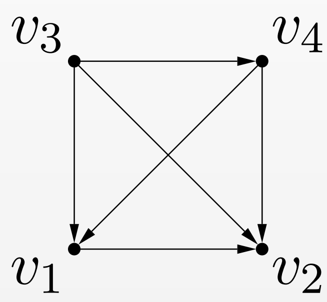
Definition: Sei ein gerichteter Graph $G = (V, E)$ mit Knotenmenge V und Kantenmenge E gegeben. Wenn für je zwei Knoten $v_1, v_2 \in V$ mit $v_1 \neq v_2$ gilt, dass entweder $(v_1, v_2) \in E$ oder $(v_2, v_1) \in E$, dann heißt G ein Turnier

Beispiele 3:



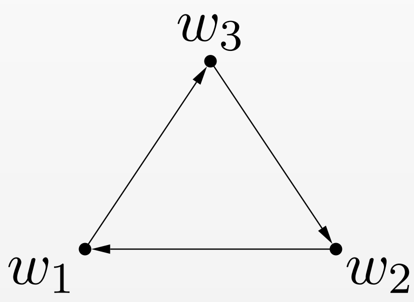
kein Turnier

(u_3, u_2, u_4, u_1)



Turnier

(v_3, v_4, v_1, v_2)

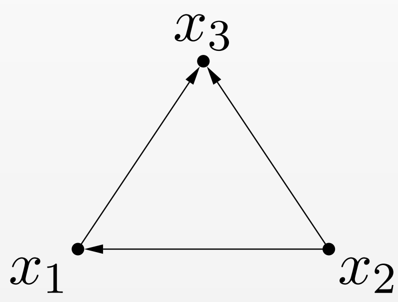


Turnier

(w_1, w_3, w_2)

(w_3, w_2, w_1)

(w_2, w_1, w_3)



Turnier

(x_2, x_1, x_3)

Anzahl von Hamiltonpfade in Turnieren (1)

Satz: [Szele] Es existiert ein Turnier T mit n Knoten, das mindestens $\frac{n!}{2^{n-1}}$ Hamiltonpfade besitzt.

Anzahl von Hamiltonpfade in Turnieren (1)

Satz: [Szele] Es existiert ein Turnier T mit n Knoten, das mindestens $\frac{n!}{2^{n-1}}$ Hamiltonpfade besitzt.

Generiere ein zufälliges Turnier $T = (V, E)$ auf der Knotenmenge $V = \{v_1, \ldots, v_n\}$.

Anzahl von Hamiltonpfade in Turnieren (1)

Satz: [Szele] Es existiert ein Turnier T mit n Knoten, das mindestens $\frac{n!}{2^{n-1}}$ Hamiltonpfade besitzt.

Generiere ein zufälliges Turnier $T = (V, E)$ auf der Knotenmenge $V = \{v_1, \ldots, v_n\}$.

Definiere zwei Zufallsvariablen

- $X(T)$ gibt die Anzahl der Hamiltonpfade von T an
- für jede Permutation $\pi \in S_n$ der Knoten sei

$$\mathbb{1}_\pi^T = \begin{cases} 1 & , \pi \textit{ ist ein Hamilton-Pfad in } T \\ 0 & , \textit{sonst} \end{cases}$$

Anzahl von Hamiltonpfade in Turnieren (1)

Satz: [Szele] Es existiert ein Turnier T mit n Knoten, das mindestens $\frac{n!}{2^{n-1}}$ Hamiltonpfade besitzt.

Generiere ein zufälliges Turnier $T = (V, E)$ auf der Knotenmenge $V = \{v_1, \ldots, v_n\}$.

Definiere zwei Zufallsvariablen

- $X(T)$ gibt die Anzahl der Hamiltonpfade von T an
- für jede Permutation $\pi \in S_n$ der Knoten sei
$$\mathbb{1}_\pi^T = \begin{cases} 1 & , (v_{\pi(j)}, v_{\pi(j+1)}) \in E \ \forall j = 1, \ldots, n-1 \\ 0 & , \textit{sonst} \end{cases}$$

Anzahl von Hamiltonpfade in Turnieren (2)

Dann läßt sich $X(T)$ darstellen als

$$X(T) = \sum_{\pi \in S_n} \mathbb{1}_\pi^T$$

Anzahl von Hamiltonpfade in Turnieren (2)

Dann läßt sich $X(T)$ darstellen als

$$X(T) = \sum_{\pi \in S_n} \mathbb{1}_\pi^T$$

Mit der Linearität folgt

$$E[X] = E[\sum_{\pi \in S_n} \mathbb{1}_\pi^T] = \sum_{\pi \in S_n} E[\mathbb{1}_\pi^T]$$

Anzahl von Hamiltonpfade in Turnieren (2)

Dann läßt sich $X(T)$ darstellen als

$$X(T) = \sum_{\pi \in S_n} \mathbb{1}_\pi^T$$

Mit der Linearität folgt

$$E[X] = E[\sum_{\pi \in S_n} \mathbb{1}_\pi^T] = \sum_{\pi \in S_n} E[\mathbb{1}_\pi^T]$$

Bleibt der Erwartungswert von $\mathbb{1}_\pi^T$ zu berechnen.

Anzahl von Hamiltonpfade in Turnieren (3)

Aus Beispiel 1 folgt

$$E[\mathbb{1}_{\pi}^{T}] = Prob(\text{'}\pi \textit{ ist Hamilton-Pfad'})$$

Anzahl von Hamiltonpfade in Turnieren (3)

Aus Beispiel 1 folgt

$$E[\mathbb{1}_\pi^T] = Prob(\text{'}\pi \textit{ ist Hamilton-Pfad'})$$

Wie groß ist die Wahrscheinlichkeit, dass $\pi = (v_{\pi(1)}, \ldots, v_{\pi(n)})$ ein Hamiltonpfad ist?

Anzahl von Hamiltonpfade in Turnieren (3)

Aus Beispiel 1 folgt

$$E[\mathbb{1}_\pi^T] = Prob(\text{'}\pi \textit{ ist Hamilton-Pfad'})$$

Wie groß ist die Wahrscheinlichkeit, dass $\pi = (v_{\pi(1)}, \ldots, v_{\pi(n)})$ ein Hamiltonpfad ist?

Alle $n-1$ Kanten müssen bei der Generierung gleichen Ausgang haben (links nach rechts).

Anzahl von Hamiltonpfade in Turnieren (3)

Aus Beispiel 1 folgt

$$E[\mathbb{1}_\pi^T] = Prob(\text{'}\pi \textit{ ist Hamilton-Pfad'}) = \left(\frac{1}{2}\right)^{n-1}$$

Wie groß ist die Wahrscheinlichkeit, dass $\pi = (v_{\pi(1)}, \ldots, v_{\pi(n)})$ ein Hamiltonpfad ist?

Alle $n-1$ Kanten müssen bei der Generierung gleichen Ausgang haben (links nach rechts).

Anzahl von Hamiltonpfade in Turnieren (4)

Zusammengefasst

$$\begin{aligned} E[X] &= \textstyle\sum_{\pi \in S_n} E[\mathbb{1}_\pi^T] \\ &= \textstyle\sum_{\pi \in S_n} \frac{1}{2^{n-1}} \\ &= \frac{n!}{2^{n-1}} \end{aligned}$$

Anzahl von Hamiltonpfade in Turnieren (4)

Zusammengefasst

$$\begin{aligned} E[X] &= \sum_{\pi \in S_n} E[\mathbb{1}_\pi^T] \\ &= \sum_{\pi \in S_n} \frac{1}{2^{n-1}} \\ &= \frac{n!}{2^{n-1}} \end{aligned}$$

Mit der Schubfacheigenschaft folgt die Existenz eines Turniers T^* mit $X(T^*) \geq \frac{n!}{2^{n-1}}$.

Anzahl von Hamiltonpfade in Turnieren (5)

Beispiele 4:

$Knoten$	n	3	4	5	6	10	20	50
$\#Pfade$	$\frac{n!}{2^{n-1}}$	1.5	3	7.5	22.5	7087.5	$4.64 \cdot 10^{12}$	$5.40 \cdot 10^{49}$
$\#Turniere$	$2^{\frac{n(n-1)}{2}}$	8	64	1024	32768	$3.5 \cdot 10^{13}$	$1.56 \cdot 10^{57}$	$5.77 \cdot 10^{369}$

Bemerkung:

Schwieriger ist die maximale Anzahl von Hamiltonpfade zu bestimmen $(c \cdot \sqrt{n^3}\frac{n!}{2^{n-1}})$.

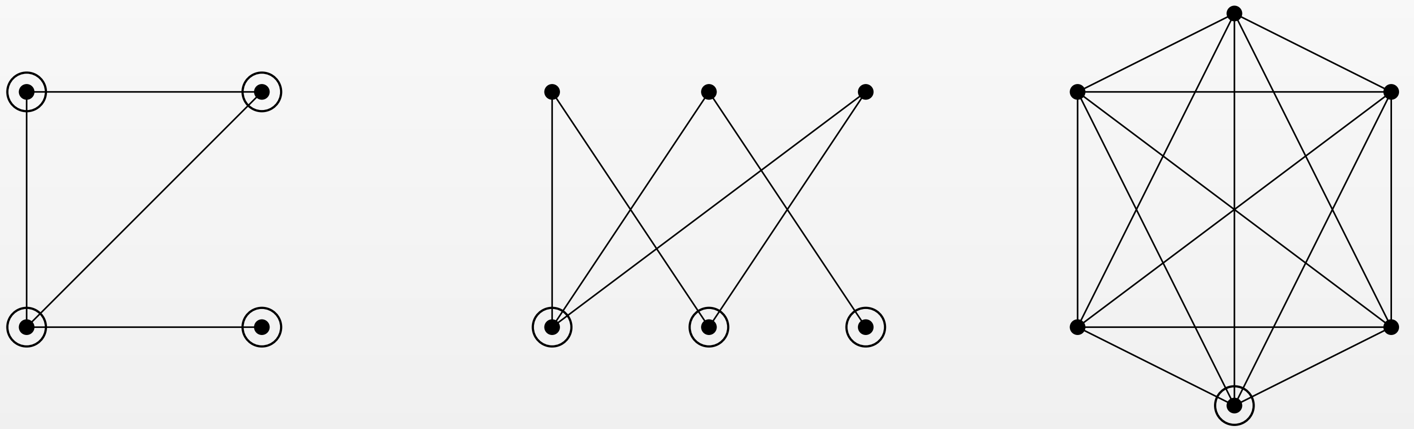
Dominierende Mengen (1)

Definition: Sei $G = (V, E)$ ein Graph. Weiterhin sei $S \subseteq V$ eine Teilmenge der Knotenmenge. Dann heißt S eine dominierende Menge, wenn für alle $v \in V$ eine der beiden Bedingungen erfüllt ist.

$$
\begin{aligned}
(i) &\quad v \in S \\
(ii) &\quad \exists w \in S \; mit \; \{v, w\} \in E
\end{aligned}
$$

Dominierende Mengen (2)

Beispiele 5:



Dominierende Mengen (2)

Satz: [Alon] Sei $G = (V, E)$ ein Graph mit $n = |V|$ Knoten und einem Minimalgrad d_{min}. Dann existiert eine dominierende Menge S mit höchstens $|S| = n\frac{1+\ln(d_{min}+1)}{d_{min}+1}$ Knoten.

Dominierende Mengen (2)

Satz: [Alon] Sei $G = (V, E)$ ein Graph mit $n = |V|$ Knoten und einem Minimalgrad d_{min}. Dann existiert eine dominierende Menge S mit höchstens $|S| = n\frac{1+\ln(d_{min}+1)}{d_{min}+1}$ Knoten.

Beweis: Zähle jeden Knoten $v \in V$ mit einer Wahrscheinlichkeit von

$$P(v \in S') =: p = \frac{ln(d_{min} + 1)}{d_{min} + 1}$$

zu einer Menge S'.

Dominierende Mengen (2)

Satz: [Alon] Sei $G = (V, E)$ ein Graph mit $n = |V|$ Knoten und einem Minimalgrad d_{min}. Dann existiert eine dominierende Menge S mit höchstens $|S| = n\frac{1+\ln(d_{min}+1)}{d_{min}+1}$ Knoten.

Beweis: Zähle jeden Knoten $v \in V$ mit einer Wahrscheinlichkeit von

$$P(v \in S') =: p = \frac{ln(d_{min} + 1)}{d_{min} + 1}$$

zu einer Menge S'.
Falls diese nicht dominierend ist, wähle $T \subset V - S'$, so dass

$$\{t, s'\} \notin E \; \forall t \in T, s' \in S'$$

Dominierende Mengen (3)

Dann ist

$$S := T \,\dot\cup\, S'$$

eine dominierende Menge als disjunkte Vereinigung.
Es ergibt sich die erwartete Größe von S als

$$E[|S|] = E[\left|S'\dot\cup T\right|] = E[\left|S'\right| + |T|] = E[\left|S'\right|] + E[|T|]$$

Für S' ist

$$E[\left|S'\right|] = E[\sum_{v\in V} \mathbb{1}_{S'}(v)] = Prob(S') = np = n\frac{\ln(d_{min}+1)}{d_{min}+1}$$

Dominierende Mengen (4)

Da für $t \in T$ gilt:

1. $t \notin S'$
2. $\nexists s' \in S'$ mit $\{t, s'\} \in E$

$$\Rightarrow Prob(t \in T) \leq (1-p)^{d_{min}+1}$$

Konstuiere die Zufallsvariable

$$\mathbb{1}_t^T = \begin{cases} 1 & , t \in T \\ 0 & , t \notin T \end{cases}$$

Dominierende Mengen (5)

Dann ergibt sich

$$\begin{aligned}
E[|T|] &= E[\sum_{t \in V} \mathbb{1}_t^T] \\
&= \sum_{t \in V} E[\mathbb{1}_t^T] \\
&= \sum_{t \in V} Prob(t \in T) \\
&\leq n(1-p)^{d_{min}+1} \\
&\leq ne^{-p(d_{min}+1)}
\end{aligned}$$

mit $(1-p) \leq e^{-p} \; \forall p \in [0,1]$

$$E[|T|] = ne^{-p(d_{min}+1)}$$

Dominierende Mengen (6)

Zusammengefasst

$$
\begin{aligned}
E[|S|] &= E[|S'|] + E[|T|] \\
&\leq np + ne^{-p(d_{min}+1)} \\
&= n\frac{\ln(d_{min}+1)}{d_{min}+1} + ne^{-\frac{\ln(d_{min}+1)}{d_{min}+1}(d_{min}+1)} \\
&= n\frac{\ln(d_{min}+1)+1}{d_{min}+1}
\end{aligned}
$$

Aus der Schubfacheigenschaft folgt die Existenz einer dominierenden Menge S^* mit

$$|S^*| \leq n\left(\frac{1+\ln(d_{min}+1)}{d_{min}+1}\right)$$

Dominierende Mengen (7)

Bemerkungen

- Wahl von p: die Funktion $f(p) := np + ne^{-p(d_{min}+1)}$ nimmt bei $p = \frac{\ln(d_{min}+1)}{d_{min}+1}$ das einzige Minimum an.

Bemerkungen

- Wahl von p: die Funktion $f(p) := np + ne^{-p(d_{min}+1)}$ nimmt bei $p = \frac{\ln(d_{min}+1)}{d_{min}+1}$ das einzige Minimum an.
- Fall $d_{min} = 0$, (isolierte Knoten): $|S| \leq n\frac{1+\overbrace{\ln(0+1)}^{=0}}{0+1} = n$

Bemerkungen

- Wahl von p: die Funktion $f(p) := np + ne^{-p(d_{min}+1)}$ nimmt bei $p = \frac{\ln(d_{min}+1)}{d_{min}+1}$ das einzige Minimum an.

- Fall $d_{min} = 0$, (isolierte Knoten): $|S| \leq n\frac{1+\overbrace{\ln(0+1)}^{=0}}{0+1} = n$

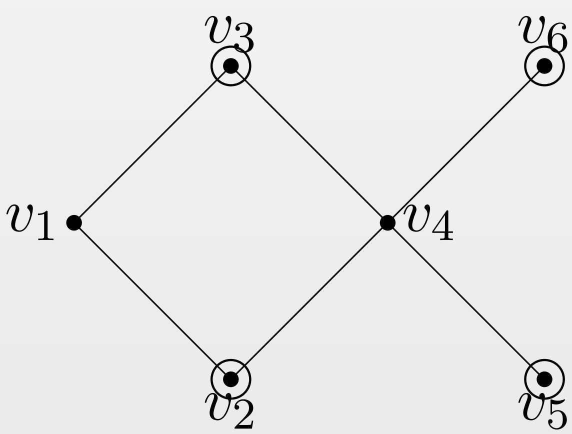
- Fall $d_{min} = 1$, (Hanteln): $|S| \leq n\frac{1+\ln(1+1)}{1+1} = n \approx 0,846n$

Unabhängigkeitszahl (1)

Definition: Sei ein Graph $G = (V, E)$ gegeben. Dann ist die Unabhängigkeitszahl $\alpha(G)$ des Graphen definiert als

$$\alpha(G) := \max_{S \subseteq V} |\{\forall s_1, s_2 \in S : \{s_1, s_2\} \notin E\}|$$

Beispiel 6:



$$d_1 = d_2 = d_3 = 2$$
$$d_4 = 4$$
$$d_5 = d_6 = 1$$
$$\alpha(G) = 4$$

Unabhängigkeitszahl (2)

Satz: [Alon,Spencer 1992] Sei $G = (V, E)$ ein Graph mit der Knotenmenge $V = \{v_1, \ldots, v_n\}$ und $d_1, \ldots, d_n$ die assoziierten Knotengrade. Dann gilt

$$\alpha(G) \geq \sum_{i=1}^{n} \frac{1}{d_i + 1}$$

Unabhängigkeitszahl (2)

Satz: [Alon,Spencer 1992] Sei $G = (V, E)$ ein Graph mit der Knotenmenge $V = \{v_1, \ldots, v_n\}$ und $d_1, \ldots, d_n$ die assoziierten Knotengrade. Dann gilt

$$\alpha(G) \geq \sum_{i=1}^{n} \frac{1}{d_i + 1}$$

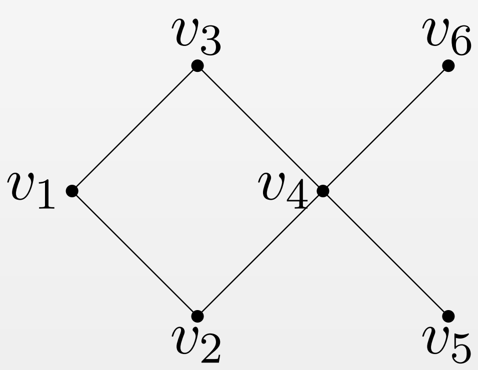
Beweis: Wähle gleichverteilt eine Permutation π der Knotenmenge. π induziert eine Ordnung der Knoten.

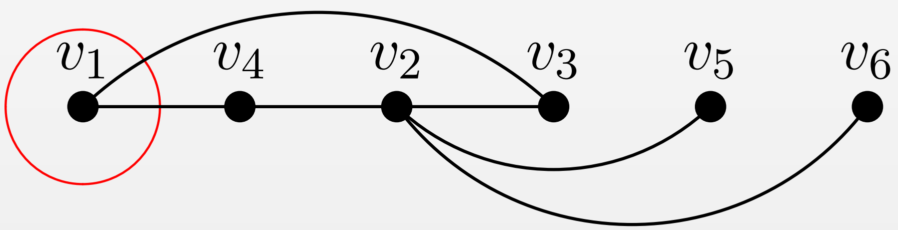
$$v_{\pi(1)} < v_{\pi(2)} < \cdots < v_{\pi(n)}$$

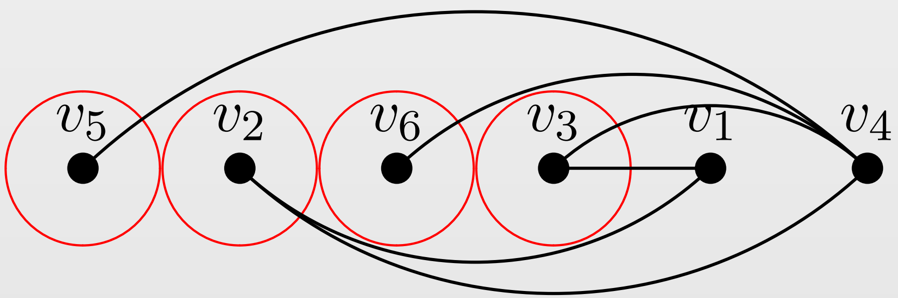
Unabhängigkeitszahl (3)

Definiere A_i als das Ereignis, dass alle d_i Nachbarn $v_{j_1}, \ldots, v_{j_{d_i}}$ in dieser Ordnung größer sind als v_i.

Beispiel 7:







Unabhängigkeitszahl (4)

Es gibt $\binom{n}{d_i+1}$ Möglichkeiten v_i und seine Nachbarn aus n Knoten zu wählen. Wenn A_i gelten soll:

1. v_i muss als erstes stehen
2. Reihenfolge der Nachbarn untereinander egal $\Rightarrow d_i!$ Anordnungen
3. Reihenfolge der Nicht-Nachbarn von v_i ebenfalls egal $\Rightarrow (n - d_i - 1)!$ Anordnungen

Die Wahrscheinlichkeit von A_i:

$$
\begin{aligned}
Prob(A_i) &= \binom{n}{d_i+1} \frac{d_i!(n-d_i-1)!}{n!} \\
&= \frac{n!d_i!(n-d_i-1)!}{n!(n-d_i-1)!(d_i+1)!} \\
&= \frac{1}{d_i+1}
\end{aligned}
$$

Unabhängigkeitszahl (5)

Definiere für alle $i = 1, \ldots, n$ die Zufallsvariable

$$\mathbb{1}^{\pi}_{A_i} = \begin{cases} 1 & , A_i \textit{ gilt} \\ 0 & , \textit{sonst} \end{cases}$$

Sei U die Vereinigung derjenigen v_i für die A_i gilt. Dann gilt

$$E[|U|] = E[\sum_{i=1}^{n} \mathbb{1}^{\pi}_{A_i}] = \sum_{i=1}^{n} E[\mathbb{1}^{\pi}_{A_i}] = \sum_{i=1}^{n} Prob(A_i) = \sum_{i=1}^{n} \frac{1}{d_i + 1}$$

Bleibt zu zeigen, dass U unabhängige Menge ist!

Unabhängigkeitszahl (6)

U ist eine unabhängige Menge, denn sei $\{v_i, v_j\} \in E$ eine Kante in G, dann gilt

$$\begin{array}{ccc} v_i < v_j & \Rightarrow & v_j \notin U \\ & oder & \\ v_j < v_i & \Rightarrow & v_i \notin U \end{array}$$

Mit dem Schubfachprinzip folgt, dass eine Permutation π^* existiert mit $|U| \geq \sum_{i=1}^{n} \frac{1}{d_i+1}$.
Für Beispiel 7:

$$\alpha(G) = 4 \geq 3 > \frac{3}{2+1} + \frac{1}{4+1} + \frac{2}{1+1} = 2\frac{1}{5}$$

Summenfreie Mengen (1)

Definition: Eine Menge M heißt summenfrei genau dann, wenn für je zwei Elemente $a, b \in M$ gilt

$$a \neq b \Rightarrow a + b \notin M$$

Summenfreie Mengen (1)

Definition: Eine Menge M heißt summenfrei genau dann, wenn für je zwei Elemente $a, b \in M$ gilt

$$a \neq b \Rightarrow a + b \notin M$$

Satz: Sei $A \subseteq \mathbb{Z} - \{0\}$ mit $|A| = n$. Dann existiert eine Teilmenge $B \subseteq A$ mit

$$|B| > \frac{n}{3}$$

Summenfreie Mengen (2)

Beweis:

- Wähle eine Primzahl $p = 3k + 2 > 2 \max_{a \in A} |a|$.

Summenfreie Mengen (2)

Beweis:

- Wähle eine Primzahl $p = 3k + 2 > 2 \max_{a \in A} |a|$.
- Kontruiere Menge $S = \{k+1, \ldots, 2k+1\} = \{k+1, \ldots, k+(k+1)\} \subset \mathbb{Z}_p$.
  $s_1, s_2 \in S$,

$$\begin{aligned} s_1 + s_2 &= (k + t_1) + (k + t_2) \\ &= 2k + t_1 + t_2 \\ \textit{mit } t_1 + t_2 &\in \{2, \ldots, 2k+2\} \\ &= \{2, \ldots, k+1\} \,\dot\cup\, \{k+2, \ldots, 2k+2\} \\ \Rightarrow s_1 + s_2 &\in \{2k+2, \ldots, 3k+1\} \,\dot\cup\, \{3k+2, \ldots, 4k+2\} \\ &= \{2k+2, \ldots, 3k+1\}_{\mathbb{Z}_p} \,\dot\cup\, \{0, \ldots, k\}_{\mathbb{Z}_p} \end{aligned}$$

Summenfreie Mengen (3)

- Wähle $t \in \mathbb{Z}_p^*$ gleichverteilt ($Prob(t) = \frac{1}{p-1}$)

Summenfreie Mengen (3)

- Wähle $t \in \mathbb{Z}_p^*$ gleichverteilt ($Prob(t) = \frac{1}{p-1}$)
- Definiere $f_t : A \rightarrow \mathbb{Z}_p$ mit $f_t(a) = at \ (mod\ p)$ sowie

$$\mathbb{1}_{a,j} = \begin{cases} 1 & , f_t(a) = j \ (mod\ p) \\ 0 & , \textit{sonst} \end{cases}$$

Summenfreie Mengen (3)

- Wähle $t \in \mathbb{Z}_p^*$ gleichverteilt ($Prob(t) = \frac{1}{p-1}$)
- Definiere $f_t : A \rightarrow \mathbb{Z}_p$ mit $f_t(a) = at \ (mod\ p)$ sowie

$$\mathbb{1}_{a,j} = \begin{cases} 1 & , f_t(a) = j \ (mod\ p) \\ 0 & , \textit{sonst} \end{cases}$$

Sei X die Anzahl der auf S abgebildeten Elemente von A.

$$X := \sum_{a \in A} \sum_{j \in S} \mathbb{1}_{a,j}$$

Da $\mathbb{Z}_p$ ein Körper im algebraischen Sinne ist, gibt es genau ein t aus $\mathbb{Z}_p - \{0\}$ mit $a \cdot t \equiv_p j$.

Summenfreie Mengen (4)

Mit

$$E[\mathbb{1}_{a,j}] = Prob(\mathbb{1}_{a,j} = 1) = Prob(t) = \frac{1}{p-1}$$

und der Linearität folgt

$$\begin{aligned} E[X] &= E[\textstyle\sum_{a\in A}\sum_{j\in S}\mathbb{1}_{a,j}] \\ &= \textstyle\sum_{a\in A}\sum_{j\in S} E[\mathbb{1}_{a,j}] \\ &= \textstyle\sum_{a\in A}\sum_{j\in S}\frac{1}{p-1} \\ &= \textstyle\sum_{a\in A}\frac{|S|}{p-1} \\ &= \textstyle\frac{n|S|}{p-1} \\ &= \textstyle\frac{nk}{3k-1} \\ &> \textstyle\frac{nk}{3k} = \frac{n}{3} \end{aligned}$$

Summenfreie Mengen (5)

Mit dem Schubfachprinzip folgt, dass ein $t^* \in \{1, \ldots, p-1\}$ existiert mit $|B| = |\{a \in A \mid a \cdot t^* = j \ (mod\ p)$ *mit* $j \in S\}| > \frac{n}{3}$.

Bemerkung:

- Konstruktion im wesentlichen unabhängig von gegebener Menge A
- Existenz von $p = 3k + 2$?
- mit $p = 3k + 1$ und $S = \{k, \ldots, 2k\}$ erhält man $|B| \geq \frac{n}{3}$

Erfüllbarkeit (1)

Definition: Eine Formel F ist k-erfüllbar, wenn mindestens eine Belegung b existiert, so dass für beliebige k Klauseln, $C_{i_1}, \ldots, C_{i_k}$, diese erfüllt sind.

Erfüllbarkeit (1)

Definition: Eine Formel F ist k-erfüllbar, wenn mindestens eine Belegung b existiert, so dass für beliebige k Klauseln, $C_{i_1}, \ldots, C_{i_k}$, diese erfüllt sind.

Satz: [Yannakakis 1994] Sei F eine Formel die 3-erfüllbar ist. Dann existiert eine Belegung b mit der mindestens $\frac{2}{3}$ aller Klauseln gleichzeitig erfüllt werden können.

Erfüllbarkeit (2)

Gegeben ist eine 3-erfüllbare Formel $F(x_1, \ldots, x_n)$. Dann konstruiere eine zufällige Belegung $b = (b_1, \ldots, b_n)$ wobei die b_i mit einer Wahrscheinlichkeit

$$p_{i,1} := Prob(b_i = 1) = \begin{cases} \frac{2}{3} & \text{, } \textit{F enthält eine Klausel } x_i \\ \frac{1}{3} & \text{, } \textit{F enthält eine Klausel } \overline{x_i} \\ \frac{1}{2} & \text{, } \textit{sonst} \end{cases}$$

und folglich

$$p_{i,0} := Prob(b_i = 0) = \begin{cases} 1 - \frac{2}{3} = \frac{1}{3} & \text{, } \textit{F enthält eine Klausel } x_i \\ 1 - \frac{1}{3} = \frac{2}{3} & \text{, } \textit{F enthält eine Klausel } \overline{x_i} \\ \frac{1}{2} & \text{, } \textit{sonst} \end{cases}$$

Beachte: wohldefiniert

Erfüllbarkeit (3)

Jede Klausel wird mit Wahrscheinlichkeit von mindestens $\frac{2}{3}$ erfüllt.

- 1. Fall (Einheitsklauseln)
  enspricht der Definition

Erfüllbarkeit (3)

Jede Klausel wird mit Wahrscheinlichkeit von mindestens $\frac{2}{3}$ erfüllt.

- 1. Fall (Einheitsklauseln)
  enspricht der Definition
- 2. Fall (Klauseln der Länge $n \geq 3$)
  Beachte jedes Literal wird mit einer W'keit von höchstens $\frac{2}{3}$ nicht erfüllt.

$$\frac{2}{3} = \frac{18}{27} < \frac{19}{27} = 1 - \left(\frac{2}{3}\right)^3 \leq 1 - \left(\frac{2}{3}\right)^n, n \geq 3$$

Erfüllbarkeit (3)

Jede Klausel wird mit Wahrscheinlichkeit von mindestens $\frac{2}{3}$ erfüllt.

- 1. Fall (Einheitsklauseln)
  enspricht der Definition
- 2. Fall (Klauseln der Länge $n \geq 3$)
  Beachte jedes Literal wird mit einer W'keit von höchstens $\frac{2}{3}$ nicht erfüllt.

$$\frac{2}{3} = \frac{18}{27} < \frac{19}{27} = 1 - \left(\frac{2}{3}\right)^3 \leq 1 - \left(\frac{2}{3}\right)^n, n \geq 3$$

- 3. Fall (Klauseln der Länge 2):
  ObdA

$$C = (x_1 \vee x_2)$$

Erfüllbarkeit (4)

Unterscheide wiederum 2 Fälle:

- mindestens ein Literal wird mit einer Wahrscheinlichkeit von $\frac{1}{2}$ erfüllt, das heißt weder das Literal, noch sein Negiertes treten in der Formel als Einheitsklauseln auf. Daraus folgt C wird mit einer Wahrscheinlichkeit von $1 - Prob(b_1 = 0) \cdot Prob(b_2 = 0) \geq 1 - \frac{1}{2} \cdot \frac{2}{3} = \frac{2}{3}$ erfüllt.

Erfüllbarkeit (4)

Unterscheide wiederum 2 Fälle:

- mindestens ein Literal wird mit einer Wahrscheinlichkeit von $\frac{1}{2}$ erfüllt, das heißt weder das Literal, noch sein Negiertes treten in der Formel als Einheitsklauseln auf. Daraus folgt C wird mit einer Wahrscheinlichkeit von $1 - Prob(b_1 = 0) \cdot Prob(b_2 = 0) \geq 1 - \frac{1}{2} \cdot \frac{2}{3} = \frac{2}{3}$ erfüllt.
- beide Literale werden nur mit einer Wahrscheinlichkeit von $\frac{1}{3}$ erfüllt. D.h. die Formel F muss die Klauseln $C' = \overline{x_1}$ und $C'' = \overline{x_2}$ enthalten. Aber die Klauseln C, C', C'' können nicht gleichzeitig erfüllt werden, was ein Widerspruch zur 3-Erfüllbarkeit darstellt.

Erfüllbarkeit (5)

Definiere die Zufallsvariable $\mathbb{1}_i$ als Indikator für: die i-te Klausel wird von der Belegung b erfüllt.
Daraus ergibt sich die Anzahl der erfüllten Klauseln

$$X := \sum_{i=1}^{m} \mathbb{1}_i$$

$\Rightarrow$

$$E[X] = E[\sum_{i=1}^{m} \mathbb{1}_i] = \sum_{i=1}^{m} E[\mathbb{1}_i] = \sum_{i=1}^{m} Prob(\mathbb{1}_i = 1) \geq m \cdot \frac{2}{3}$$

D.h. es existiert eine Belegung b^*, so dass mindestens $\frac{2}{3}$ der Klauseln einer Formel erfüllt sind.

Erfüllbarkeit (6)

Bemerkung:

- Für die 2-Erfüllbarkeit ergibt sich die Konstante $\frac{2}{1+\sqrt{5}} \approx 0.618$
- Die Konstante konvergiert für große k gegen $\frac{3}{4}$

Zusammenfassung

- Wahrscheinlichkeitstheorie (insb. Erwartungswert) mächtiges Werkzeug
- Beweise kurz und einfach
- Nur Existenzbeweise
- Modellierung oft nicht einfach