Permanents and the power of entropy

Stefan Haan

2023-01-09

Definition (Entropie)

Sei X eine Zufallsvariable mit Support $\{x_1, x_2, \ldots, x_n\}$ und Wahrscheinlichkeiten $p_i = \text{Prob}(X = x_i)$. Die Entropie von X ist definiert als

$$H(X) := \sum_i p_i \log_2 \frac{1}{p_i} = -\sum_i p_i \log_2 p_i.$$

Definition (Entropie)

Sei X eine Zufallsvariable mit Support $\{x_1, x_2, \ldots, x_n\}$ und Wahrscheinlichkeiten $p_i = \text{Prob}(X = x_i)$. Die Entropie von X ist definiert als

$$H(X) := \sum_i p_i \log_2 \frac{1}{p_i} = -\sum_i p_i \log_2 p_i.$$

- Erwartungswert der "Überraschung", dass X den Wert x_i annimmt $h(x_i) := \log_2 \frac{1}{p_i}$.
- Maß für die Unsicherheit welchen Wert X annimmt.

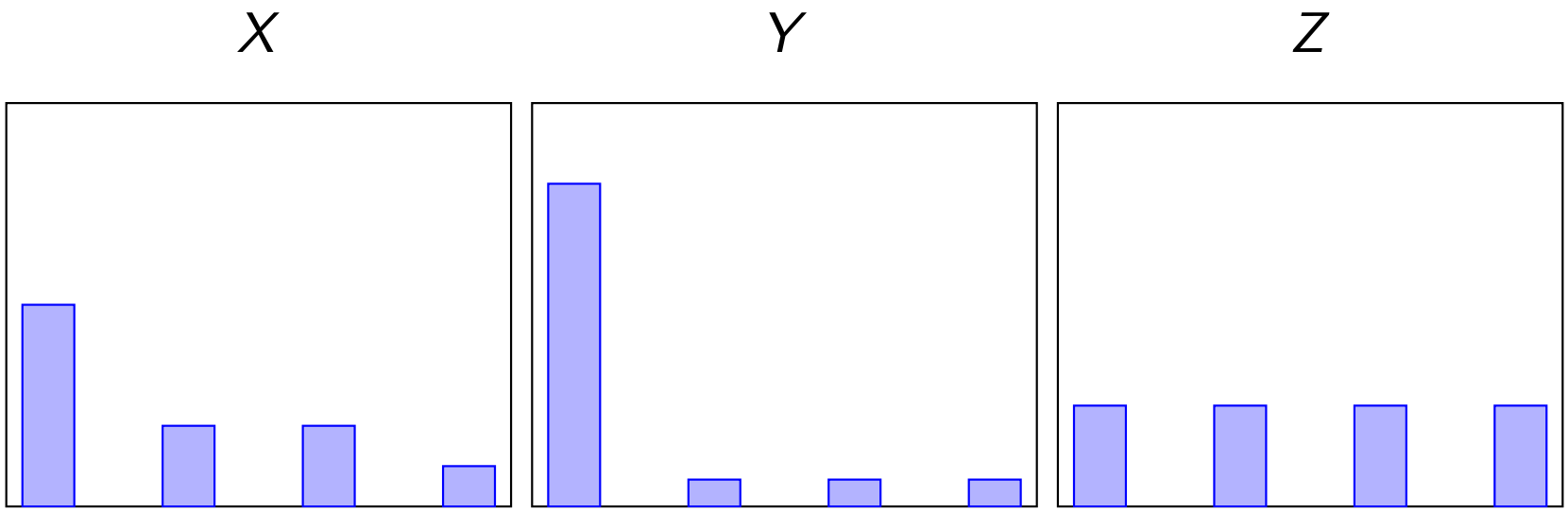
Definition (Entropie)

Sei X eine Zufallsvariable mit Support $\{x_1, x_2, \ldots, x_n\}$ und Wahrscheinlichkeiten $p_i = \text{Prob}(X = x_i)$. Die Entropie von X ist definiert als

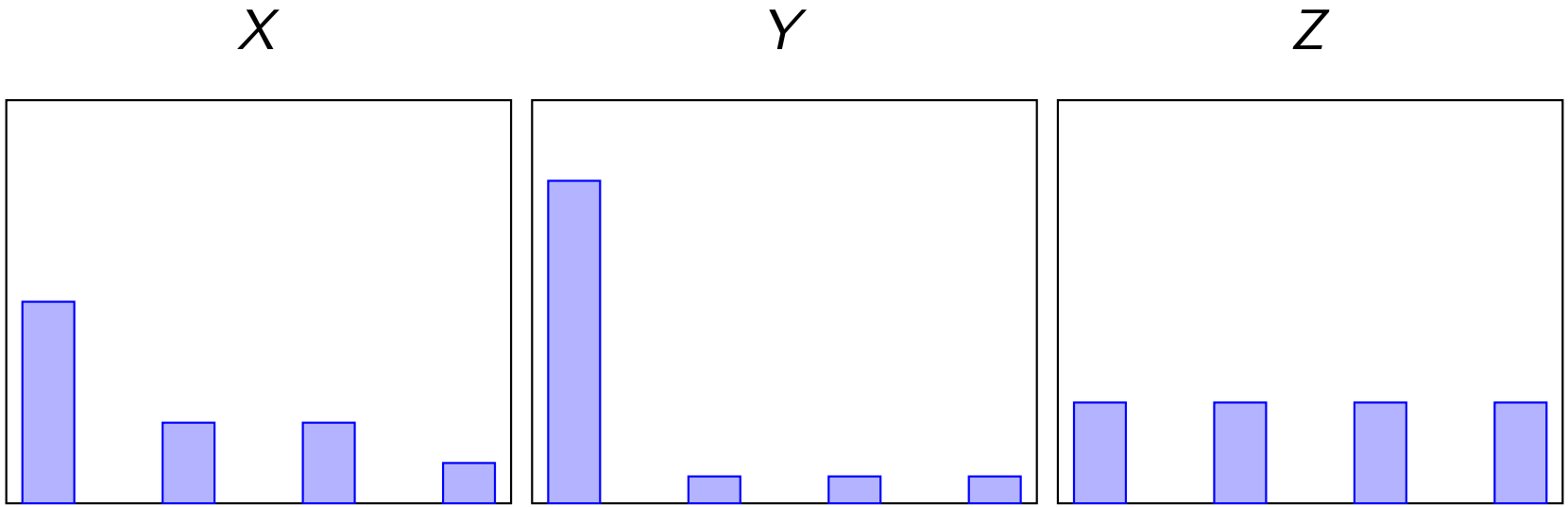
$$H(X) := \sum_i p_i \log_2 \frac{1}{p_i} = -\sum_i p_i \log_2 p_i.$$

- Erwartungswert der "Überraschung", dass X den Wert x_i annimmt $h(x_i) := \log_2 \frac{1}{p_i}$.
- Maß für die Unsicherheit welchen Wert X annimmt.
- Nur von der Verteilung abhängig.

Bei welcher Zufallsvariable gibt es am meisten Unsicherheit über den Ausgang?



Bei welcher Zufallsvariable gibt es am meisten Unsicherheit über den Ausgang?



- Antwort: Z.
- $H(X) \approx 1.76$, $H(Y) \approx 1.04$, $H(Z) = 2$.
- Vermutung: Die Entropie ist maximal bei Gleichverteilung.

Lemma (Entropie maximal bei Gleichverteilung)

Es gilt $H(X) \leq \log_2 |\operatorname{supp} X|$. *Die Schranke wird genau dann angenommen, wenn* X *gleichverteilt ist.*

Beweis "$\Longleftarrow$".

Jensensche Ungleichung für die konkave Funktion $\log_2$ liefert

$$H(X) = \sum_i p_i \log_2 \left(\frac{1}{p_i}\right) \leq \log_2 \left(\sum_i p_i \frac{1}{p_i}\right) = \log_2 n.$$

Bei Gleichverteilung ist $p_1 = p_2 = \cdots = \frac{1}{n}$ und es gilt Gleichheit. □

Eine andere Sichtweise...

- Jemand versteckt vor uns den Ausgang der Zufallsvariable X.
- Sie beantwortet nur Ja/Nein Fragen über das Ergebnis.

Eine andere Sichtweise...

- Jemand versteckt vor uns den Ausgang der Zufallsvariable X.
- Sie beantwortet nur Ja/Nein Fragen über das Ergebnis.
- Wieviele Ja/Nein Fragen müssen wir durchschnittlich stellen, um das Ergebnis mit Sicherheit zu kennen?
- Interpretation: *Informationsgehalt* in Bits.

$$\text{Informationsgehalt von } X \longleftrightarrow \text{Unsicherheit über } X$$

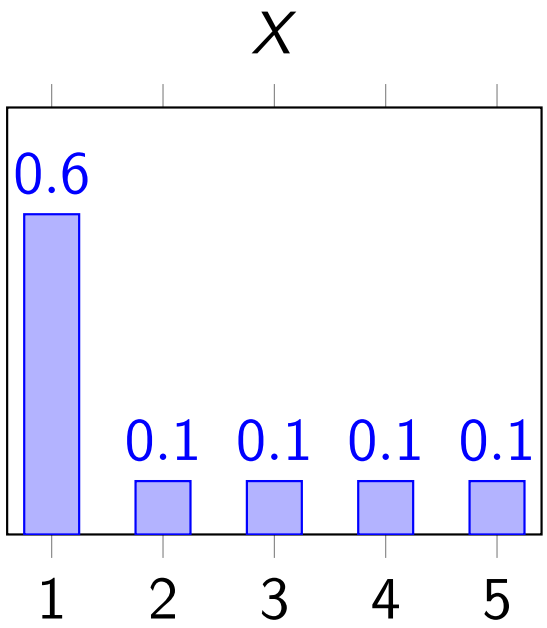
Eine andere Sichtweise...

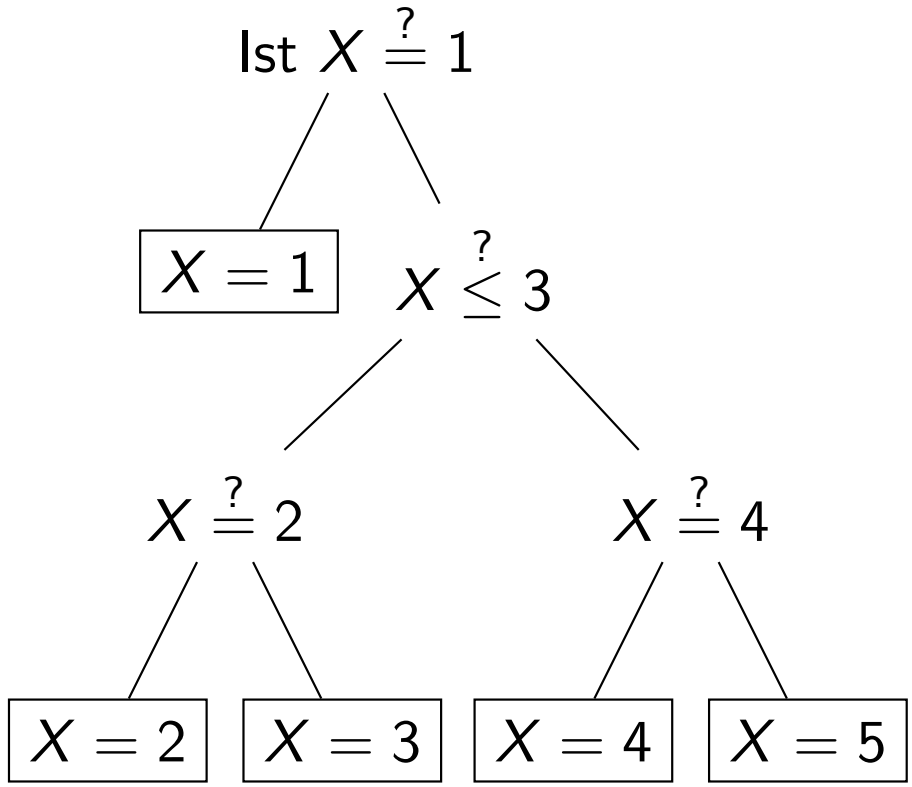
- Jemand versteckt vor uns den Ausgang der Zufallsvariable X.
- Sie beantwortet nur Ja/Nein Fragen über das Ergebnis.
- Wieviele Ja/Nein Fragen müssen wir durchschnittlich stellen, um das Ergebnis mit Sicherheit zu kennen?
- Interpretation: *Informationsgehalt* in Bits.

$$\text{Informationsgehalt von } X \longleftrightarrow \text{Unsicherheit über } X$$

Zusammenhang

Die Entropie ist eine untere Schranke für die durchschnittliche Anzahl Fragen, die wir stellen müssen!





$$1.77 \approx H(X) \leq \mathbb{E}[\#\text{Fragen}] = \frac{6}{10} \cdot 1 + \frac{4}{10} \cdot 3 = 1.8$$

Definition (Bedingte Entropie)

Seien X, Y Zufallsvariablen die die Werte $\{x_1, x_2, \ldots, x_m\}$ bzw. $\{y_1, y_2, \ldots, y_n\}$ annehmen. Die Entropie von Y, falls $X = x_i$ bekannt ist, ist definiert als

$$H(Y \mid x_i) := -\sum_j p(y_j \mid x_i) \log_2 p(y_j \mid x_i).$$

Definition (Bedingte Entropie)

Seien X, Y Zufallsvariablen die die Werte $\{x_1, x_2, \ldots, x_m\}$ bzw. $\{y_1, y_2, \ldots, y_n\}$ annehmen. Die Entropie von Y, falls $X = x_i$ bekannt ist, ist definiert als

$$H(Y \mid x_i) := -\sum_j p(y_j \mid x_i) \log_2 p(y_j \mid x_i).$$

Über alle Werte von X erhalten wir die *bedingte Entropie von Y unter X*:

$$H(Y \mid X) := \sum_i p(x_i)\, H(Y \mid x_i).$$

Definition (Bedingte Entropie)

Seien X, Y Zufallsvariablen die die Werte $\{x_1, x_2, \ldots, x_m\}$ bzw. $\{y_1, y_2, \ldots, y_n\}$ annehmen. Die Entropie von Y, falls $X = x_i$ bekannt ist, ist definiert als

$$H(Y \mid x_i) := -\sum_j p(y_j \mid x_i) \log_2 p(y_j \mid x_i).$$

Über alle Werte von X erhalten wir die *bedingte Entropie von Y unter X*:

$$H(Y \mid X) := \sum_i p(x_i)\, H(Y \mid x_i).$$

- $H(Y \mid X) = 0$, wenn der Ausgang von Y keine Überraschung mehr ist, sobald X bekannt ist.
- Wenn X und Y unabhängig: $H(Y \mid X) = H(Y)$.

Lemma

Für die Entropie der gemeinsamen Verteilung von X und Y gilt

$$H(X, Y) = H(X) + H(Y \mid X).$$

- Beweis: Definition und $p(y, x) = p(x)\, p(y \mid x)$.
- Iterierte Anwendung des Lemmas liefert $H(X_1, \ldots, X_n) = H(X_1) + H(X_2 \mid X_1) + \cdots + H(X_n \mid X_1, \ldots, X_{n-1})$.

- Wenn X bekannt ist, wieviele Werte kann Y noch annehmen?
- Partitioniere $\operatorname{supp} X = \{x_1, \ldots, x_n\}$ je nach Anzahl in $E_1, \ldots, E_d$:

$$x \in E_j \iff |\operatorname{supp}(Y \mid x)| = j.$$

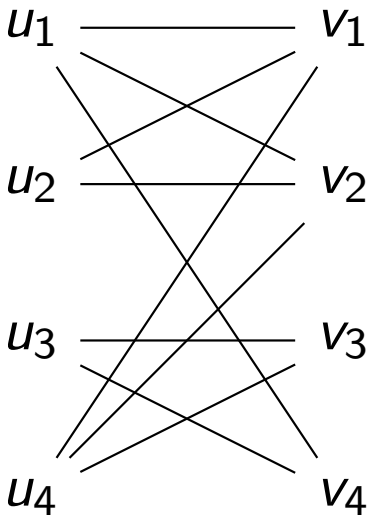
- Wenn X bekannt ist, wieviele Werte kann Y noch annehmen?
- Partitioniere $\operatorname{supp} X = \{x_1, \ldots, x_n\}$ je nach Anzahl in $E_1, \ldots, E_d$:

$$x \in E_j \iff |\operatorname{supp}(Y \mid x)| = j.$$

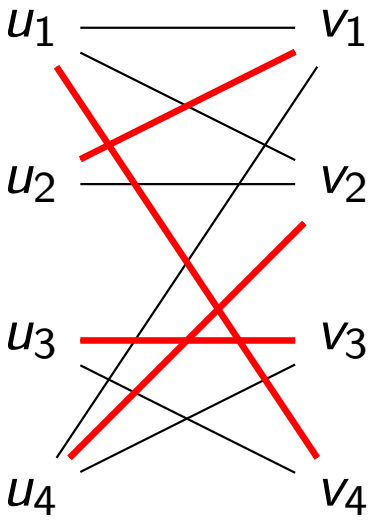
- Da Entropie $\leq \log_2 |\operatorname{supp}(\ldots)|$ ist, gilt

$$\begin{aligned} H(Y \mid X) &= \sum_{x \in \operatorname{supp} X} p(x)\, H(Y \mid x) \\ &= \sum_{j=1}^{d} \sum_{x \in E_j} p(x)\, \underbrace{H(Y \mid x)}_{\leq \log_2 j} \\ &\leq \sum_{j=1}^{d} \operatorname{Prob}(X \in E_j)\, \log_2 j. \end{aligned}$$

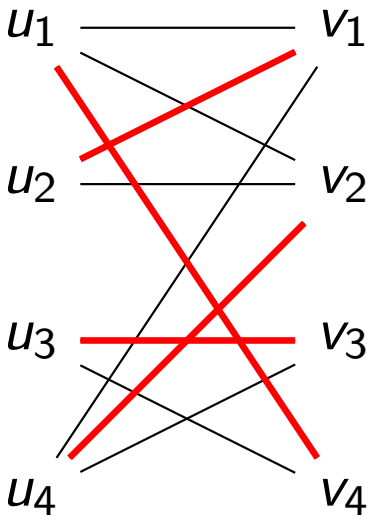
And now for something completely different. . .



- G bipartiter Graph mit n Knoten links und rechts



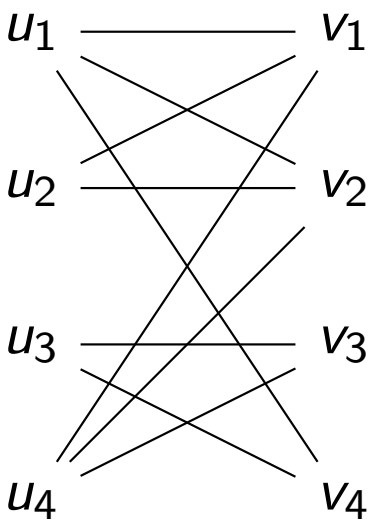
- G bipartiter Graph mit n Knoten links und rechts
- Matching: Menge von Kanten ohne gemeinsame Knoten
- . . . perfekt, wenn alle Knoten gematcht werden
- perfektes Matching weist jedem u_i genau ein $v_{\sigma(i)}$ zu.

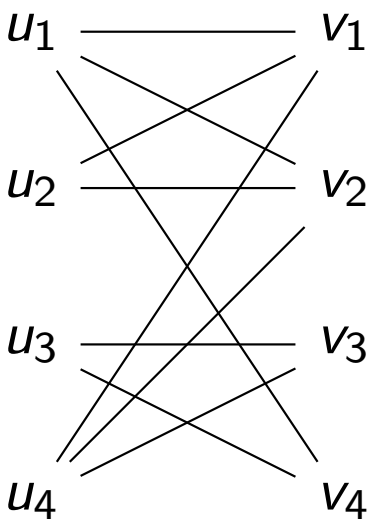


$\sigma = 4132$

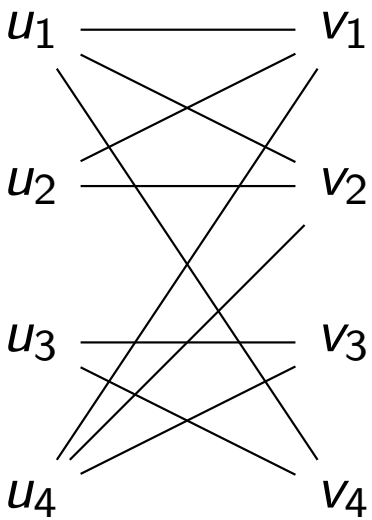
- G bipartiter Graph mit n Knoten links und rechts
- Matching: Menge von Kanten ohne gemeinsame Knoten
- . . . perfekt, wenn alle Knoten gematcht werden
- perfektes Matching weist jedem u_i genau ein $v_{\sigma(i)}$ zu.
- entspricht einer Permutation!

- Permutationen $\longrightarrow$ perfekte Matchings?
- $\sigma = 3142$ liefert kein Matching, da keine Kante zwischen u_1 und v_3.





- Permutationen $\longrightarrow$ perfekte Matchings?
- $\sigma = 3142$ liefert kein Matching, da keine Kante zwischen u_1 und v_3.
- Sei M_G die 0/1-Matrix mit $m_{ij} = 1 \Longleftrightarrow (u_i, v_j) \in E$.
- Zeilensumme $d_i = \sum_j m_{ij}$ ist Knotengrad von u_i.



- Permutationen $\longrightarrow$ perfekte Matchings?
- $\sigma = 3142$ liefert kein Matching, da keine Kante zwischen u_1 und v_3.
- Sei M_G die 0/1-Matrix mit $m_{ij} = 1 \iff (u_i, v_j) \in E$.
- Zeilensumme $d_i = \sum_j m_{ij}$ ist Knotengrad von u_i.

Definition (Permanente)

$$\operatorname{per} M := \sum_{\sigma \in S_n} m_{1\sigma(1)} m_{2\sigma(2)} \cdots m_{n\sigma(n)}$$

Produkt ist $1 \iff$ alle für σ benötigten Kanten in G.
Daher: $\boxed{\operatorname{per} M_G = \text{Anzahl perfekte Matchings in } G}$.

$$\operatorname{per} M = \sum_{\sigma \in S_n} m_{1\sigma(1)} m_{2\sigma(2)} \cdots m_{n\sigma(n)}$$

Theorem (Brégman 1973)

Sei M eine $n \times n$ Matrix mit Einträgen aus $\{0, 1\}$. Es gilt

$$\operatorname{per} M \leq \prod_{i=1}^{n} (d_i!)^{(1/d_i)}$$

wobei d_i die Zeilensumme $d_i = \sum_j m_{ij}$ ist.

$$\operatorname{per} M = \sum_{\sigma \in S_n} m_{1\sigma(1)} m_{2\sigma(2)} \cdots m_{n\sigma(n)}$$

Theorem (Brégman 1973)

Sei M eine $n \times n$ Matrix mit Einträgen aus $\{0, 1\}$. Es gilt

$$\operatorname{per} M \leq \prod_{i=1}^{n} (d_i!)^{(1/d_i)}$$

wobei d_i die Zeilensumme $d_i = \sum_j m_{ij}$ ist.

- per M (=Anzahl perfekter Matchings) für große 0/1-Matritzen sehr schwer zu bestimmen.
- Theorem liefert uns eine einfach zu bestimmende Abschätzung!

Wir wollen zeigen:

$$\operatorname{per} M \leq \prod_{i=1}^{n} (d_i!)^{(1/d_i)}$$

- per M ist die Anzahl perfekter Matchings in G_M.
- Sei σ eine gleichverteilte Zufallsvariable über den perfekten Matchings, dann gilt

$$H(\sigma(1), \ldots, \sigma(n)) = \log_2(\operatorname{per} M).$$

- . . . umformuliert ist also zu zeigen, dass

$$H(\sigma(1), \ldots, \sigma(n)) \leq \sum_{i=1}^{n} \frac{1}{d_i} \log_2(d_i!).$$

- Aus $H(X, Y) = H(X) + H(Y \mid X)$ erhalten wir

$$H(\sigma(1), \ldots, \sigma(n)) = \sum_{i=1}^{n} H(\sigma(i) \mid \sigma(1), \ldots, \sigma(i-1)).$$

- Aus $H(X, Y) = H(X) + H(Y \mid X)$ erhalten wir

$$H(\sigma(1), \ldots, \sigma(n)) = \sum_{i=1}^{n} H(\sigma(i) \mid \sigma(1), \ldots, \sigma(i-1)).$$

- Wir müssen $\sigma(i)$ aber nicht in der Reihenfolge $1, \ldots, n$ aufdecken!
- Sei $\tau \in S_n$ eine andere Reihenfolge und $k_i := \tau^{-1}(i)$ der neue Index von i unter τ, dann

$$H(\sigma(1), \ldots, \sigma(n)) = \sum_{i=1}^{n} H(\sigma(i) \mid \sigma(\tau(1)), \ldots, \sigma(\tau(k_i - 1))).$$

- Aus $H(X, Y) = H(X) + H(Y \mid X)$ erhalten wir

$$H(\sigma(1), \ldots, \sigma(n)) = \sum_{i=1}^{n} H(\sigma(i) \mid \sigma(1), \ldots, \sigma(i-1)).$$

- Wir müssen $\sigma(i)$ aber nicht in der Reihenfolge $1, \ldots, n$ aufdecken!
- Sei $\tau \in S_n$ eine andere Reihenfolge und $k_i := \tau^{-1}(i)$ der neue Index von i unter τ, dann

$$H(\sigma(1), \ldots, \sigma(n)) = \sum_{i=1}^{n} H(\sigma(i) \mid \sigma(\tau(1)), \ldots, \sigma(\tau(k_i - 1))).$$

- Da dies für alle $n!$ Reihenfolgen τ gilt,

$$\cdots = \frac{1}{n!} \sum_{\tau \in S_n} \sum_{i=1}^{n} H(\sigma(i) \mid \sigma(\tau(1)), \ldots, \sigma(\tau(k_i - 1))).$$

Fixiert: σ, τ, i

- Wir wollen $H(\sigma(i) \mid \overbrace{\sigma(\tau(1)), \ldots, \sigma(\tau(k_i - 1))}^{=:X})$ abschätzen.
- Partitioniere $\operatorname{supp} X$ in Mengen $E_{i,j}^{(\tau)}$:

  $x \in E_{i,j}^{(\tau)} :\Longleftrightarrow$ genau j Nachbarn von u_i sind noch frei unter x,

  wobei ein Nachbar v_ℓ von u_i noch frei ist, wenn er noch nicht von $u_{\tau(1)}, \ldots, u_{\tau(k_i-1)}$ gematcht wurde.

Fixiert: σ, τ, i

- Wir wollen $H(\sigma(i) \mid \overbrace{\sigma(\tau(1)), \ldots, \sigma(\tau(k_i - 1))}^{=:X})$ abschätzen.
- Partitioniere $\operatorname{supp} X$ in Mengen $E_{i,j}^{(\tau)}$:

  $x \in E_{i,j}^{(\tau)} :\Longleftrightarrow$ genau j Nachbarn von u_i sind noch frei unter x,

  wobei ein Nachbar v_ℓ von u_i noch frei ist, wenn er noch nicht von $u_{\tau(1)}, \ldots, u_{\tau(k_i-1)}$ gematcht wurde.
- Da σ einmal ein perfektes Matching sein will, muss mindestens ein Nachbar von u_i noch frei geblieben sein, also $1 \leq j \leq d_i$.

Fixiert: σ, τ, i

- Wir wollen $H(\sigma(i) \mid \overbrace{\sigma(\tau(1)), \ldots, \sigma(\tau(k_i - 1))}^{=:X})$ abschätzen.
- Partitioniere $\operatorname{supp} X$ in Mengen $E_{i,j}^{(\tau)}$:

  $x \in E_{i,j}^{(\tau)} :\Longleftrightarrow$ genau j Nachbarn von u_i sind noch frei unter x,

  wobei ein Nachbar v_ℓ von u_i noch frei ist, wenn er noch nicht von $u_{\tau(1)}, \ldots, u_{\tau(k_i-1)}$ gematcht wurde.
- Da σ einmal ein perfektes Matching sein will, muss mindestens ein Nachbar von u_i noch frei geblieben sein, also $1 \leq j \leq d_i$.
- Gleichzeitig muss $\sigma(i)$ einer der j noch freien Nachbarn werden, also $\operatorname{supp}(\sigma(i) \mid x) \leq j$.

$$\Longrightarrow H(\sigma(i) \mid x) \leq \log_2 j$$

$$\Longrightarrow H(\sigma(i) \mid X) \leq \sum_{j=1}^{d_i} \operatorname{Prob}(X \in E_{i,j}^{(\tau)}) \log_2 j.$$

Zusammenfassend

$$H(\sigma(1),\ldots,\sigma(n)) = \frac{1}{n!}\sum_{\tau\in S_n}\sum_{i=1}^{n} H(\sigma(i) \mid \overbrace{\sigma(\tau(1)),\ldots,\sigma(\tau(k_i-1))}^{=:X})$$

$$\leq \frac{1}{n!}\sum_{\tau\in S_n}\sum_{i=1}^{n}\sum_{j=1}^{d_i} \mathrm{Prob}(X \in E_{i,j}^{(\tau)}) \log_2 j$$

$$= \frac{1}{n!}\sum_{i=1}^{n}\sum_{j=1}^{d_i} \log_2 j \sum_{\tau\in S_n} \mathrm{Prob}(X \in E_{i,j}^{(\tau)}).$$

Zusammenfassend

$$H(\sigma(1),\dots,\sigma(n)) = \frac{1}{n!}\sum_{\tau\in S_n}\sum_{i=1}^{n} H(\sigma(i) \mid \overbrace{\sigma(\tau(1)),\dots,\sigma(\tau(k_i-1))}^{=:X})$$

$$\leq \frac{1}{n!}\sum_{\tau\in S_n}\sum_{i=1}^{n}\sum_{j=1}^{d_i} \mathrm{Prob}(X \in E_{i,j}^{(\tau)}) \log_2 j$$

$$= \frac{1}{n!}\sum_{i=1}^{n}\sum_{j=1}^{d_i} \log_2 j \sum_{\tau\in S_n} \mathrm{Prob}(X \in E_{i,j}^{(\tau)}).$$

Gelingt es uns zu zeigen, dass die innerste Summe $\frac{n!}{d_i}$ ist, so vereinfacht sich der Ausdruck zu

$$\cdots = \sum_{i=1}^{n} \frac{1}{d_i} \underbrace{\sum_{j=1}^{d_i} \log_2 j}_{\log_2 d_i!}$$

und wir sind fertig!

Fixiert: σ, τ, i

- Betrachte, in Reihenfolge τ, die Liste von Knoten, die mit den Nachbarn von u_i gematcht werden.
- Diese Liste enthält auch u_i selbst. Aber an welcher Stelle?

Fixiert: σ, τ, i

- Betrachte, in Reihenfolge τ, die Liste von Knoten, die mit den Nachbarn von u_i gematcht werden.
- Diese Liste enthält auch u_i selbst. Aber an welcher Stelle?
- Finden wir u_i an erster Stelle, so sind noch alle d_i Nachbarn von u_i frei geblieben, bis wir in τ zu i gekommen sind.
- . . . an zweiter Stelle, $d_i - 1$ Nachbarn frei geblieben usw.

Fixiert: σ, i

- Betrachte, in Reihenfolge τ, die Liste von Knoten, die mit den Nachbarn von u_i gematcht werden.
- Diese Liste enthält auch u_i selbst. Aber an welcher Stelle?
- Finden wir u_i an erster Stelle, so sind noch alle d_i Nachbarn von u_i frei geblieben, bis wir in τ zu i gekommen sind.
- . . . an zweiter Stelle, $d_i - 1$ Nachbarn frei geblieben usw.
- Über alle Reihenfolgen τ tritt u_i aber an allen d_i Stellen mit gleicher Wahrscheinlichkeit auf!
- Da $x \in E_{i,j}^{(\tau)} :\Longleftrightarrow$ genau j Nachbarn sind noch frei geblieben, haben wir deshalb

$$\frac{1}{n!} \sum_{\tau \in S_n} \text{Prob}(X \in E_{i,j}^{(\tau)}) = \frac{1}{d_i}.$$

□