Die AGM-Schranke besagt, dass die Ungleichung (***) sogar für jede Abbildung $x : [1,m] \longrightarrow \mathbb{Q}_{\geq 0}$, die (**) erfüllt, gilt.

<u>Notation</u>: $\mathbb{Q}_{\geq 0}$ bezeichnet die Menge aller nicht-negativen rationalen Zahlen

Präzise Formulierung der AGM-Schranke:

<u>Definition 2.8</u> (fraktionale Kantenüberdeckung)

Sei G ein beliebiger Hypergraph mit Knotenmenge V und Hyperkantenmenge E. Eine fraktionale Kantenüberdeckung (kurz: fKü, engl.: fractional edge cover) von G ist eine Abbildung $x : E \longrightarrow \mathbb{Q}_{\geq 0}$, so dass für <u>jeden</u> Knoten $v \in V$ gilt:

$$\sum_{\substack{e \in E: \\ v \in e}} x(e) \geq 1 .$$

<u>Notation</u>:

Wenn wir Join-Anfragen Q betrachten, sprechen wir von "fraktionalen Kantenüberdeckungen von Q" und meinen damit fraktionale Kantenüberdeckungen des Anfrage-Hypergraphen G_Q. Statt "$x(R_i(\bar{x}_i))$" schreiben wir auch kurz: "$x(i)$".

Satz 2.9 (Die AGM-Schranke)

Sei Q eine Join-Anfrage der Form
$Q(\overline{x}_0) \leftarrow R_1(\overline{x}_1), \ldots, R_m(\overline{x}_m)$.

Für jede fraktionale Kantenüberdeckung x von Q und jede Datenbank D gilt:

$$|Q(D)| \leq \prod_{i=1}^{m} N_i^{x(i)},$$

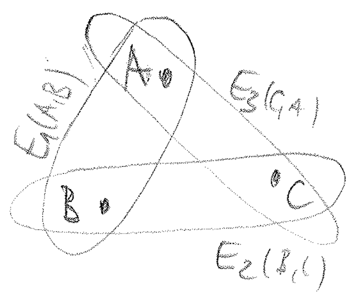
wobei $N_i := |R_i^D|$ für alle $i \in [1,m]$ ist.

Bevor wir die AGM-Schranke beweisen, schauen wir uns zunächst einige Beispiele an.

Beispiel 2.10

(a) Betrachte die Anfrage $Q'_\Delta(A,B,C) \leftarrow E_1(A,B), E_2(B,C), E_3(C,A)$

Skizze von $G_{Q'_\Delta}$:



Die Abbildung x mit $x(1) = x(2) = x(3) = \frac{1}{2}$ ist eine f.Kü von Q'_Δ.
Laut AGM-Schranke gilt: $|Q'_\Delta(D)| \leq \prod_{i=1}^{3} N_i^{1/2} = \sqrt{N_1}\cdot\sqrt{N_2}\cdot\sqrt{N_3}$
f.a. DBen D und $N_i := E_i^D$ für $i \in \{1,2,3\}$.

Dies liefert auch eine leicht verbesserte Variante von Satz Δ:

Für die Anfrage $Q_\Delta(A,B,C) \leftarrow E(A,B), E(B,C), E(C,A)$ und jede $\{E\}$-DB D gilt:

$$|Q_\Delta(D)| \leq (N^D)^{3/2},$$

wobei $N^D := |E^D|$ ist

Beweis:

Sei D eine $\{E\}$-DB und sei D' die $\{E_1, E_2, E_3\}$-DB mit $E_i^{D'} := E^D$ f.a. $i \in [1,3]$. Dann gilt:

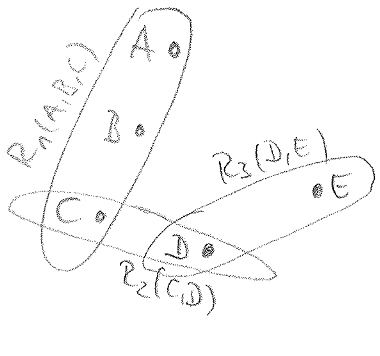
$Q_\Delta(D) = Q'_\Delta(D')$. Gemäß AGM-Schranke folgt:

$$|Q_\Delta(D)| = |Q'_\Delta(D')| \leq \sqrt{N^D} \cdot \sqrt{N^D} \cdot \sqrt{N^D} = (N^D)^{3/2}.$$

(b) Betrachte die Anfrage $Q_{(b)}$ aus Beispiel 2.5:

$$Q_{(b)}(A,B,C,D,E) \leftarrow R_1(A,B,C), R_2(C,D), R_3(D,E)$$

$G_{Q_{(b)}}$:

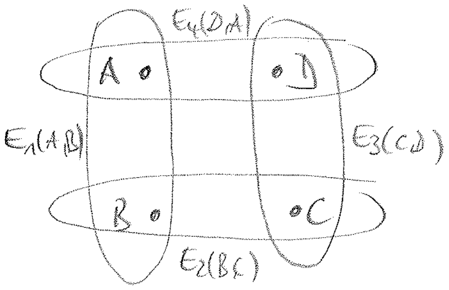


Für jede f.Kü x von $Q_{(b)}$ gilt: $x(1) \geq 1$ und $x(3) \geq 1$, da der Knoten A (bzw. E) nur in einer einzigen Hyperkante, nämlich $R_1(A,B,C)$ (bzw. $R_3(D,E)$) vorkommt.

(c) Betrachte die Anfrage

$$Q_4(A,B,C,D) \leftarrow E_1(A,B),\ E_2(B,C),\ E_3(C,D),\ E_4(D,A)$$

G_{Q_4}:



einige fKü von Q_4:

	1	2	3	4
$x^{(1)}$	1	0	1	0
$x^{(2)}$	0	1	0	1
$x^{(3)}$	$1/2$	$1/2$	$1/2$	$1/2$
$x^{(4)}$	$1/3$	$2/3$	$1/3$	$2/3$

(d) Sei $k \in \mathbb{N}$ mit $k \geq 2$. Betrachte die Loomis-Whitney Join-Anfrage LW_k mit

$$LW_k(X_1,\dots,X_k) \leftarrow \underbrace{R_1(X_2,\dots,X_k),\ R_2(X_1,X_3,\dots,X_k),\ \dots,\ R_k(X_1,\dots,X_{k-1})}_{\text{d.h. für jedes } i\in[1,k] \text{ ein Atom der Form } R_i(X_1,\dots,X_{i-1},X_{i+1},\dots,X_k)}$$

Eine fKü von LW_k ist die Abbildung $x : [1,k] \to \mathbb{Q}_{\geq 0}$ mit $x(i) := 1/_{k-1}$ f.a. $i \in [1,k]$.

Die AGM-Schranke liefert: $\left| LW_k(\mathcal{D}) \right| \leq \prod_{i=1}^{k} \sqrt[k-1]{N_i}$,

wobei $N_i := |R_i^{\mathcal{D}}|$ f.a. $i \in [1,k]$ ist.

Insbes.: Falls $N_1 = \dots = N_k = N$, so ist $|LW_k(\mathcal{D})| \leq N^{\frac{k}{k-1}} = N \cdot \sqrt[k-1]{N}$.

Betrachten wir nun folgendes Szenario:

Gegeben sei eine Join-Anfrage

$$Q(\overline{X}_0) \leftarrow R_1(\overline{X}_1), \ldots, R_m(\overline{X}_m)$$

mit $vars(Q) = \{A_1, \ldots, A_n\}$.

Außerdem seien Zahlen $N_1, \ldots, N_m$ bekannt, von denen wir wissen, dass die Datenbank D, auf der Q ausgewertet wird in der Relation R_i genau N_i Tupel enthält (d.h. $N_i = |R_i^D|$), f.a. $i \in [1,m]$

Die AGM-Schranke besagt, dass für jede fKü x von Q gilt: $$|Q(D)| \leq \prod_{i=1}^{m} N_i^{x(i)}$$

Frage: Wie können wir eine fKü x finden, für die $\prod_{i=1}^{m} N_i^{x(i)}$ möglichst klein ist?

Dazu lösen wir folgendes Optimierungsproblem:

Statt $x(i)$ schreibe x_i für $i \in [1,m]$ und setze $x := \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$.

Ziel:

minimiere $\prod_{i=1}^{m} N_i^{x_i}$

unter der Bedingung, dass Folgendes gilt:

(1). f.a. $j \in [1,n]$ ist $\sum_{\substack{i \in [1,m]: \\ A_j \in \{\overline{X}_i\}}} x_i \geq 1$

und

(2). f.a. $i \in [1,m]$ ist $x_i \geq 0$, $x_i \in \mathbb{Q}$

<u>Beachte:</u>

- $\prod_{i=1}^{m} N_i^{x_i}$ ist minimal

$$\iff \underbrace{\log\left(\prod_{i=1}^{m} N_i^{x_i}\right)}_{} \text{ ist minimal}$$

$$= \sum_{i=1}^{m} \log\left(N_i^{x_i}\right) = \sum_{i=1}^{m} x_i \cdot \log N_i = \sum_{i=1}^{m} x_i \cdot c_i$$

mit $c_i := \log N_i$ f.a. $i \in [1,m]$.

Als Zielfunktion für unser Minimierungsproblem können wir statt $\prod_{i=1}^{m} N_i^{x_i}$ also auch $c^T x$ verwenden.

- Die Bedingung (1) lässt sich darstellen als "$Ax \geq b$" für die Matrix $A_Q := A := (a_{ji})_{j\in[1,n],\, i\in[1,m]}$ mit

$$a_{ji} := \begin{cases} 1 & \text{falls } A_j \in \{\overline{X}_i\} \\ 0 & \text{sonst} \end{cases}$$

und den Vektor $b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$

<u>Unser Ziel ist also:</u>

(OPT:)

$$\boxed{\begin{array}{l} \text{minimiere } c^T x \quad \left(\text{mit } c_i := \log N_i \text{ f.a. } i \in [1,m]\right) \\ \text{unter der Bedingung, dass} \\ \quad Ax \geq b, \quad \left(\text{mit } A = (a_{ji}) \text{ und } a_{ji} := \begin{cases} 1 & \text{falls } A_j \in \{\overline{X}_i\} \\ 0 & \text{sonst} \end{cases}\right) \\ \quad x \geq 0, \quad x \in \mathbb{Q}^m \end{array}}$$

Dies ist ein lineares Optimierungproblem, für das wir eine rationale Lösung suchen.

Dazu stehen verschiedene Lösungsverfahren zur Verfügung:

- der Simplex-Algorithmus (exponentiell im worst-case, aber in der Praxis i.d.R. effizient)
- das Innere-Punkte-Verfahren (Karmarkar-Verfahren (1984) — läuft in Polynomialzeit)
- die Ellipsoid-Methode (Khachiyan (1979) — läuft in Polynomialzeit)

Beachte:

Ist x eine gemäß AGM-Schranke Lösung des Optimierungsproblems OPT, so gilt für jede DB D mit $N_i = |R_i^D|$ f.a. $i \in [1,m]$, dass $|Q(D)| \leq 2^{c^T x}$ ist.

Denn:
$$2^{c^T x} = 2^{\sum_{i=1}^{m} c_i x_i} = 2^{\sum_{i=1}^{m} x_i \cdot \log N_i} = 2^{\sum_{i=1}^{m} \log(N_i^{x_i})}$$
$$= 2^{\log(\prod_{i=1}^{m} N_i^{x_i})} = \prod_{i=1}^{m} N_i^{x_i}.$$

Beweis der AGM-Schranke

Zum Beweis benötigen wir einige Grundlagen aus der Wahrscheinlichkeitsrechnung und der Informationstheorie, die wir hier zunächst bereitstellen.

Wahrscheinlichkeitsräume

Definition: Ein endlicher Wahrscheinlichkeitsraum (Ω, P) besteht aus einer endlichen, nicht-leeren Menge Ω von Ergebnissen bzw. Elementarereignissen, denen Wahrscheinlichkeiten $P(\omega) = p_\omega \in \mathbb{R}$ für jedes $\omega \in \Omega$ zugeordnet sind, so dass gilt:

$$0 \leq p_\omega \leq 1, \text{ für jedes } \omega \in \Omega, \quad \text{und} \quad \sum_{\omega \in \Omega} p_\omega = 1.$$

Ereignisse

Definition: Ein Ereignis ist eine Menge von Ergebnissen, d.h. eine Teilmenge von Ω.
Die Wahrscheinlichkeit eines Ereignisses $A \subseteq \Omega$ ist definiert als $P(A) := \sum_{\omega \in A} P(\omega)$.

Wir schreiben $\overline{A}$ um das Komplement von A zu bezeichnen, d.h. $\overline{A} = \Omega \setminus A$

Regeln zum Rechnen mit Wahrscheinlichkeiten:

Für jeden endlichen Wahrscheinlichkeitsraum (Ω, P) und für alle Ereignisse A und B gilt:

- $P(\overline{A}) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Für $s \in \mathbb{N}_{\geq 1}$ und beliebige Ereignisse $A_1, \dots, A_s \subseteq \Omega$ gilt:
$$P(A_1 \cup \dots \cup A_s) \leq \sum_{i=1}^{s} P(A_i)$$

Falls $A_i \cap A_j = \emptyset$ für alle $i,j \in [1,s]$ mit $i \neq j$, so gilt sogar:
$$P(A_1 \cup \dots \cup A_s) = \sum_{i=1}^{s} P(A_i).$$

Zufallsvariablen

Definition:

Sei (Ω, P) ein endlicher Wahrscheinlichkeitsraum.
Eine Zufallsvariable ist eine Funktion $Y: \Omega \to M$, für eine beliebige nicht-leere Menge M.
Für jedes $a \in M$ definieren wir die Wahrscheinlichkeit dafür, dass die Zufallsvariable Y den Wert a annimmt, durch
$$P(Y=a) := P(A) \text{ für das Ereignis } A := \{\omega \in \Omega : Y(\omega) = a\}$$

Beispiel 2.11

Betrachte eine Join-Anfrage Q der Form

$$Q(\overline{X_0}) \leftarrow R_1(\overline{X_1}), \dots, R_m(\overline{X_m})$$

mit $\overline{X_0} = A_1, \dots, A_n$.

Sei D eine Datenbank vom Schema $\{R_1, \dots, R_m\}$.

Wir betrachten das "Zufallsexperiment", bei dem wir zufällig, gleichverteilt ein beliebiges Tupel aus dem Anfrageergebnis $Q(D)$ ziehen.

Dies entspricht dem endlichen Wahrscheinlichkeitsraum (Ω, P) mit $\Omega := Q(D)$ und $P(\omega) := \frac{1}{|Q(D)|}$ f.a. $\omega \in Q(D)$.

Sei $M := adom(D)$.

Für jedes $j \in [1,n]$ sei $Y_j : \Omega \to M$ die Zufallsvariable, die jedem Tupel $t = (t_1, \dots, t_n) \in Q(D)$ den Wert t_j zuordnet — d.h. $Y_j(t) = \pi_j(t)$ f.a. $t \in Q(D)$

Für jedes feste Element $a \in adom(D)$ ist dann

$P(Y_j = a)$ die Wahrscheinlichkeit dafür, dass ein zufällig gleichverteilt aus $Q(D)$ gewähltes Tupel in der j-ten Spalte den Eintrag a hat. D.h.

$$P(Y_j = a) = \frac{|\{t \in Q(D) : \pi_j(t) = a\}|}{|Q(D)|}.$$

Informationsgehalt und Entropie

Definition:

Sei (Ω, P) ein endlicher Wahrscheinlichkeitsraum

Der <u>Informationsgehalt</u> eines Elementarereignisses $\omega \in \Omega$ ist definiert als

$$\log\left(\frac{1}{p_\omega}\right) \qquad \left(= -\log(p_\omega) \right)$$

Die <u>Entropie</u> von (Ω, P) ist definiert als

$$H((\Omega,P)) := \sum_{\omega\in\Omega} p_\omega \cdot \log\left(\frac{1}{p_\omega}\right) \qquad \left(= -\sum_{\omega\in\Omega} p_\omega \cdot \log(p_\omega) \right)$$

Konvention für den Fall, dass $p_\omega = 0$ ist:

$$0 \cdot \log\left(\frac{1}{0}\right) := -0 \cdot \log(0) := 0 .$$

Bemerkung:

Man kann beweisen, dass Folgendes gilt:

$$H((\Omega,P)) \leq \begin{bmatrix} \text{durchschnittliche Länge des Bitstrings,} \\ \text{der ein gemäß } P \text{ zufällig aus } \Omega \\ \text{gewähltes Elementarereignis kodiert} \\ \text{(bei optimaler Kodierung)} \end{bmatrix} \leq H((\Omega,P))+1$$

Wegen $H((\Omega,P)) \stackrel{\text{Def}}{=} \sum_{\omega\in\Omega} p_\omega \cdot \log\left(\frac{1}{p_\omega}\right)$ gilt daher

$$\begin{bmatrix} \text{Informationsgehalt} \\ \text{von } \omega\in\Omega \end{bmatrix} \stackrel{\text{Def}}{=} \log\left(\frac{1}{p_\omega}\right) \approx \begin{bmatrix} \text{Länge des Bitstrings,} \\ \text{der } \omega \text{ kodiert} \\ \text{(bei optimaler Kodierung)} \end{bmatrix}$$

Definition (Entropie einer Zufallsvariablen)

Sei (Ω, P) ein endlicher Wahrscheinlichkeitsraum, sei M eine beliebige nicht-leere Menge und sei $Y: \Omega \longrightarrow M$ eine Zufallsvariable.

Die Entropie von Y ist definiert als

$$H(Y) := \sum_{a \in M} P(Y=a) \cdot \log\left(\frac{1}{P(Y=a)}\right).$$

Bemerkung:

Wir können uns $H(Y)$ vorstellen als ein Maß für die "Unsicherheit" ~~bzw.~~ "Unbestimmtheit" von Y.

Falls es ein $b \in M$ gibt, so das $P(Y=b) = 1$, so wissen wir, dass die Zufallsvariable, unabhängig vom Ausgang des Zufallsexperiments, stets den Wert b annimmt — die "Unsicherheit" bzw. "Unbestimmtheit" von Y ist also 0.

Die spiegelt sich in der Entropie wider, da

$$H(Y) \overset{\text{Def}}{=} \sum_{a \in M} P(Y=a) \cdot \log\left(\frac{1}{P(Y=a)}\right) = P(Y=b) \cdot \log\left(\frac{1}{P(Y=b)}\right)$$
$$= 1 \cdot \log\left(\frac{1}{1}\right) = 0$$

ist.

Falls jedoch für jedes $b \in M$ gilt, dass $P(Y=b) = \frac{1}{|M|}$, so ist die "Unsicherheit / Unbestimmtheit" von Y größtmöglich, und es gilt: $H(Y) = \sum_{a \in M} \frac{1}{|M|} \cdot \log(|M|) = \log(|M|)$

Wichtige Eigenschaften der Entropie:

Sei (Ω, P) ein endlicher Wahrscheinlichkeitsraum

(1) Für jede Zufallsvariable $Y: \Omega \to M$ mit endlichem M gilt:

$$0 \leq H(Y) \leq \log(|M|)$$

Außerdem gilt:

- $H(Y) = 0 \iff$ ex $b \in M$ s.d. $P(Y=b) = 1$
- $H(Y) = \log(|M|) \iff$ f.a. $b \in M$ ist $P(Y=b) = \frac{1}{|M|}$.

(2) Seien $Y_1: \Omega \to M_1$ und $Y_2: \Omega \to M_2$ zwei Zufallsvariablen.
Die zusammengesetzte Zufallsvariable (Y_1, Y_2) ist die Zufallsvariable $Y: \Omega \to M_1 \times M_2$ mit $Y(\omega) := (Y_1(\omega), Y_2(\omega))$ f.a. $\omega \in \Omega$.

- Die <u>gemeinsame Entropie von Y_1 und Y_2</u> ist die Entropie von (Y_1, Y_2). D.h.:

$$H(Y_1, Y_2) = \sum_{\substack{a_1 \in M_1, \\ a_2 \in M_2}} P(Y_1 = a_1, Y_2 = a_2) \cdot \log\left(\frac{1}{P(Y_1 = a_1, Y_2 = a_2)}\right)$$

Komma $\hat{=}$ "und"

- Für $a_1 \in M_1$ mit $P(Y_1 = a_1) \neq 0$ ist die <u>bedingte Entropie von Y_2 unter der Voraussetzung, dass $Y_1 = a_1$ ist,</u>

$$H(Y_2 \mid Y_1 = a_1) := \sum_{a_2 \in M_2} P(Y_2 = a_2 \mid Y_1 = a_1) \cdot \log\left(\frac{1}{P(Y_2 = a_2 \mid Y_1 = a_1)}\right)$$

Hierbei ist $P(Y_2 = a_2 \mid Y_1 = a_1)$ die bedingte Wahrscheinlichkeit dafür, dass Y_2 den Wert a_2 annimmt unter der Voraussetzung, dass Y_1 den Wert a_1 angenommen hat.

D.h. $P(Y_2 = a_2 \mid Y_1 = a_1) := \dfrac{P(Y_2 = a_2, Y_1 = a_1)}{P(Y_1 = a_1)}$.

- Die <u>bedingte Entropie von Y_2 bei gegebenem Y_1</u> ist

$$H(Y_2 \mid Y_1) := \sum_{a_1 \in M_1} P(Y_1 = a_1) \cdot H(Y_2 \mid Y_1 = a_1)$$

$$= \sum_{a_1 \in M_1} \left(P(Y_1 = a_1) \cdot \sum_{a_2 \in M_2} P(Y_2 = a_2 \mid Y_1 = a_1) \cdot \log\left(\frac{1}{P(Y_2 = a_2 \mid Y_1 = a_1)}\right)\right)$$

- Es gilt: $\boxed{H(Y_1, Y_2) = H(Y_1) + H(Y_2 \mid Y_1)}$ (1)

<u>Beweis:</u>

$$H(Y_1, Y_2) \overset{\text{Def}}{=} \sum_{\substack{a_1 \in M_1, \\ a_2 \in M_2}} \overbrace{P(Y_1 = a_1, Y_2 = a_2)}^{= P(Y_1 = a_1) \cdot P(Y_2 = a_2 \mid Y_1 = a_1)} \cdot \log\left(\frac{1}{P(Y_1 = a_1, Y_2 = a_2)}\right)$$

$$= \sum_{a_1 \in M_1} \left(P(Y_1 = a_1) \cdot \sum_{a_2 \in M_2} P(Y_2 = a_2 \mid Y_1 = a_1) \cdot \left(\log\left(\frac{1}{P(Y_1 = a_1)}\right) + \log\left(\frac{1}{P(Y_2 = a_2 \mid Y_1 = a_1)}\right)\right)\right)$$

$$= \sum_{a_1 \in M_1} P(Y_1 = a_1) \cdot \log\left(\frac{1}{P(Y_1 = a_1)}\right) \cdot \underbrace{\left(\sum_{a_2 \in M_2} P(Y_2 = a_2 \mid Y_1 = a_1)\right)}_{= 1}$$

$$+ \sum_{a_1 \in M_1} P(Y_1 = a_1) \cdot \left(\sum_{a_2 \in M_2} P(Y_2 = a_2 \mid Y_1 = a_1) \cdot \log\left(\frac{1}{P(Y_2 = a_2 \mid Y_1 = a_1)}\right)\right)$$

$$= H(Y_1) + H(Y_2 \mid Y_1) \qquad \square$$

- Es gilt: $\boxed{H(Y_2 | Y_1) \leq H(Y_2)}$ (2)

Intuition dazu: Die "Unsicherheit von Y_2" kann durch die zusätzliche Information hinsichtlich Y_1 höchstens abnehmen, aber nicht größer werden.

Beweis von (2): Übung!

Hinweis: Verwende die so genannte Jensensche Ungleichung, die Folgendes besagt:

Für jede konkave Funktion f und für Zahlen $\lambda_1, \dots, \lambda_n \in \mathbb{R}_{\geq 0}$ mit $\sum_{i=1}^{n} \lambda_i = 1$ gilt

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \geq \sum_{i=1}^{n} \lambda_i f(x_i) \quad \text{für alle } x_1, \dots, x_n \in \mathbb{R}$$

Eine Funktion $f: \mathbb{R} \to \mathbb{R}$ heißt konkav, wenn für alle $x, y \in \mathbb{R}$ und alle $\lambda \in \mathbb{R}$ mit $0 \leq \lambda \leq 1$ gilt:

$$f(\lambda x + (1-\lambda) y) \geq \lambda f(x) + (1-\lambda) f(y)$$

Insbesondere ist die Funktion $f(x) := \log x$ konkav.

(3) Die Begriffe aus (2) können für zusammengesetzte Zufallsvariablen $(Y_1, \dots, Y_n)$ verallgemeinert werden:

Sei $n \geq 2$. Für jedes $j \in [n]$ sei $Y_j : \Omega \to M_j$ eine Zufallsvariable.

Die zusammengesetzte Zufallsvariable $(Y_1, \dots, Y_n)$ ist die Zufallsvariable $Y: \Omega \to M_1 \times \dots \times M_n$ mit

$$Y(\omega) := (Y_1(\omega), \dots, Y_n(\omega)) \quad \text{f.a. } \omega \in \Omega.$$

- Die gemeinsame Entropie von $Y_1, \dots, Y_n$ ist

$$H(Y_1, \dots, Y_n) := \sum_{(a_1,\dots,a_n) \in M_1 \times \dots \times M_n} P\left(\bigwedge_{j=1}^{n} Y_j = a_j\right) \cdot \log\left(\frac{1}{P\left(\bigwedge_{j=1}^{n} Y_j = a_j\right)}\right)$$

- Analog zu (1) gilt

$$\boxed{H(Y_1, \dots, Y_n) = H(Y_1) + H(Y_2 \mid Y_1) + H(Y_3 \mid Y_1, Y_2) + \dots + H(Y_n \mid Y_1, \dots, Y_{n-1})} \qquad (1')$$

- Analog zu (2) gilt f.a. $J \subseteq J' \subseteq [n]$ und f.a. $k \in [n]$:

$$\boxed{H\left(Y_k \mid (Y_{j'})_{j' \in J'}\right) \leq H\left(Y_k \mid (Y_j)_{j \in J}\right)} \qquad (2')$$

Das zentrale Werkzeug, das wir zum Beweis der AGM-Schranke verwenden werden, ist

Shearer's Lemma: Sei (Ω, P) ein endlicher Wahrscheinlichkeitsraum. Sei $n \in \mathbb{N}_{\geq 1}$ und sei $I := \{1, \dots, n\}$. Für jedes $j \in I$ sei $Y_j : \Omega \to M_j$ eine Zufallsvariable.
Für jedes $J \subseteq I$ sei $Y_J := (Y_j)_{j \in J}$.
Sei $\ell \in \mathbb{N}_{\geq 1}$, und seien $J_1, \dots, J_\ell$ Teilmengen von I, so dass für eine Zahl $q \in \mathbb{N}$ gilt: Jedes $j \in I$ kommt in mindestens q der Mengen $J_1, \dots, J_\ell$ vor.
Dann gilt: $H(Y_1, \dots, Y_n) \leq \frac{1}{q} \cdot \sum_{k=1}^{\ell} H(Y_{J_k})$

Beweis:

Gemäß (1) gilt für jedes $J \subseteq I$, dass

$$H(Y_J) = \sum_{j \in J} H\left(Y_j \,\middle|\, (Y_i)_{i \in J \text{ mit } i < j}\right). \quad \boxed{*}$$

Somit gilt:

$$\sum_{k=1}^{\ell} H(Y_{J_k})$$

$$\overset{\boxed{*}}{=} \sum_{k=1}^{\ell} \sum_{j \in J_k} H\left(Y_j \,\middle|\, (Y_i)_{i \in J_k \text{ mit } i < j}\right)$$

$$\overset{(2)}{\geq} \sum_{k=1}^{\ell} \sum_{j \in J_k} H\left(Y_j \,\middle|\, (Y_i)_{i \in I \text{ mit } i < j}\right)$$

$$\geq q \cdot \sum_{j \in I} H\left(Y_j \,\middle|\, (Y_i)_{i \in I \text{ mit } i < j}\right) \overset{\boxed{*}}{=} q \cdot H(\underbrace{Y_I}_{=(Y_1,\dots,Y_n)})$$

denn laut Voraussetzung gilt:
jedes $j \in I$ kommt in
mindestens q der Mengen $J_1, \dots, J_\ell$ vor.

$\square$