

Aufgabe 1: Optimierung und Kostenmodelle (1 P.)

Gegeben die Relationen $R(A, B, C)$, $S(D, E, F)$ und $U(A, D, G)$ mit folgenden Statistiken:

$T(R) = 1000$	$V(R, A) = 4$	$V(S, D) = 5$	$V(U, A) = 5$
$T(S) = 1000$	$V(R, B) = 10$	$V(S, E) = 50$	$V(U, D) = 4$
$T(U) = 20$	$V(R, C) = 200$	$V(S, F) = 10$	$V(U, G) = 20$

$R.A$ ist Fremdschlüssel auf $U.A$, $S.D$ auf $U.D$. Die Einträge pro Attribut sind zwischen 1 und Kardinalität, z.B. nimmt $S.F$ Werte zwischen 1 und 10 an. Weiterhin gibt es keine NULL-Werte und Sie können von gleichverteilten und unabhängigen Werten ausgehen.

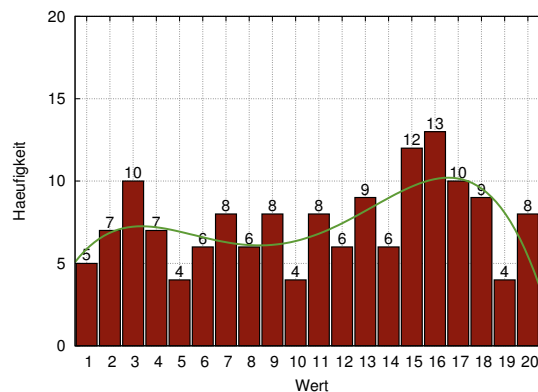
Betrachten Sie die folgenden Anfragen:

- i) $\sigma_{R.A=U.A}(\sigma_{U.D=S.D}(\sigma_{R.B=6}(\sigma_{U.G \neq 10}(R \times (S \times U))))))$
- ii) $\pi_{R.A,R.B}(\sigma_{R.B=2}(\sigma_{U.G \leq 10}(\sigma_{U.A=R.A}((R \times U) \cup (U \times R))))))$

Optimieren Sie die Anfragen nach der in der Vorlesung vorgestellten Vorgehensweise (“Regelbasierte Optimierung”). Benutzen Sie dabei die im Kapitel über Relationale Algebra vorgestellten Äquivalenzregeln. Geben Sie für Ihren finalen Plan die Größe des Ergebnisses sowie der Zwischenergebnisse in Tupeln an.

Aufgabe 2: Histogramme (1 P.)

Gegeben eine Sequenz von Integer-Zahlen aus $[1, 20]$ mit folgender Verteilung:



Berechnen Sie das dazugehörige Equi-Width Histogramm \mathcal{H}_2 mit Zellenbreite 2 sowie das Equi-Width Histogramm \mathcal{H}_4 mit Zellenbreite 4 und betrachten Sie zusätzlich folgendes Polynom vierten Grades, das an die obige Verteilung angepasst wurde (siehe Linie im Plot)

$$p(x) := -0.00125 * x^4 + 0.0473 * x^3 - 0.5606 * x^2 + 2.355 * x + 4.034$$

Berechnen Sie je anhand von (1.) \mathcal{H}_2 , (2.) \mathcal{H}_4 , (3.) unter Annahme der Gleichverteilung ($g(x) = 7.5$), sowie (4.) anhand von p Abschätzungen für:

- Die Häufigkeit von Zahlen kleiner oder gleich 13.

Informationssysteme SS 2017

Prof. Dr.-Ing. Stefan Deßloch

Dipl.-Inf Weiping Qu

TU Kaiserslautern, FB Informatik – Lehrgebiet Informationssysteme <http://www.lgis.informatik.uni-kl.de>

Übungsblatt 11: Ausgabe 26.06.2017, Abgabe 03.07.2017, 16:00 Uhr



- Die Häufigkeit von Zahlen in $[17, 20]$.

Und geben Sie jeweils den absoluten sowie den relativen Fehler gegenüber der tatsächlichen Häufigkeiten an.