Algoritm de coliziune probabilistic

Fie G un grup şi $g \in G$ un element al sau de ordin $ord(g) = N$. Presuounem că PLD $g^x = h$ are soluţie. Căutăm soluţia x astfel ca $x = y - z$ şi ne uităm acum după numerele y şi z astfel ca $g^y = h \cdot g^z$. Pentru aceasta vom forma o listă cu valori de tipul g^y şi o listă cu valori de tipul $h \cdot g^z$ şi căutăm potriviri între cele două liste (coliziuni).

- alegem la întâmplare exponenţii $y_1, \ldots, y_n$ în intervalul $[1, N]$ şi calculăm valorile

$$l_1 = \{g^{y_1}, \ldots, g^{y_n}\} \subset G.$$

Să remarcăm că $l_1 \subset \{g^1, \ldots, g^N\} = S$ şi deci avem o alegere a (aproximativ) n elemente din S.

- alegem la întâmplare exponenţii $z_1, \ldots, z_n$ în intervalul $[1, N]$ şi calculăm valorile

$$l_2 = \{h \cdot g^{z_1}, \ldots, h \cdot g^{z_n}\} \subset G.$$

Deoarece am presupus că ecuaţia $g^x = h$ are soluţie, rezultă că $l_2 \subset S$ (h este o putere a lui g).

Care este probabilitatea să avem coliziuni între cele două liste şi cât de mare ar trebui să fie n pentru a avea o probabilitate rezonabilă de coliziune?

Reformulăm problema: Avem o cutie cu N bile din care (aproximativ) n bile sunt roşii (adică lista l_1) şi $N - n$ albastre. Scoatem o bilă la întâplare, ne uităm la ea şi o punem la loc în cutie. Procedăm tot aşa de n ori (bilele extrase sunt analogul listei l_2). Care este probabilitatea ca să fi extras cel puţin o bilă roşie (adica să avem coliziuni).

Probabilitatea căutată este egală cu $1 - P'$, unde P' este probabilitatea ca toate bilele extrase să fie albastre. Este clar că

$$P = 1 - \left(\frac{N-n}{N}\right)^n = 1 - \left(1 - \frac{n}{N}\right)^n.$$

Aici $\frac{N-n}{N}$ este probabilitatea ca bila i extrasă să fie albastră.

Deoarece $1 - x \leq e^{-x}$ pentru orice $x \in \mathbb{R}$ pentru $x = \frac{n}{N}$ avem

$$P \geq 1 - e^{-n^2/N}.$$

Pe de altă parte, dacă N este foarte mare iar n nu este mult mai mare decât $\sqrt{N}$ (adică $n < 10\sqrt{N}$) atunci inegalitatea de mai înainte devine (aproape) egalitate.

Revenind, probabilitatea pentru a avea coliziuni între cele două liste este aproximativ egală cu

$$P = 1 - e^{-n^2/N}.$$

Observaţia 1. *Dacă n este aproximativ egal cu $3\sqrt{N}$, atunci probabilitatea este aproape $99,98\%$. Dacă suntem "multumiţi" cu o probabilitate de aproximativ $P = 0,64\%$ va fi suficient sa luăm $n = \sqrt{N}$.*

Problema are legătură cu celebrul paradox al zilelor de naştere: Într-o sală se află n persoane. Probabilitatea ca cel puţin două persoane să aibă aceeaşi zi de naştere este

$$P(n) = 1 - e^{n^2/(2\cdot 365)}.$$