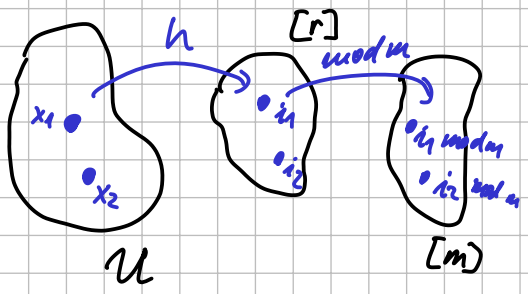
Recall: $\mathcal{L} := \{ h_{a,b} \mid a,b \in [p] \}$ family of functions from $[p] \to [m]$ (this is (2,1)-indep.)

$h_{a,b}(x) := ((ax+b) \bmod p) \bmod m$ (p prime)

$[p] \to [p]$ is (2,1)-indep.

Lemma M (modulo):

Let $\mathcal{H}$ be a $(2,c)$-indep. family from $\mathcal{U} \to [r]$ and $m < r$.
Then $\mathcal{H} \bmod m = \{ h \bmod m \mid h \in \mathcal{H} \}$ (fam. $\mathcal{U} \to [m]$) is $2c$-universal and $(2, 4c)$-indep.



Proof: ① universality

We want: if $x_1 \neq x_2$, then $\Pr_{h \in \mathcal{H} \bmod m}[h(x_1) = h(x_2)] \leq \frac{2c}{m}$.

$\Pr_{h \in \mathcal{H}}[h(x_1) \bmod m = h(x_2) \bmod m]$

$= \Pr_h \Big[\bigvee_{i_1, i_2 \in [r] \text{ s.t. } i_1 \equiv i_2 \pmod m} h(x_1) = i_1 \ \& \ h(x_2) = i_2 \Big]$

$\leq \sum_{i_1 \equiv i_2} \Pr_h[h(x_1) = i_1 \ \& \ h(x_2) = i_2]$ $\leq \frac{c}{r^2}$ by $(2,c)$-indep. of $\mathcal{H}$

$\leq r \cdot \lceil \frac{r}{m} \rceil$

$\leq r \cdot \lceil \frac{r}{m} \rceil \cdot \frac{c}{r^2} \leq \frac{2r \cdot c}{m \cdot r} = \frac{2c}{m}$ ($\lceil \frac{r}{m} \rceil \leq \frac{r+m-1}{m} \leq \frac{2r}{m}$)

② independence

Given $x_1 \neq x_2$, $j_1, j_2 \in [m]$

want $\Pr_{h \in \mathcal{H} \bmod m}[h(x_1) \bmod m = j_1 \ \& \ h(x_2) \bmod m = j_2] \leq \frac{4c}{m^2}$

$\leq \sum_{i_1 \bmod m = j_1, \ i_2 \bmod m = j_2} \Pr[h(x_1) = i_1 \ \& \ h(x_2) = i_2]$ $\leq \frac{c}{r^2}$

$\leq \lceil \frac{r}{m} \rceil^2 \leq \left(\frac{r+m-1}{m} \right)^2 \leq \frac{4r^2}{m^2}$

$\leq \frac{4r^2 c}{m^2 r^2} = \frac{4c}{m^2}$.

---

better analysis:

$\lceil \frac{r}{m} \rceil^k \leq \left(\frac{2r}{m} \right)^k = \frac{2^k r^k}{m^k}$

$\leq \left(\frac{r+m-1}{m} \right)^k \leq \left(\frac{r+m}{m} \right)^k$

for (k,c)-indep.:

$x_1 \dots x_k$ distinct, $j_1 \dots j_k$

$\sum_{i_1 \dots i_k \text{ s.t. } \forall t \ i_t \bmod m = j_t} \Pr_h[\bigwedge_t h(x_t) = i_t] \leq \frac{2^k r^k c}{m^k r^k}$ (⊛) $= \frac{2^k c}{m^k}$ $\leq \frac{c}{r^k}$

$\mathcal{H} \bmod m$ is $(k, 2^k c)$-indep.

(⊛) $\leq \left(\frac{r+m}{m} \right)^k \cdot \frac{c}{r^k} = \frac{c}{m^k} \cdot \left(\frac{r+m}{r} \right)^k \leq \frac{2c}{m^k}$

$\left(1 + \frac{m}{r} \right)^k \leq e^{\frac{mk}{r}} \leq e^{\frac{1}{2}} \leq 2$ ($\leq e^{\frac{m}{r}}$)

if we require $r \geq 2mk$

Tool: $e^x \geq 1 + x$ for all $x \in \mathbb{R}$

Lemma K: Let $\mathcal{H}$ be a (k,c)-indep. system from $\mathcal{U} \to [r]$ and m such that $r \geq 2mk$.
Then $\mathcal{H} \bmod m$ (from $\mathcal{U}$ to $[m]$) is $(k, 2c)$-indep.

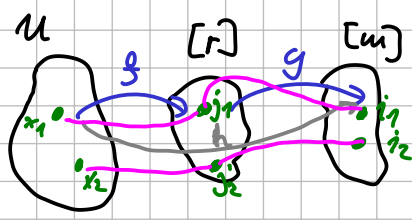
↳ **Applications:** ① $\mathcal{L}$ again ... if $p \geq 4m$, then $\mathcal{L}$ is $(2,2)$-independent.

② $\mathcal{P}$ family of polynomials from $[p] \to [p]$ is $(k,1)$-indep. of degree $< k$

$\mathcal{P}$ mod m is $(k,2)$-indep. for $p \geq 2mk$.

general composition

Lemma 6: Let $\mathcal{F}$ be a c-univ. family from $\mathcal{U}$ to $[r]$
$\mathcal{G}$ be a $(2,d)$-indep. family from $[r]$ to $[m]$.
Then $\mathcal{H} := \mathcal{F} \circ \mathcal{G} = \{ f \circ g \mid f \in \mathcal{F}, g \in \mathcal{G} \}$
is $(2,c')$-indep. for $c' = \left(\frac{cm}{r} + 1\right) d$



Proof: we want: given $x_1, x_2 \in \mathcal{U}$, $x_1 \neq x_2$, $i_1, i_2 \in [m]$

show $\Pr_{h \in \mathcal{H}}[h(x_1) = i_1 \ \& \ h(x_2) = i_2] \leq \frac{c'}{m^2}$

This is $\Pr_{f \in \mathcal{F}, g \in \mathcal{G}}[g(\underbrace{f(x_1)}_{j_1}) = i_1 \ \& \ g(\underbrace{f(x_2)}_{j_2}) = i_2] \leq \frac{d}{m^2}$ ~~wrong~~ since $j_1 \neq j_2$ not guaranteed

Events: C ... collision $f(x_1) = f(x_2) \iff j_1 = j_2$

M ... match $g(j_1) = i_1 \ \& \ g(j_2) = i_2$

Want $\Pr[M] = \underbrace{\Pr[M|C]}_{\leq \frac{d}{m}} \cdot \underbrace{\Pr[C]}_{\leq \frac{c}{r}} + \underbrace{\Pr[M|\neg C]}_{\leq \frac{d}{m^2}} \cdot \underbrace{\Pr[\neg C]}_{\leq 1}$

trivial for $\Pr[C] = 0$ and $\Pr[C] = 1$ does not happen

$\mathcal{G}$ is $(2,d)$-indep. $\Rightarrow$ $(1,d)$-indep.

$$\leq \frac{cdm}{rm^2} + \frac{d}{m^2} = \frac{1}{m^2} \underbrace{d\left(\frac{cm}{r} + 1\right)}_{c' \text{ from the statement}}$$

Hashing of vectors

$\mathcal{U} := \mathbb{Z}_p^d$ families hashing $\mathbb{Z}_p^d \to \mathbb{Z}_p$ $\mathbb{Z}_{257}$

Df: Scalar product family $\mathcal{G}$ from $\mathbb{Z}_p^d$ to $\mathbb{Z}_p$:

$\mathcal{G} := \{ h_a \mid a \in \mathbb{Z}_p^d \}$

$h_a(x) := a \cdot x$ (in $\mathbb{Z}_p$)

$x \cdot y = \sum_i x_i \cdot y_i \pmod p$

Theorem: $\mathcal{G}$ is 1-universal.

Proof: we have x, y $x \neq y$ WLOG $x_1 \neq y_1$

we want $\Pr_a[\underbrace{x \cdot a \overset{\text{mod } p}{\equiv} y \cdot a}_{(x-y)\cdot a \equiv 0}]$

$\sum_i (x_i - y_i) \cdot a_i \equiv 0$

$(x_1 - y_1) \cdot a_1 \equiv -\sum_{i>1} (x_i - y_i) \cdot a_i$

for every $a_2, \dots, a_d$ $\exists!$ a_1 satisfying the equation

$\Rightarrow \Pr_a[\dots] \leq \frac{1}{p}$.

Use Lemma 6 to compose g with $\mathcal{H} \to \mathcal{H}$ $(2,c')$-indep.

g: 1-univ. $\mathcal{H}$: $(2,2)$-indep. for $p \geq 4m$

for $c' = \left(\frac{2m}{p}+1\right)\cdot 2$

$= \left(\frac{m}{p}+1\right)\cdot 2$

$\leq \left(\frac{1}{4}+1\right)\cdot 2 = \frac{5}{2}$

Bloom Filter

probabilistic DS for representing $X \subseteq U$

Insert(x) reliable

Member(x) can give false positive answers

if $x \in X$: always says YES

$x \notin X$: { NO / YES with small probability

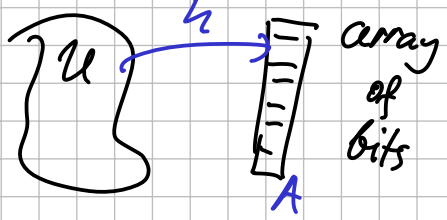
(1) Simple B.F.

$h: U \to [m]$ taken at random from 1-universal family

for every bucket: store just the "non-empty" bit

Insert(x): $A[h(x)] \leftarrow 1$

Member(x): YES iff $A[h(x)] = 1$



Analysis: consider $X = \{x_1 \text{ — } x_n\}$, $y \notin X$

$$\Pr[\text{Member}(y) = \text{YES}] = \Pr\left[\bigvee_i h(x_i) = h(y)\right] \leq \sum_i \underbrace{\Pr[h(x_i) = h(y)]}_{\leq \frac{1}{m} \text{ by 1-univ.}} \leq \frac{n}{m}$$

Consider error prob. ε ... we want $\frac{n}{m} < \varepsilon$ ... $m \geq \frac{1}{\varepsilon}\cdot n$

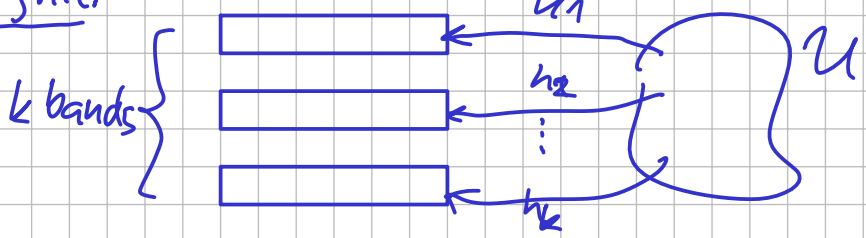
Good: m is indep. of U Bad: Scales badly with ε

e.g. for $n = 10^6$, $\varepsilon = 0.01$ ($k=7$) ... $m \geq 10^8$ = 100 Mb (14 Mb)

$\varepsilon = 0.001$ ... $m \geq 10^9$ = 1 Gb

$\varepsilon = 10^{-6}$ ($k=10$, $k=20$) 20 Mb 40 Mb

(2) Multi-band filter



set $m := 2n$

$\Rightarrow \Pr[\text{FP of 1 band}] \leq \frac{1}{2}$

$\Rightarrow \Pr[\text{all bands FP}] \leq \frac{1}{2^k}$

For given ε: $\frac{1}{2^k} \leq \varepsilon$

$\Rightarrow k := \lceil \log_2 \frac{1}{\varepsilon} \rceil$

$\Rightarrow$ in total $2n \cdot \lceil \log_2 \frac{1}{\varepsilon} \rceil$ bits

k indep. filters

Insert to all

Member answers YES only if all k bands say YES