Ex 6.7. Give an example of a set H of hash functions such that $h(x)$ is equally likely to be any element of $\{0, \ldots, M-1\}$ (H is 1-universal) but H is not 2-universal

Solution:

Let $H = \{h \mid h: \{1, 2, \ldots, m\} \to \{0, 1, 2, \ldots, M-1\}\}$

which means x ~~and~~ or y is a integer in $\{1, \ldots, m\}$.

Let M be a prime greater than m, and we construct a set of hash functions:

$$H' = \{h \mid h(x) = ax \pmod M, a \in \{0, 1, \ldots, M-1\}\}$$

Claim: H' is 1-universal but not 2-universal

Proof. we show that H' is 1-universal first.

we want to show for any x in range $[1, m]$:

$$Pr_{h \in H}[h(x) = w] = Pr_{h \in H}[ax \pmod M = w] = 1/M$$

for a certain x_i, w_i

$$ax_i \pmod M = w_i$$

$$ax_i = w_i + \left\lfloor \frac{ax_i}{M} \right\rfloor \cdot M$$

there's certain a_i satisfies the above equation, so the probability of $h(x_i) = w_i$ is equal to the probability of $a = a_i$, which is $1/M$.

Hence, H' is 1-universal.

Now, we want to prove that H' is not 2-universal.

$$Pr[h(x)=w, h(y)=z] \neq \frac{1}{M^2}$$

$$\Rightarrow Pr[ax=w, ay=z \pmod{M}] \neq \frac{1}{M^2}$$

The probability of ~~sete~~ $h(x)=w$ and $h(y)=z$ is equal to the probability of selected a satisfies $ax \pmod{M} = w$ and $ay \pmod{M} = z$, which is $1/M$ (there's only one value in range $[0, M-1]$ qualifies the requirement above for any certain x, y and w, z.

$$Pr[ax=w, ay=z \pmod{M}] = \frac{1}{M} > \frac{1}{M^2}$$

So H' is not 2-universal.

Ex 6.8

(a) No

Proof:

$$\begin{cases} h_{ab}(x) = u \\ h_{ab}(y) = v \\ h_{ab}(z) = w \end{cases} \Rightarrow \begin{pmatrix} x & 1 \\ y & 1 \\ z & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} u \\ v \\ w \end{pmatrix} \pmod p$$

suppose we solve a, b w.r.t. first two equations, then the last equation must hold for a and b we solved, therefore there's no randomness from the third equation, the probability for equations hold for x, y, z, u, v, w ~~if~~ is the probability of selected a, b satisfies first two equations, which is $1/p^2$.

note that if the third equation doesn't hold ~~for~~ w.r.t a, b solved from the first two, then

$$\Pr(h(x)=u, h(y)=v, h(z)=w) = 0.$$

~~so~~ Above all, $\{h_{ab}(x) = ax+b \mod p \mid 0 \le a, b < p\}$ is not 3-universal

(b) Claim: $\{h_{abc}(x) = ax^2+bx+c \mid 0 \le a, b, c < p\}$ is 3-universal

Proof:

$$\begin{pmatrix} x^2 & x & 1 \\ y^2 & y & 1 \\ z^2 & z & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} u \\ v \\ w \end{pmatrix}$$

since x, y, z are distinct, $\begin{pmatrix} x^2 & x & 1 \\ y^2 & y & 1 \\ z^2 & z & 1 \end{pmatrix}$ is invertibl

$$\Rightarrow \quad \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} x^2 & x & 1 \\ y^2 & y & 1 \\ z^2 & z & 1 \end{pmatrix}^{-1} \begin{pmatrix} u \\ v \\ w \end{pmatrix}$$

there is unique solution for a, b, c,
each ~~one~~ of them was selected from range $[0, p-1]$
so

$$Pr[h(x)=u, h(y)=v, h(z)=w]$$

$$= Pr\left[\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} x^2 & x & 1 \\ y^2 & y & 1 \\ z^2 & z & 1 \end{pmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix} \right]$$

$$= 1/p^3$$

$\therefore \{ h_{abc}(x) = ax^2 + bx + c \mid 0 \le a, b, c < p \}$ is 3-universal

3.5.1

(a) Not a distance measure, because when $x=y$ (where $x\neq 0$), we have $d(x,y) = \max(x,y) \neq 0$. ∎

(b). $d(x,y) = |x-y|$ is a distance measure.

Proof: We prove the (4) axioms:

(1). $d(x,y)=|x-y| \geq 0 \quad \forall x,y$

(2) if $x=y$, then $d(x,y) = |x-y| = |x-x| = 0$

if $d(x,y)=0$, then $d(x,y)=|x-y| = 0$

$\rightarrow x-y=0$

$x=y$

Hence $x=y \iff d(x,y)=0$

(3) $d(x,y) = |x-y| = |y-x| = d(y,x)$

(4) $d(x,y) = |x-y|$

$\cancel{= |x+z|}$

$= |(x-z)+(z-y)|, \quad (\forall z)$

$\leq |x-z| + |z-y|$

$= d(x,z) + d(z,y)$

Hence $d(x,y) \leq d(x,z) + d(z,y)$

$d(x,y)$ satisfies all (4) axioms, hence it is a distance measure. ∎

(c). Not a distance measure, because when $x=y$ (where $x\neq 0$) we have $d(x,y) = x+y = 2x \neq 0$.

3.5.4

(a) $d = 1 - \frac{3}{5} = \frac{2}{5}$ (b) $d = 1 - 0 = 1$

3.5.7

(a) $\cancel{a}b\cancel{c}def \xrightarrow{+a.c} bd\underline{a}ef\underline{c}$ $d = 4$

(b) $ab\cancel{c}\cancel{c}d\cancel{a}\cancel{b}c \xrightarrow{+c.a.b} a\underline{c}bdc\underline{a}\underline{b}$ $d = 7$

(c) $\cancel{a}b\cancel{c}\cancel{d}ef \xrightarrow{+a.d.c.} b\underline{a}e\underline{d}f\underline{c}$ $d = 6$

$\boxed{3.6.1.}$ (a) Let F be a (d_1, d_2, p_1, p_2) - LSH family.

then, after a 2-way AND, we will get F': (d_1, d_2, p_1^2, p_2^2) - LSH

applying 3-way OR on F', we get F'': $(d_1, d_2, 1-(1-p_1^2)^3, 1-(1-p_2^2)^3)$ - LSH

Hence we get a $(d_1, d_2, 1-(1-p_1^2)^3, 1-(1-p_2^2)^3)$ - LSH.

(b). Let F be a (d_1, d_2, p_1, p_2) - LSH family

applying a 3-OR, we get F' : $(d_1, d_2, 1-(1-p_1)^3, (1-(1-p_2)^3))$ - LSH

applying a 2-AND on F', we get F'': $(d_1, d_2, [1-(1-p_1)^3]^2, [1-(1-p_2)^3]^2)$ - LSH

Hence we get a $(d_1, d_2, [1-(1-p_1)^3]^2, [1-(1-p_2)^3]^2)$ - LSH.

$\boxed{4.3.3}$ let false prob. rate be $f(k) = (1-e^{-\frac{m}{n}k})^k$

taking log of both sides, $\ln f(k) = k \ln(1-e^{-\frac{m}{n}k})$

$$\frac{d}{dk}[\ln f(k)] = \frac{d}{dk}\left[k \ln(1-e^{-\frac{m}{n}k})\right]$$

$$\frac{f'(k)}{f(k)} = k\left[\frac{1}{1-e^{-\frac{m}{n}k}} \cdot \frac{m}{n} e^{-\frac{m}{n}k}\right] + \ln(1-e^{-\frac{m}{n}k})$$

~~So $f'(k) =$~~

set $f'(k) = 0$, so that

$$(1) \longrightarrow 0 = \frac{km}{n} e^{-\frac{m}{n}k} + (1-e^{-\frac{m}{n}k}) \cdot \ln(1-e^{-\frac{m}{n}k})$$

let $b = e^{\frac{-m}{n}k}$,

then $\ln b = \frac{-m}{n}k$

$$k = \frac{-n}{m}\ln b \quad \text{②}$$

Substituting ② into ①, $0 = -\ln b \cdot b + (1-b)\ln(1-b)$

$$b\ln b = (1-b)\ln(1-b)$$

So $b = 1-b$

$$2b = 1$$

$$b = \frac{1}{2}$$

Hence $k = \frac{-n}{m}\ln\frac{1}{2}$

$\boxed{k = \frac{n}{m}\ln 2}$ is the value that minimises $f(k)$, or the false prob. rate.

4.3.2

The probability of a false positive is $(1 - e^{-km/n})^k$, the same as having k hash functions and 1 array because the odds of collisions are the same and therefore the odds of a false positive is the same.