Alex Quispe

Problem Set 1

Problem 1

a) Claim: let $\hat{n} = \frac{1}{\delta}((1+\delta)^{X_n} - 1)$ then $\mathbb{E}[\hat{n}] = n$

Proof by induction:

Base case $(n=0)$

$$\mathbb{E}[\hat{n}] = \mathbb{E}[(1+\delta)^{X_n} - 1] = (1+\delta)^0 - 1 = 0 = n$$

Inductive case $(n = n+1)$

Assume $\mathbb{E}[\hat{n}] = n$

$$\mathbb{E}\left[\frac{1}{\delta}((1+\delta)^{X_{n+1}} - 1)\right] = \mathop{\mathbb{E}}_{X_0 \cdots X_n}\left[\mathop{\mathbb{E}}_{X_{n+1}}\left[\frac{1}{\delta}((1+\delta)^{X_{n+1}} - 1)\right]\right]$$

$$= \mathop{\mathbb{E}}_{X_0 \cdots X_n}\left[\frac{1}{(1+\delta)^{X_n}}\left(\frac{1}{\delta}((1+\delta)^{X_n+1} - 1)\right) + \left(1 - \frac{1}{(1+\delta)^{X_n}}\right)\left(\frac{1}{\delta}((1+\delta)^{X_n} - 1)\right)\right]$$

$$= \mathop{\mathbb{E}}_{X_0 \cdots X_n}\left[\frac{1}{\delta}\left((1+\delta) - \frac{1}{(1+\delta)^{X_n}}\right) + \frac{1}{\delta}\left((1+\delta)^{X_n} - 1 - 1 + \frac{1}{(1+\delta)^{X_n}}\right)\right]$$

$$= \mathop{\mathbb{E}}_{X_0 \cdots X_n}\left[\frac{1}{\delta}((1+\delta) + (1+\delta)^{X_n} - 2)\right]$$

$$= \mathop{\mathbb{E}}_{X_0 \cdots X_n}\left[\frac{1}{\delta}(\delta + (1+\delta)^{X_n} - 1)\right]$$

$$= \mathop{\mathbb{E}}_{X_0 \cdots X_n}\left[1 + \frac{1}{\delta}((1+\delta)^{X_n} - 1)\right]$$

$$= 1 + n \quad \boxtimes$$

Problem Set 1

Problem 1

b) Claim 1: $\mathbb{V}[\hat{n}] \leq \frac{\delta}{2} n^2$

Proof:

$$\mathbb{V}[\hat{n}] = \mathbb{E}_{x_0 \cdots x_n}[\hat{n}^2] - \mathbb{E}[\hat{n}]^2$$

$$= \mathbb{E}_{x_0 \cdots x_n}\left[\frac{1}{\delta^2}\left((1+\delta)^{x_n}-1\right)^2\right] - n^2$$

$$= \frac{1}{\delta^2} \mathbb{E}_{x_0 \cdots x_n}\left[(1+\delta)^{2x_n} + 1 - 2(1+\delta)^{x_n}\right] - n^2$$

$$\mathbb{E}_{x_0 \cdots x_n}\left[(1+\delta)^{x_n}\right] = \mathbb{E}_{x_0 \cdots x_{n-1}}\left[\frac{1}{(1+\delta)^{x_{n-1}}}(1+\delta)^{x_{n-1}+1} + \left(1 - \frac{1}{(1+\delta)^{x_{n-1}}}\right)(1+\delta)^{x_{n-1}}\right]$$

$$= \mathbb{E}_{x_0 \cdots x_{n-1}}\left[1+\delta-1+(1+\delta)^{x_{n-1}}\right]$$ → recursion where $\mathbb{E}_{x_0}[(1+\delta)^{x_0}] = 1$

$$= \delta n + 1$$

$$\mathbb{E}_{x_0 \cdots x_n}\left[(1+\delta)^{2x_n}\right] = \mathbb{E}\left[\frac{1}{(1+\delta)^{x_{n-1}}}(1+\delta)^{2(x_{n-1}+1)} + \left(1 - \frac{1}{(1+\delta)^{x_{n-1}}}\right)(1+\delta)^{2x_{n-1}}\right]$$

$$= \mathbb{E}_{x_0 \cdots x_{n-1}}\left[(1+\delta)^2(1+\delta)^{x_{n-1}} + (1+\delta)^{2x_{n-1}} - (1+\delta)^{x_{n-1}}\right]$$

$$= \mathbb{E}_{x_0 \cdots x_{n-1}}\left[(\delta^2+2\delta)(1+\delta)^{x_{n-1}} + (1+\delta)^{2x_{n-1}}\right]$$

$$= (\delta^2+2\delta)(\delta(n-1)+1) + \mathbb{E}\left[(1+\delta)^{2x_{n-1}}\right]$$ → recursion where $\mathbb{E}[(1+\delta)^{2x_0}] = 1$

$$= (\delta^2+2\delta)\left[\sum_{i=1}^{n-1}[\delta(n-i)+1]\right] + 1$$

from prev. calculation — from final term

$$= (\delta^2+2\delta)\left(\frac{\delta n(n-1)}{2} + (n-1) + 1\right)$$

$$= \frac{\delta^2(\delta+2)n(n-1)}{2} + (\delta^2+2\delta)n$$

$$\mathbb{V}[\hat{n}] = \frac{1}{\delta^2}\left(\frac{\delta^2(\delta+2)(n-1)n}{2} + (\delta^2+2\delta)n + 1 - 2(\delta n+1)\right) - n^2$$

$$= \frac{n}{2}(\delta n - \delta + 2n - 2) + n + \frac{2n}{\delta} + \frac{1}{\delta^2} - \frac{2n}{\delta} - \frac{2}{\delta^2} - n^2$$

$$= \frac{\delta n^2}{2} - \frac{\delta n}{2} + n^2 - n + n - \frac{1}{\delta^2} - n^2$$

$$= \frac{\delta}{2}n^2 - \frac{\delta}{2}n - \frac{1}{\delta^2}$$

$$\leq \frac{\delta}{2}n^2$$

□

Problem set 1

Problem 1

b) Claim 2: $\mathbb{P}[|\hat{n}-n| \leq \varepsilon n] \geq .9$ as long as $\delta \leq \frac{\varepsilon^2}{5}$

Proof:

From Chebyshev's Inequality,

$$\mathbb{P}[|\hat{n}-n| \geq \lambda] \leq \frac{\mathbb{V}[\hat{n}]}{\lambda^2}$$

$$\mathbb{P}[|\hat{n}-n| \leq \lambda] \geq 1-\frac{\mathbb{V}[\hat{n}]}{\lambda^2}$$

Let $\lambda^2 = 10\mathbb{V}[\hat{n}]$

$$\mathbb{P}[|\hat{n}-n| \leq \sqrt{10\mathbb{V}[\hat{n}]}] \geq .9$$

$$\sqrt{10\mathbb{V}[\hat{n}]} \leq \sqrt{10\tfrac{\delta}{2}n^2} = \sqrt{5\delta}\, n$$

$$\varepsilon = \sqrt{5d}$$

$$\delta = \frac{\varepsilon^2}{5} = \boxed{.2\varepsilon^2}$$

c) Claim: space required is $O(\lg\lg n + \lg\frac{1}{\varepsilon})$

Proof: space $\leq O(\lg X_n)$

$$\frac{(1+\delta)^{X_n}-1}{\delta} \leq (1+\varepsilon)n$$

$$X_n \leq \lg_{(1+\delta)} (1+\varepsilon)n\delta+1$$

so space $\leq O(\lg\lg_{(1+\delta)}(1+\varepsilon)n\delta+1)$

$$= O\left(\lg\left(\frac{\lg(n\delta+\varepsilon n\delta+1)}{\lg(1+\delta)}\right)\right)$$

$$= O(\lg\lg n - \lg\lg(1+\varepsilon^2))$$

$$= O(\lg\lg n - \lg\varepsilon^2) \quad \text{from Taylor Series}$$

$$= O(\lg\lg n + \lg\tfrac{1}{\varepsilon})$$

$\blacksquare$

We know $\varepsilon \in (0,1)$
ε^2 is small and close to 0. The Taylor Approximation is below.

$$\lg(1+x) \approx \lg(1) + \frac{1}{1+0}(x) - \frac{1}{2}\frac{1}{(1+0)^2}(x^2)\ldots$$

$$= O(x - \tfrac{x^2}{2}+\ldots)$$

$$= O(x)$$

Problem Set 1

Problem 2

a) Claim: $\forall x \neq y \quad x, y \in U \quad \mathbb{P}[h(x) = h(y)] = \frac{1}{n}$

Proof: Let $x, y \in U$ s.t. $x \neq y$

Consider 4 cases

1. $x_H = y_H$, $x_L = y_L$
2. $x_H \neq y_H$, $x_L = y_L$
3. $x_H = y_H$, $x_L \neq y_L$
4. $x_H \neq y_H$, $x_L \neq y_L$

Case 1:

Impossible since $x \neq y$ by definition

Case 2. $\mathbb{P}[h(x) = h(y)] = \mathbb{P}[H[x_H] \oplus L[x_L] = H[y_H] \oplus L[y_L]]$

$= \mathbb{P}_H[H[x_H] = H[y_H]]$ since we know $L[x_L] = L[y_L]$

there are $n^{2^{u/2}}$ possible $H[\cdot]$ arrays

$n^{2^{u/2}-1}$ possible where $H[x_H] = H[y_H]$

$= \frac{n^{u-1}}{n^u} = \frac{1}{n}$

Case 3.

Same as case 2 by symmetry

Case 4. $\mathbb{P}[H[x_H] \oplus L[x_L] = H[y_H] \oplus L[y_L]]$

$= \mathbb{P}[H[x_H] \oplus H[y_H] = L[x_L] \oplus L[y_L]]$ bc. of how XOR works

Consider the $\mathbb{P}[H[x_H] \oplus H[y_H] = a]$ where $a \in \{0, \ldots, n-1\}$

This is essentially the same as case 2. since for any $H[x_H]$ and a there is only one $H[y_H]$ that works

The same reasoning applies to $L[\cdot]$ by symmetry.

$$\mathbb{P}[H[x_H] \oplus H[y_H] = L[x_L] \oplus L[y_L]]$$

$$= \sum_{a=1}^{n-1} \mathbb{P}_H[H[x_H] \oplus H[y_H] = a] \cdot \mathbb{P}[L[x_L] \oplus L[y_L] = a]$$

$$= \sum_{a=0}^{n-1} \frac{1}{n} \cdot \frac{1}{n} = \frac{n}{n^2} = \frac{1}{n}$$

In every case, the claim holds true $\blacksquare$

Problem Set 1

Problem 2

b) Claim: $\forall x \neq y \quad x, y \in U \quad \mathbb{P}[h(x) = h(y)] = \frac{1}{n}$

Proof: let $b \in \{2, \dots, u\}$ be the number of $H[\cdot]$ arrays
this is also the number of pieces an input to $h()$
will be split into.

Base case: $b = 2$.

This is the hash function from part a which we have proven is universal.

Inductive case: $b = b+1$

Assume a hash function with b arrays is universal

$$\mathbb{P}[h(x) = h(y)] = \mathbb{P}[H_1[x_1] \oplus \dots \oplus H_b[x_b] \oplus H_{b+1}[x_{b+1}] = H_1[y_1] \oplus \dots \oplus H_b[y_b] \oplus H_{b+1}[y_{b+1}]]$$

let $x_h = x[0:b]$ bits, $y_h = y[0:b]$ bits

Case 1. $x_h = y_h$, $x_{b+1} = y_{b+1}$: Impossible by definition

Case 2. $x_h = y_h$, $x_{b+1} \neq y_{b+1}$

$$\mathbb{P}[h(x) = h(y)] = \mathbb{P}_{H_{b+1}}[H_{b+1}[x_{b+1}] = H_{b+1}[y_{b+1}]]$$

number of possible $H_{b+1} = n^{2^{\frac{u}{b}}}$

number of possible where $H_{b+1}[x_{b+1}] = H_{b+1}[x_{b+1}] = n^{2^{\frac{u}{b}}-1}$

$$\mathbb{P}[h(x) = h(y)] = \frac{n^{2^{\frac{u}{b}}-1}}{n^{2^{\frac{u}{b}}}} = \frac{1}{n}$$

Case 3. $x_h \neq y_h$, $x_{b+1} = y_{b+1}$

$\mathbb{P}[h(x) = h(y)] = \frac{1}{n}$ by induction hypothesis

Problem Set 1

Problem 2

b) Case 4: $X_n \neq Y_n$, $X_{b+1} \neq Y_{b+1}$

$$\mathbb{P}[h(x) = h(y)] = \mathbb{P}[H_1[X_1] \oplus \dots \oplus H_b[X_b] \oplus H_1[Y_1] \oplus \dots \oplus H_b[Y_b] = H_{b+1}[X_{b+1}] \oplus H_{b+1}[Y_{b+1}]]$$

$$= \sum_{a=0}^{n-1} \mathbb{P}[H_1[Y_1] \oplus \dots \oplus H_b[Y_b] = a] \cdot \mathbb{P}[H_{b+1}[X_{b+1}] \oplus H_{b+1}[Y_{b+1}]]$$

$$= \sum_{a=0}^{n-1} \frac{1}{n} \cdot \frac{1}{n} = \frac{n}{n^2} = \frac{1}{n}$$

This is because the first term is a restatement of our induction hypothesis. Checking equivalence is just the special case where $a=0$. The second term is $\frac{1}{n}$ following the same logic in case 2.

The claim holds in every case, so by induction it will be true when $b=u$ $\boxtimes$

The space to specify the hash function is

$$O(u \cdot 2 \cdot \lg n) = O(u \lg n)$$

from u arrays with two numbers taking $\lg n$ bits each.

Problem Set 1

Problem 3

a) Let X_i be the number of elements mapped to bucket i

Claim: $\mathbb{P}[X_1 \geq w] \geq \frac{1}{\sqrt{n}}$ where $w = c\frac{\lg n}{\lg\lg n}, c>0$

Proof: $\mathbb{P}[X_1 \geq w] = \sum_{i=w}^{n} \mathbb{P}[X_1 = i] = \sum_{i=w}^{n} \binom{n}{i}\left(\frac{1}{n}\right)^i\left(\frac{n-1}{n}\right)^{n-i}$

$\geq \binom{n}{w}\left(\frac{1}{n}\right)^w\left(\frac{n-1}{n}\right)^{n-w}$ just the largest term

$\geq \left(\frac{n}{w}\right)^w\left(\frac{1}{n}\right)^w\left(\frac{n-1}{n}\right)^{n-w}$ from Stirlings inequality

$\geq \left(\frac{1}{w}\right)^w\left(\frac{n-1}{n}\right)^{n-w}$

$\left(\frac{n-1}{n}\right)^{n-w} \geq \left(1-\frac{1}{n}\right)^n \rightarrow e^{-1}$ as $n \to \infty$ by definition of e

when $n \geq 2$ $\left(\frac{n-1}{n}\right)^n \geq \frac{1}{4}$ and is upper-bounded by $\frac{1}{e}$

Therefore $\mathbb{P}[X_1 \geq w] \geq \frac{1}{4}\left(\frac{1}{w}\right)^w = \frac{1}{4}w^{-w}$

$\geq 2^{\lg\left(\frac{1}{4}w^{-w}\right)} = 2^{-w\lg\left(\frac{1}{4}w\right)}$

Let c be a small constant such as $\frac{1}{1000}$

$\geq 2^{-\frac{1}{1000}\frac{\lg n}{\lg\lg n}\left(\lg\lg n - \lg 4000\lg\lg n\right)}$

$\geq 2^{-\frac{1}{1000}\lg n}$ by dropping the positive exponent term

$\geq n^{-1/1000} \geq n^{-\frac{1}{2}}$

$\square$

Problem Set 1

Problem 3

b) Claim: Let X be the number of heavy buckets

$$\mathbb{E}[X] = \Omega(\sqrt{n})$$

Proof:
$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} \mathbb{1}[X_i \geq w]\right]$$
$$= \sum_{i=1}^{n} \mathbb{P}[X_i \geq w]$$
$$\geq \sum_{i=1}^{n} \frac{1}{\sqrt{n}}$$
$$\geq \frac{n}{\sqrt{n}} = \sqrt{n}$$

Therefore, X is lower bounded by $\sqrt{n}$ and

$\mathbb{E}[X] = \Omega(\sqrt{n})$

$\blacksquare$