

# CSE 521 HW4

Rohan Mukherjee

November 16, 2023

1. (a) Notice for  $\alpha > 0$  and some fixed  $y \in \mathbb{R}^n$ ,

$$\mathbb{P}\left[\sum_{i=1}^n y_i \alpha Z_i \geq x\right] = \mathbb{P}\left[\sum_{i=1}^n y_i Z_i \geq \frac{x}{\alpha}\right] = \mathbb{P}\left[\|y\|_p Z \geq \frac{x}{\alpha}\right] = \mathbb{P}\left[\|y\|_p \alpha Z \geq x\right]$$

- (b) Since the pdf is continuous and bounded, we see that  $\int_{1-\varepsilon}^1 p(x)dx = p(c_\varepsilon) \cdot \varepsilon$  for some suitable choice of  $c_\varepsilon \in (1 - \varepsilon, 1)$ . The pdf is also symmetric, since  $-Z$  has the same distribution as  $\| - 1\|_p Z = Z$ . Thus

$$\mathbb{P}[-1 + \varepsilon < Z < 1 - \varepsilon] = 2\left(\int_0^1 p(x)dx - \int_{1-\varepsilon}^1 p(x)dx\right) = \frac{1}{2} - 2p(c_\varepsilon)\varepsilon \leq \frac{1}{2} - c\varepsilon$$

Similarly  $\int_1^{1+\varepsilon} p(x)dx = p(d_\varepsilon) \cdot \varepsilon$  for a suitable choice of  $c_\varepsilon$ . Thus,

$$\mathbb{P}[-1 - \varepsilon < Z < 1 + \varepsilon] = 2\left(\int_0^1 p(x)dx + \int_1^{1+\varepsilon} p(x)dx\right) = \frac{1}{2} + 2p(d_\varepsilon)\varepsilon \geq \frac{1}{2} + c\varepsilon$$

Since  $p(1) > 0$  we can find a sufficiently small neighborhood so that  $p(x) > p(1)/2 = c$  for all  $x$  close to 1.

- (c) Notice that  $y_i = (Px)_i = \sum_j x_j Z_{ij}$  just by matrix multiplication. Define  $a_i = |y_i| < \|x\|_p(1 - \varepsilon)$  and  $b_i = |y_i| > \|x\|_p(1 + \varepsilon)$ . Clearly  $\mathbb{E}a_i = \mathbb{P}[|y_i| < \|x\|_p(1 - \varepsilon)] = \mathbb{P}[|Z| < 1 - \varepsilon] \leq \frac{1}{2} - c\varepsilon$ , and  $\mathbb{E}b_i \geq \frac{1}{2} + c\varepsilon$ . Let  $S_\ell = \sum_{i=1}^\ell a_i$ . Notice that

$$\mathbb{P}[S_\ell \geq \ell/2] = \mathbb{P}[S_\ell - \mathbb{E}S_\ell \geq \ell/2 - \mathbb{E}S_\ell] \leq \mathbb{P}[|S_\ell - \mathbb{E}S_\ell| \geq 2C\varepsilon\ell] \leq 2e^{-2C^2\varepsilon^2\ell}$$

Similarly if we let  $T_\ell = \sum_{i=1}^\ell b_i$  we get

$$\mathbb{P}[T_\ell \geq \ell/2] \leq 2e^{-2C^2\varepsilon^2\ell}$$

Now we want

$$\mathbb{P}[S_\ell \leq \ell/2, T_\ell \leq \ell/2] \geq 1 - 4e^{-2C^2\varepsilon^2\ell} \geq 1 - \delta$$

It suffices to choose  $\ell = \frac{\log(4/\delta)}{C^2\varepsilon^2} = O\left(\frac{\log 1/\delta}{\varepsilon^2}\right)$ . Notice that if both  $S_\ell \leq \ell/2$  and  $T_\ell \leq \ell/2$ , and  $\ell$  is odd, then we certainly have that the median is in  $[\|x\|_p(1 - \varepsilon), \|x\|_p(1 + \varepsilon)]$ .

(d) Notice that, if  $Z \sim \text{Cauchy}(0, 1)$ ,

$$\mathbb{P}[1 + \varepsilon < Z < 1 - \varepsilon] = \frac{1}{\pi} \arctan(1 - \varepsilon) - \frac{1}{\pi} \arctan(1 + \varepsilon) \approx \frac{1}{2\pi} 2\varepsilon = \frac{\varepsilon}{\pi}$$

By the mean value theorem. So we must use  $\ell = \pi^2 \log(4/0.01)/\varepsilon^2$  to get a  $1 \pm \varepsilon$  approximation with probability 99/100. I shall choose  $\varepsilon = 0.01$  as well, so we need  $\ell \approx 591334$ . Here is my code:

---

```
import numpy as np
from numpy import random as rand

x = [int(x) for x in open("p4.in", "r")]
P = rand.standard_cauchy((591334, len(x)))
y = np.absolute(P.dot(x))
print(np.median(y))
```

---

I get a value of 4788.158226259304.

2. (a) For  $a = 0$ ,

$$\mathbb{P}\left[\left\lfloor \frac{-s}{w} \right\rfloor = \left\lfloor \frac{b-s}{w} \right\rfloor\right] = \mathbb{P}\left[-1 = \left\lfloor \frac{b-s}{w} \right\rfloor\right]$$

This happens precisely when  $-w < b - s < 0$ , or  $b < s < b + w$ .  $s$  can only be in this range if  $|b| < w$ , else the probability is 0. If  $b$  is positive then we need  $b < s < w$ , which happens with probability  $(w - b)/w = 1 - b/w$ . Similarly if it is negative we need  $1 + b/w$ . So the probability is  $\max\{0, 1 - \frac{|b|}{w}\}$ . Similarly, if

$$\left\lfloor \frac{a-s}{w} \right\rfloor = \left\lfloor \frac{b-s}{w} \right\rfloor$$

Then,

$$\begin{aligned} \frac{a-s}{w} - 1 &\leq \frac{b-s}{w} \leq \frac{a-s}{w} + 1 \\ \iff a-s-w &\leq b-s \leq a-s+w \\ \iff a-w &\leq b \leq a+w \\ \iff |b-a| &\leq w \end{aligned}$$

So if  $|b-a| > w$  the probability is already 0. Now WLOG  $b > a$ . Since  $\lfloor x \rfloor = \lfloor y \rfloor$  iff  $\lfloor x-k \rfloor = \lfloor y-k \rfloor$  (for integer  $k$ ), and since  $|b-a| < 1$ , we can assume that  $0 < a, b < 2$ . Now either  $(a, b) \subset (0, 1)$  or  $1 \in (a, b)$ . In the first case, the only way the floors won't be equal is if  $s \in (a, b)$ , since then  $\lfloor a-s \rfloor = -1$  while  $\lfloor b-s \rfloor = 0$  (Else they are either both 0 or both -1). In the second case, the only way they won't be equal is if  $s$  takes away too little from  $b$ , or too much from  $a$ . That is, if  $s \in (0, b-1)$ , or  $s \in (a, 1)$ . The probability this bad event happens is just  $b-a$ . Similarly, in the general case, for  $k$  an

integer,  $\lfloor x/w \rfloor = \lfloor y/w \rfloor$  iff  $\lfloor x/w + kw/w \rfloor = \lfloor y/w + kw/w \rfloor$ . We can again assume that  $0 < a, b < 2w$ . The exact same logic as before works again but just having dividing by  $w$  everywhere (indeed: we are just stretching the number line by a factor of  $1/w$ ), so we can conclude the probability is  $\max\{0, 1 - \frac{|a-b|}{w}\}$ .

(b) We see that

$$\begin{aligned}\mathbb{P}[h(p) = h(q)] &= \mathbb{P}\left[\bigcap_{i=1}^d \left\lfloor \frac{p_i - s_i}{w} \right\rfloor = \left\lfloor \frac{q_i - s_i}{w} \right\rfloor\right] = \prod_{i=1}^d \left(1 - \frac{\alpha_i}{w}\right) \\ &\approx \exp\left(-\frac{1}{w} \sum_{i=1}^d \alpha_i\right) = \exp\left(-\frac{1}{w} \|p - q\|_1\right)\end{aligned}$$

So if  $d(p, q) = \|p - q\|_1 \leq r$ , we have

$$\mathbb{P}[h(p) = h(q)] \geq \exp\left(-\frac{r}{w}\right)$$

And when  $d(p, q) \geq c \cdot r$ , we have

$$\mathbb{P}[h(p) = h(q)] \leq \exp\left(-\frac{cr}{w}\right)$$

So this hash function is  $(r, c \cdot r, e^{-r/w}, e^{-cr/w})$ -sensitive.