An Intensive Introduction to Cryptography: Notes

Rushil Surti

July 14, 2024

0 Mathematical Background

Exercise 5 (Random Hash Function)**.** Let $H : \{1, \ldots, n\} \to \{1, \ldots, m\}$ represent a hash function, with each entry for the function chosen randomly (this is equivalent to uniformly choosing over all m^n functions). We say that there is a *collision* if for some $i < j$, $H(i) = H(j)$. Let $X_{i,j} := \mathbf{1}_{H(i)=H(j)}$.

1. For every $i < j$, compute $\mathbf{E}[X_{i,j}]$.
2. Let $Y := \sum_{i<j} X_{i,j}$, representing the total collisions. Compute $\mathbf{E}[Y]$.
3. Prove that if $m > 1000 \cdot n^2$, the probability that H is injective is at least 0.9.
4. Prove that if $m < n^2/1000$, the probability that H is injective is at most 0.1.

Solution. We shall proceed with each part as follows:

1. By symmetry, it stands that each $\mathbf{E}[X_{i,j}]$ is the same. We see that $\mathbf{E}[X_{i,j}] = m \cdot 1/m^2 = 1/m$.
2. We know that
$$\mathbf{E}[Y] = \sum_{i=2}^{n} \sum_{j=1}^{i-1} \frac{1}{m},$$
which tells us that $\mathbf{E}[Y] = n(n-1)/(2m)$.
3. The probability that H is injective is given by
$$\mathbf{P}(H \text{ is injective}) = \frac{\binom{m}{n} n!}{m^n} = \prod_{k=0}^{n-1} \left(1 - \frac{k}{m}\right).$$
This function is strictly increasing with respect to m so that if $m > 1000n^2$,
$$\prod_{k=0}^{n-1} \left(1 - \frac{k}{m}\right) > \prod_{k=0}^{n-1} \left(1 - \frac{k}{1000n^2}\right) > \prod_{k=0}^{n-1} \left(1 - \frac{1}{1000n}\right) = \left(1 - \frac{1}{1000n}\right)^n.$$
By Bernoulli's inequality, we have that
$$\left(1 - \frac{1}{1000n}\right)^n \geq 1 - \frac{1}{1000} = 0.999 > 0.9,$$
which shows the desired quality.

4. Observe that, in order for H to be injective, we must have $n \leq m < n^2/1000$, which tells us that at the very least $n > 1000$. By AM-GM, we have that

$$\begin{aligned}
\mathbf{P}(H \text{ is injective}) &= \prod_{k=0}^{n-1}\left(1 - \frac{k}{m}\right) \leq \left(\frac{1}{n}\sum_{k=0}^{n-1}\left(1 - \frac{k}{m}\right)\right)^n \\
&= \left(1 - \frac{n-1}{2m}\right)^n \leq \left(\frac{1}{2} - \frac{1}{2n}\right)^n \\
&= \frac{1}{2^n}\left(1 - \frac{1}{n}\right)^n \\
&\leq \frac{1}{2^{1000}} < 0.1.
\end{aligned}$$

Exercise 12. The *Shannon entropy* of a distribution μ formed over a finite set S is given by

$$H(\mu) := \sum_{x \in S} \mu(x) \log_2(1/\mu(x)).$$

We wish to prove the intution that, in the amortized sense, $H(\mu)$ bits are needed to encode members of the distribution (not quite sure what this is referring to exactly).

1. Prove that for every injective function $F : S^* \to \{0,1\}^*$,

$$\mathbf{E}_{x\sim\mu} \left|F(x)\right| = \sum_{x \in S} \left|F(x)\right| \mu(x) \geq H(\mu).$$

2. Prove that for every ε, there is some n and an injective function $F : S^n \to \{0,1\}^*$ such that (note: I'm not sure this is what the problem is exactly asking. Perhaps I'm being a bit smooth brain, but the notation isn't exactly clear to me)

$$\mathbf{E}_{\mathbf{x}\sim\mu^n} \left|F(\mathbf{x})\right| = \sum_{\mathbf{x} \in S^n} \left|F(\mathbf{x})\right| \mu(x_1)\mu(x_2)\cdots\mu(x_n) \leq n(k + \varepsilon).$$

Solution. As per the MSE question I asked on this, it's likely that there is a typo in the original source for the question, which would have been very nice to know from the start, but oh well.

1.