Exercise sheet - Week 16 (Balls in bins)

EXERCISE 1 -Fingerprints [1]

Let $\mathcal{W}$ be a (large) set and $S \subset \mathcal{W}$ and set $m = |S|$. We want to design an algorithm which answers the following query : does $w \in S$? In order to save space, we consider a randomized algorithm and allow it to do occasional mistakes.

Let $b \geq 1$ and $h : \mathcal{W} \to \{0,1\}^b$ a hash function which maps every word onto a b-bit string. We make very strong assumptions on h :

— for all $w \in W$, $h(w)$ is uniform in $\{0,1\}^b$;

— for all $w \neq w' \in W$, $h(w)$ and $h(w')$ are independent.

Consider now the following algorithm :

```
Precomputation: compute and store the set h(S)
input: w in W
if h(w) is in h(S):
   return 'MAYBE' (w belongs probably to S)
else:
   return 'NO' (w does not belong to S)
```

The only possible mistake is when the algorithm returns `MAYBE` while $w \notin S$, we would like to control the probability error.

Let $\varepsilon > 0$, find b as small as possible such that for all $w \notin S$,

$$\mathbb{P}\,(\text{the algorithm returns 'MAYBE'}) \leq \varepsilon.$$

(Hint : you may use the inequalities $(1-x)^m \geq \exp(-2mx)$ and $\exp(-x) \geq 1 - x$ both valid for all $0 \leq x \leq 1/2$.)

(For the remaining exercises we take the same model as in the Lecture : throw n balls uniformly at random and independently into r bins.)

EXERCISE 2 -Multiple collisions : the easy way

We say that there is a b-collision if at least b balls land in the same bin. Using the 1st moment method, prove that if $n = o(r^{1-1/b})$ then

$$\mathbb{P}(\text{at least one } b\text{-collision at time } n) \to 0.$$

EXERCISE 3 -Two balls in each bin

In the Lecture we proved that if $C = \min\{n \geq 1; \text{at least one ball in each bin at time } n\}$, then for all $\varepsilon > 0$ then $\mathbb{P}(C \geq r\log(r)(1+\varepsilon)) \to 0$.

Let $\tilde{C} = \min\{n \geq 1; \text{at least two balls in each bin at time } n\}$, clearly $\tilde{C} > C$. However, prove that the same estimate as above still holds :

$$\mathbb{P}(\tilde{C} \geq r\log(r)(1+\varepsilon)) \to 0.$$

EXERCISE 4 -Multiple collisions : the constant

Prove that

$$\int_0^{+\infty} \exp(-s^b/b!)\mathrm{d}s = \sqrt[b]{b!}\,\Gamma(1+1/b).$$

(Recall $\Gamma(z) = \int_0^{+\infty} t^{z-1}e^{-t}\mathrm{d}t$.)

1. (Source : Sec.5.5.2 in : M.Mitzenmacher, E.Upfal. Probability and computing : randomized algorithms and probabilistic analysis. Cambridge University Press (1995).)