Compression

By: Maya Kalai


?
HAHAHA
HAHAHA
19


45
?
HA
HA
HA
HA
HA
HA


We've all been here so long that we assigned numbers to our jokes so we wouldn't have to say them each time


32
UM, NO.
UR A LOSER


Why did no one laugh? Is joke 32 not funny?
It’s all in the delivery

The Goal

Get the computer to compress jokes efficiently!!!

Note: each joke has some probability

Let's say we have some file $f = \{0,1\}^*$. Our compression algorithm is $c : \{0,1\}^* \to \mathbb{N}$, and our decompression algorithm is $d : \mathbb{N} \to \{0,1\}^*$. Where $d(c(f)) = f$.

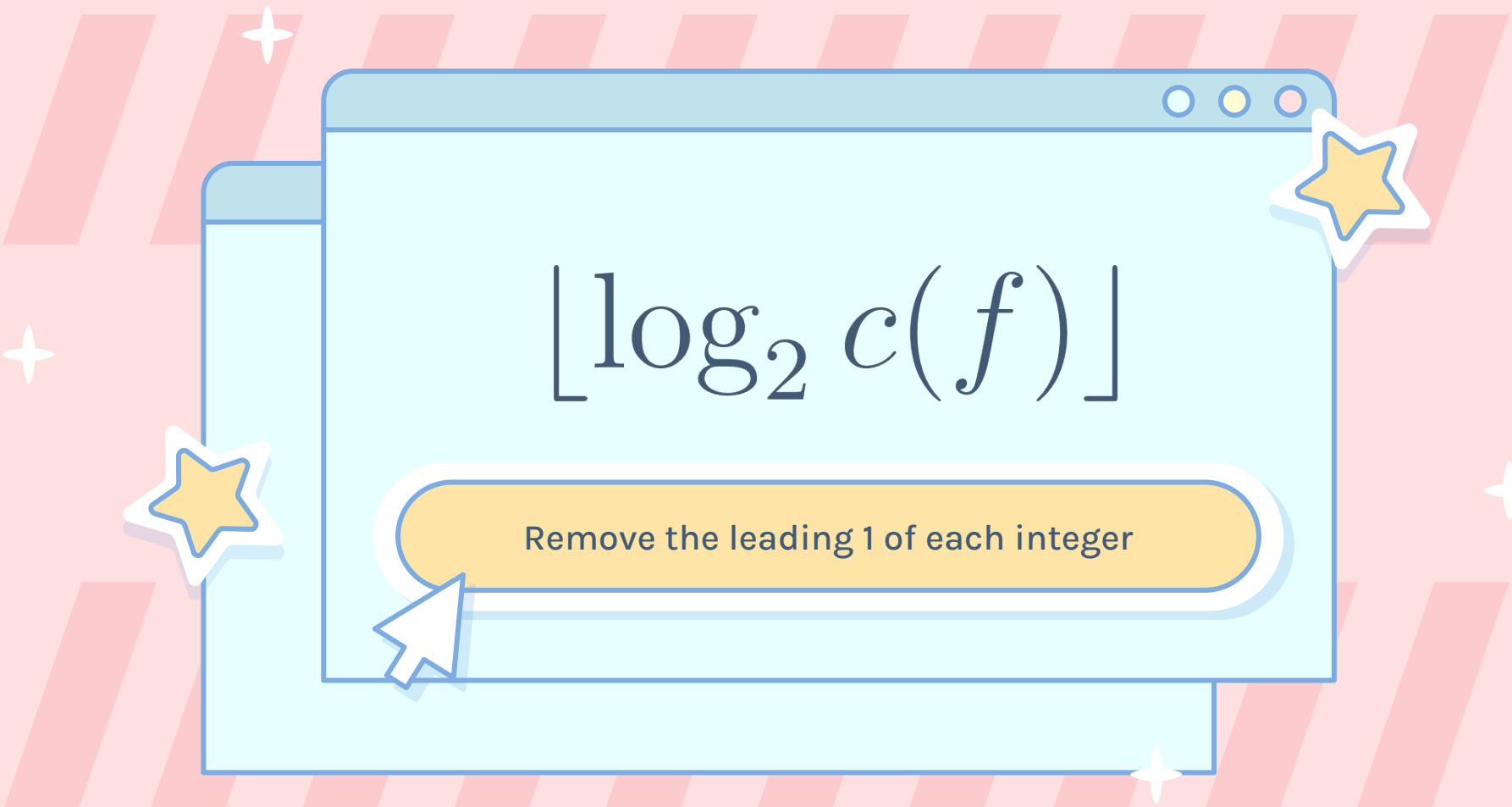
f	$p(f)$	$c(f)$
“What does a baby computer call his father? Data”	0.5	1
“Where do pirates get their hooks? Second hand stores”	0.3	2
“Wanna hear a joke about paper? Never mind. It's tearable.”	0.2	3
“lhftfrduu;o jygjytr fdtrdjtrd jnklijp”	0	4

Let's say it takes x bits to represent c(f) in binary, then $2^{x-1} \leq c(f) < 2^x$ or

$x - 1 \leq \log_2 c(f) < x$ so $x - 1 = \lfloor \log_2 c(f) \rfloor$ or

$x = \lfloor \log_2 c(f) \rfloor + 1$

Converting an integer into binary


$\lfloor \log_2 c(f) \rfloor$
Remove the leading 1 of each integer

Upper bound for the Expected Value

$$\sum_{f \in \{0,1\}^*} p(f) * \lfloor \log_2 c(f) \rfloor$$

$$c(f) \leq \frac{1}{p(f)} \forall f$$

Shannon Entropy

Entropy

The entropy is the average amount of information needed to store a file, Claude Elwood Shannon invented Shannon Entropy which essentially finds how much data is required to store a file.


QUESTIONS?

Thank You!!!