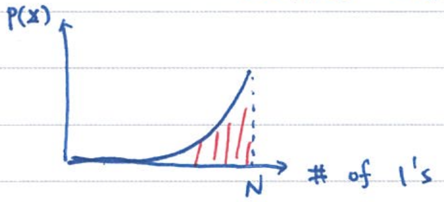
* Implication about part (a) proof

$$\frac{1}{N} H_\delta(X) \le H + \epsilon$$

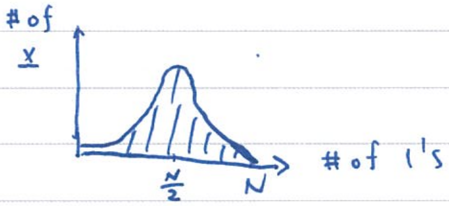
sufficient to only consider T_ϵ to get $> 1 - \delta$ prob.

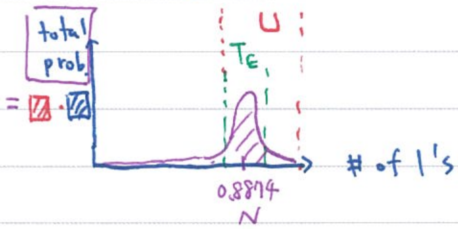
* case study $X = \{0, 1\}$

$p_0 = 0.1126$, $p_1 = 0.8874$



for $x \in X^N$





* Given ϵ, δ, if $N \geq N_0$

can do block coding of T_ϵ to achieve $\frac{1}{N} H_\delta \le H + \epsilon$ bits

physicall

$$\boxed{\lceil \log T_\epsilon \rceil \text{ bits to encode every } x \in T_\epsilon \text{ correctly}}$$

$x \to \nu(x)$

* block coding & "δ" not always useful in practice — computation/storage, partial decoding

next question: can we do $\boxed{\text{symbol}}$ coding to achieve $\approx H$?

$$x \to \nu(x) \in \underbrace{\{0,1\}^+}_{B}$$

* $X \xrightarrow{\nu} Y$ needs $H_0(X)$ if only alternatives $\in Y$ used to represent X

$H_\delta(X)$ w/ prob. $> 1-\delta$

$X \xrightarrow{\nu} B^+$ uses some alternatives $\in B^+$ w/ different length to represent X

* will start w/ error-free encoder / decoder first

if X, N known (see notes: can be unknown as well)

error-free coding $\iff$ $\nu(x_1 x_2 \cdots x_N)$ one-to-one

$\nu(x_1 x_2 \cdots x_N) = \nu(x_1)\cdot\nu(x_2) \cdots \nu(x_N)$

error-free coding $\equiv$ uniquely decodable (UD)

avg. block code length per symbol $= \frac{1}{N} H_\delta(X^N) \approx H(X)$

*

X	P	$\nu(X)$
a_1	P_1	$\lvert\nu(a_1)\rvert = n_1$
a_2	P_2	$\lvert\nu(a_2)\rvert = n_2$
$\vdots$	$\vdots$	
a_K	P_K	$\lvert\nu(a_K)\rvert = n_K$

$H(X) = \sum P_k \log \frac{1}{P_k}$

avg. symbol code length $= \sum P_K n_K$

next: relate (optimal) UD code length to $H(X)$

* UD $\longrightarrow$ $\sum_{k=1}^{K} 2^{-n_k} \le 1$

$$\sum_{k=1}^{K} 2^{-n_k} = \sum_{i=1}^{\boxed{L}} \underbrace{\left|\{k : n_k = i\}\right|}_{\le 2^{n_k}} 2^{-n_k} \le \boxed{L}$$

UD for $N=1$ (Hello! Pigeon Hole Principle)

UD for $N=2$

$$\underbrace{\sum_{i=1}^{2\cdot\boxed{L}} \underbrace{\left|\{(k,l) : n_k + n_l = i\}\right|}_{\le 2^{n_k+n_l}} 2^{-(n_k+n_l)}}_{= \sum_{k=1}^{K} 2^{-n_k} \cdot \sum_{l=1}^{K} 2^{-n_l}} \le 2\cdot\boxed{L}$$

$\vdots$

UD $\underbrace{\left(\sum_{k=1}^{K} 2^{-n_k}\right)^N}_{\text{exponential}} \le \underbrace{N\cdot\boxed{L}}_{\text{linear}}$ for all $N \Rightarrow (\quad) \le 1$

called Kraft's Ineq.