Note Title 25-Jun-19

$(X_1, X_2, \ldots, X_n) \in \{0,1\}^n$

$X_i = 1$ w.p. p

0 $1-p$ $\quad p \in (0,1)$

$B^{(n)} \subseteq \{0,1\}^n$

① $P_e^{(n)}(B^{(n)}) = Pr((B^{(n)})^c) = 1 - Pr(B^{(n)})$ prob. of error (we would like this value to be small)

② $\text{Ambiguity}(B^{(n)}) = \dfrac{|B^{(n)}|}{|\{0,1\}^n|}$ (we would like ambiguity to be small)

Suppose that we require $P_e^{(n)}(B^{(n)}) \leq \varepsilon$ for some fixed $\varepsilon > 0$, $\varepsilon < 1$

Now we will ask, how small can the ambiguity be?

Find a set $\boxed{B_\varepsilon^{(n)} = \arg\min_{B^{(n)}:\, P_e^{(n)}(B^{(n)}) \leq \varepsilon} \text{Ambiguity } B^{(n)}}$

Question: How does $\frac{|B_\varepsilon^{(n)}|}{|\{0,1\}^n|}$ behave?

$$\frac{|B_\varepsilon^{(n+1)}|}{|\{0,1\}^{n+1}|} \leq \frac{|B_\varepsilon^{(n)}|}{|\{0,1\}^n|}$$

Proof: Consider the set $B_\varepsilon^{(n)} \times \{0,1\} = B^{(n+1)}$

$$P_e^{(n)}(B^{(n+1)}) = \sum_{x^{n+1} \notin B^{(n+1)}} p(x^n) = \sum_{x^n \notin B_\varepsilon^{(n)}} [p(x^n)p(0) + p(x^n)p(1)]$$

$$= \sum_{x^n \notin B_\varepsilon^{(n)}} p(x^n)\,[p(0)+p(1)] = P_e^{(n)}\left((B_\varepsilon^{(n)})^c\right) \leq \varepsilon$$

$$\frac{|B_\varepsilon^{(n+1)}|}{|\{0,1\}^{n+1}|} \leq \frac{|B^{(n+1)}|}{|\{0,1\}^{n+1}|} = \frac{|B_\varepsilon^{(n)}| \times 2}{2^{n+1}} = \frac{|B_\varepsilon^{(n)}|}{2^n} = \text{Ambiguity}(B_\varepsilon^{(n)})$$

$\square$

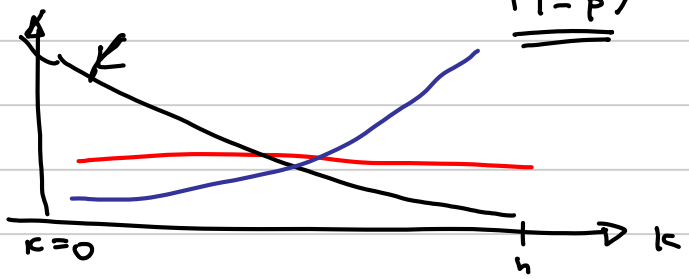
Ambiguity $(B_\varepsilon^{(n)})$ is non-increasing in n.

Ambiguity $(B_\varepsilon^{(n)}) \geq 0$

Let's try to reason it out.

Strings x^n with k ones and $(n-k)$ 0's all have probability

$$(1-p)^{n-k} p^k = (1-p)^n \left(\frac{p}{1-p}\right)^k$$

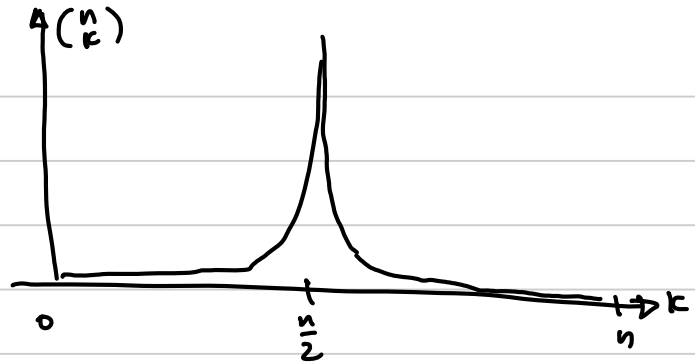


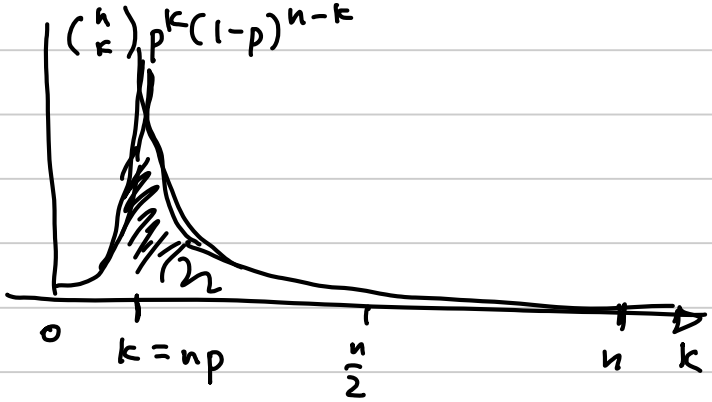
$p < \frac{1}{2}$

$p > \frac{1}{2}$

~~$p = \frac{1}{2}$~~

There are $\binom{n}{k}$ strings with $(n-k)$ 0's and k 1's.





WLLN

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to E[X] \quad \text{in prob}$$

$$= p$$

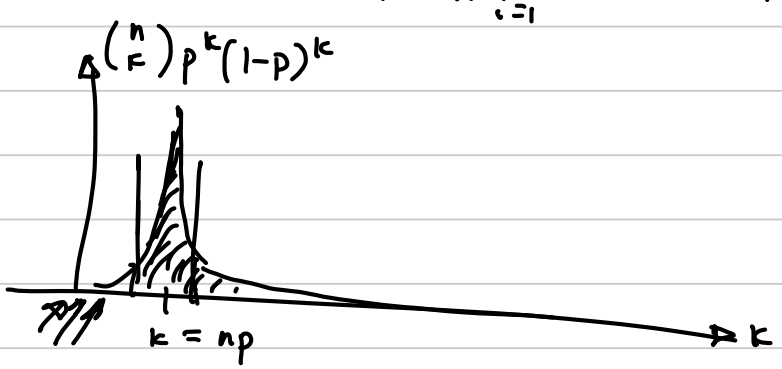
$$\sum_{i=1}^{n} X_i = \#\ 1\text{s in string}$$

$$\to np$$

Weak Law of Large Numbers (WLLN)

Given $X_1, X_2, \ldots \sim iid\ p(x)$ on $\mathcal{X}$

$$\Pr\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - E[X] \right| \geq \sqrt{a} \right) \leq \frac{\mathrm{var}(X)}{a \cdot n} \to 0 \quad \text{as } n \to \infty$$



Given that with high probability we expect to see a string with np '1's an $n(1-p)$ '0's, it should also be true that

$$p(X^n) \approx p^{np}(1-p)^{n(1-p)}$$

Is that true??

$$p(X^n) = \prod_{i=1}^{n} p(x_i) = 2^{\log \prod_{i=1}^{n} p(x_i)} = 2^{\sum_{i=1}^{n} \log p(x_i)} = 2^{n \cdot \frac{1}{n} \sum_{i=1}^{n} \log p(x_i)}$$

$$= 2^{-n \sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{p(x_i)}}$$

$$p(x^n) = 2^{-n \cdot \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{1}{p(x_i)}\right)}$$

$$\frac{1}{n}\sum_{i=1}^{n} \log \frac{1}{p(x_i)} \longrightarrow E\left[\log \frac{1}{p(x_i)}\right] \text{ in probability}$$

$$H(x) \triangleq E\left[\log \frac{1}{p(x)}\right] = \text{"entropy"}$$

$$p(x^n) = 2^{-n \cdot \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{1}{p(x_i)}\right)} \longrightarrow 2^{-nH(x)} \text{ in probability}$$

$$\Pr\left(2^{-n(H(x)+\varepsilon)} \leq p(x^n) \leq 2^{-n(H(x)-\varepsilon)} \right) \longrightarrow 1 \text{ as } n \to \infty.$$

$$p^{np}(1-p)^{n(1-p)} = 2^{\log\left[p^{np}(1-p)^{n(1-p)}\right]}$$

$$= 2^{np\log p + n(1-p)\log(1-p)}$$

$$= 2^{-n\left[p\log\frac{1}{p} + (1-p)\log\frac{1}{1-p}\right]} = 2^{-nH(x)}$$

$A_\varepsilon^{(n)} = \{x^n : 2^{-n(H(x)+\varepsilon)} \leq p(x^n) \leq 2^{-n(H(x)-\varepsilon)}\}$ "typical set"

$\Pr(A_\varepsilon^{(n)}) \longrightarrow 1$ as $n \to \infty$ by WLLN

As long as our $P_e^{(n)}$ constraint was greater than 0, this set will meet that error constraint for n sufficiently large.

$$P_e^{(n)}(B_\varepsilon^{(n)}) \leq \varepsilon$$

Let's bound $|A_\varepsilon^{(n)}|$

$$1 \geq \Pr(A_\varepsilon^{(n)}) = \sum_{x^n \in A_\varepsilon^{(n)}} p(x^n) \geq |A_\varepsilon^{(n)}| \, 2^{-n(H(x)+\varepsilon)} \Rightarrow |A_\varepsilon^{(n)}| \leq \frac{1}{2^{-n(H(x)+\varepsilon)}} = 2^{n(H(x)+\varepsilon)}$$

$$1-\varepsilon \leq \Pr(A_\varepsilon^{(n)}) \leq |A_\varepsilon^{(n)}| 2^{-n(H(x)-\varepsilon)} \Rightarrow |A_\varepsilon^{(n)}| \geq (1-\varepsilon)\, 2^{n(H(x)-\varepsilon)}$$

for n suff. large

for n suff. large

$$\text{Ambiguity}\left(A_\varepsilon^{(n)}\right) \leq \frac{2^{n(H(X)+\varepsilon)}}{|\mathcal{X}|^n} = \frac{2^{n(H(X)+\varepsilon)}}{2^{n\log|\mathcal{X}|}} = 2^{-n(\log|\mathcal{X}|-H(X))}$$

$H(X) \leq \log|\mathcal{X}|$ with equality iff $X \sim \text{Unif}(\mathcal{X})$.

$\therefore$ $\text{Ambiguity}\left(A_\varepsilon^{(n)}\right) \longrightarrow 0$ as $n \to \infty$ provided $X \not\sim \text{Unif}(\mathcal{X})$.