Recap

$K(x) = \min_{p: \mathcal{U}(p)=x} l(p)$ $\quad P_{\mathcal{U}}(x) = \sum_{p:\mathcal{U}(p)=x} 2^{-l(p)}$

$K(x|l(x)) = \min_{p: \mathcal{U}(p, l(x)) = x} l(p)$

Universal gambling

$\underbrace{S(x)}_{\text{wealth}} = \underbrace{2^{l(x)} \cdot b(x)}$, $\sum b(x) = 1$

modified $S(x) = \sum_{x' \sqsupseteq x} 2^{l(x')} \cdot b(x')$

universal gambling :

$b(x) = 2^{-K(x)}$

c.f. $W^* + H = \log m$. (Chapter 6)

Thm) $\log S(x) + K(x) \geq l(x)$

pf) $S(x) = \sum_{x' \sqsupseteq x} 2^{l(x')} \cdot b(x') \geq 2^{l(x) - K(x)}$

For inf. seq., $S_n = S(x^n) \geq 2^{n-c}$

Occam's razor

Q. Sun has risen on day 1 to n. Probability of that sun raises at day $n+1$?

A. (Laplace) Assume i.i.d. Bernoulli(θ).

$$P(X_{n+1}=1 \mid X_n=1, \cdots, X_1=1)$$
$$= \frac{P(X_{n+1}=1, \cdots, X_1=1)}{P(X_n=1, \cdots, X_1=1)}$$
$$= \frac{\int_0^1 \theta^{n+1} d\theta}{\int_0^1 \theta^n d\theta}$$
$$= \frac{n+1}{n+2}$$

$$\sum_y p(1^n y) \approx p(1^\infty) = c > 0$$
$$\sum_y p(1^n 0 y) \approx p(1^n 0) \approx 2^{-\log n} \approx \frac{1}{n}$$
$$p(0 \mid 1^n) = \frac{p(1^n 0)}{p(1^n 0) + p(1^\infty)} \approx \frac{1}{cn+1}$$

"It is similar to $\frac{1}{n+2}$ of Laplace" (?)

K and P

Thm. $\exists$ a constant c s.t. $\forall x$

$$2^{-K(x)} \leq P_{\mathcal{U}}(x) \leq c\, 2^{-K(x)}$$

That is,

$K(x)$ is almost equal to the $\underbrace{\text{ideal codeword length}}_{\log \frac{1}{p(x)}}$ w.r.t. $P(x)$

Recall.

$H(X)$ is an average of the <u>ideal codeword length</u>

$2^{-K(x)} \le P_U(x) \le c\,2^{-K(x)}$

$\downarrow$

trivial $\quad P_U(x) = \sum_{p:U(p)=x} 2^{-l(p)}$

WTS :

$$K(x) \le \log \frac{1}{P_U(x)} + C$$

If we can compute $P_U(x)$, then we can make the following program.

" Print $\underbrace{x}_{\text{a string}}$ whose ideal wordcode is y w.r.t. P_U".

y ↳ length is $\log \frac{1}{P_U(x)}$

---

stage	Λ	0	1	00	01	...
1	1					
2	2	2				
3	3	3	3			
		$\ddots$				

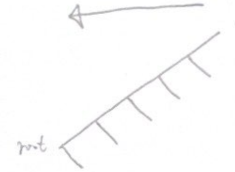
$\Downarrow$

for each program and output (p_k, x_k), compute $n_k = \lceil \log \frac{1}{P_U(x_k)} \rceil$

We assign (p_k, x_k, n_k) at $\underbrace{\text{tree}}_{\text{depth } n_k+1}$ only if n_k "jumps". $\rightarrow$ we assign at most one nodes for the same depth for each x.

We can do this process if the Kraft inequ. holds. $\rightarrow \sum_i 2^{-l_i} \le 1$

order of assignment



$$\begin{aligned}
\sum_{k=1} 2^{-(n_k+1)} &= \sum_x \sum_{k:x_k=x} 2^{-(n_k+1)} \\
&= \sum_x 2^{-1} \sum_{k:x_k=x} 2^{-n_k} \\
&\le \sum_x 2^{-1} \cdot \left(2^{\lfloor \log P_U(x) \rfloor} + 2^{\lfloor \log P_U(x) \rfloor - 1} + \cdots \right) \\
&= \sum_x 2^{-1} \cdot 2^{\lfloor \log P_U(x) \rfloor} \cdot 2 \\
&\le \sum_x P_U(x) \\
&\le 1
\end{aligned}$$

For each string x, there is a corresponding node at depth $\le \lceil \log \frac{1}{P_U(x)} \rceil + 1$

" Print a string of a node $\tilde{p}$ "

$\tilde{p}$ ↳ length is $\le \lceil \log \frac{1}{P(x)} \rceil + 1$

Kolmogorov Sufficient Statistic

Def) The Kolmogorov structure func.

$$K_k(x^n|n) = \min_{\substack{p:\,l(p)\le k\\ \mathcal{U}(p,n)=S\\ x^n\in S\subseteq\{0,1\}^n}} \log_2|S|$$

Def) For a given small constant c, let k^* be the least k s.t.

$$K_k(x^{(n)}|n) + k \le K(x^{(n)}|n) + c$$

Let S^{**} be the corresponding set and let p^{**} be the program that prints out the indicator func. of S^{**}
Then, p^{**} is a Kolmogorov minimal sufficient statistic for $x^{(n)}$.

$$\underbrace{K_k(x^{(n)}|n)}_{\text{index of } x^{(n)} \text{ in } S} + \underbrace{k}_{\text{desc. of } p^{**}} = K(x^{(n)}|n)$$

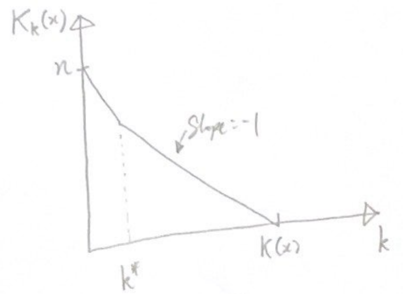
Note that a statistic T is sufficient for a parameter θ.

$$\theta \to T(X) \to X$$

forms a Markov Chain.

Example: $X_i \sim \text{normal}(\mu, \sigma^2)$, $i=1,\cdots,n$.

$$(\mu,\sigma^2) \to (\sum x, \sum x^2) \to X_1,\cdots,X_n$$



Minimum Description Length principle

Minimize:

$$\underbrace{K(p)}_{\text{model}} + \underbrace{\log\frac{1}{p(X_1,X_2,\cdots,X_n)}}_{\text{description}}$$