1) $X = \{H, T\}$ $\quad x_1, x_2, \dots, x_N \in \{H, T\}$ random variables that are i.i.d.

$(N = 100)$

i.i.d. = independent identically distributed

↳ according to Bernoulli($\cancel{\alpha}\ w$)

$$p(x_n = H) = \overset{w}{\cancel{\alpha}} \quad \Rightarrow \quad p(x_n = T) = \overset{1-w}{1-\alpha}$$

---

2) Maximum Likelihood Estimation (MLE) principle:

Data X whose prob. depends on some params. W.

MLE $\quad \hat{W} = \arg\max\limits_{W} p(X \mid W)$ $\qquad L(W) = p(X \mid W)$.

$$= \arg\max_{W} L(W; X)$$

↳ the likelihood $L(W)$

Going back to 1): We have $N \rightarrow 100$ coin tosses, out of which $K \rightarrow 40$ are H.

$$\underbrace{\begin{matrix} x_1 & x_2 & x_3 & \dots & x_N \\ H & T & T & \dots & H \end{matrix}}_{K \text{ Heads, } N-K \text{ tails}}$$

A, B indep $\Rightarrow p(A, B) = p(A)\,p(B)$.

MLE: $\hat{W} = \arg\max\limits_{W} L(W; x_1, x_2, \dots, x_N) = \arg\max\limits_{W} p(x_1, x_2, \dots, x_N \mid W) =$

$$\hat{w} = \arg\max_w \prod_{i=1}^{N} p(x_i | w) = \arg\max_w \underbrace{\ln \prod_{i=1}^{N} p(x_i|w)}_{\text{log-likelihood}}$$

$$\arg\max_W f(w) = \arg\max_w \ln f(w)$$

$$= \arg\max_w \underbrace{\sum_{i=1}^{N} \ln p(x_i | w)}$$

$$\sum_{\{i : x_i = H\}} \ln \underbrace{p(x_i|w)}_{w} + \sum_{\{j : x_j = T\}} \ln \underbrace{p(x_j|w)}_{1-w}$$

$$\hat{w} = \arg\max_w \underbrace{k \ln w + (N-k) \ln(1-w)}_{J(w)}$$

$$\frac{\partial J}{\partial w} = 0 \Rightarrow \frac{k}{w} + (N-k)\frac{-1}{1-w} = 0 \quad \Big| \; w(1-w)$$

$$k(1-w) - (N-k)\cdot w = 0$$

$$w(-k - (N-k)) + k = 0$$

$$w(-N) + k = 0 \Rightarrow k = wN \Rightarrow \boxed{w = \frac{k}{N} = 0.4}$$