# PARAMETER ESTIMATION

ESTIMATE PARAMETER VALUES THAT EXPLAINS DATA.
CALCULATE PROBABILITY OF UNSEEN DATA USING PARAMETERS

$$P(Q|x) = \frac{P(x|Q)\,P(Q)}{P(x)}$$

$$L(Q|x) = P(\underset{\uparrow}{x}|Q) - \boxed{\text{LIKELIHOOD OF PARAMETER 'Q'.}}$$

$\boxed{\text{NOT LIKELIHOOD OF 'X'}}$

x is FIXED (INPUT)
Q is VARIABLE.

WHY TO LOOK AT LIKELYHOOD RATHER THAN $P(Q|x)$

- $P(x)$ is FIXED.
- $P(Q)$ is CONSTANT ACROSS All $Q$. AND We DON'T HAVE INFO ABOUT $P(Q)$
- Goal: To Find the "$Q$" which probability is more given "x". TRANFORMS INTO $\underline{P(x|Q)}$.

ASSUMPTION: "x" is IID. (independent Identical DISTRIBUTION)

$$P(X|Q) = \prod_{i=1}^{n} P(x_i|Q)$$

LOG-LIKELIHOOD: $L(Q|X) = \log(P(X|Q)) = \log\left(\prod_i P(x_i|Q)\right)$

$$\hat{Q} = \underset{Q}{\operatorname{argmax}} \sum_{i=1}^{n} \log P(x_i|Q)$$

$$\Rightarrow P(\hat{x}|X) = P(\text{New data}|\text{Training DATA}) = P(\text{New DATA}|\hat{Q})$$

$$P(C=c|f) = f^{c}(1-f)^{1-c} \rightarrow \text{Bernoulli DISTRIBUTION}$$

$c=0 \quad f$
$c=1 \quad 1-f$

For 1 Example.

$$L\cancel{(P(f|c))} = l(P(c=c_i|f)) = \log\left(f^{c}(1-f)^{1-c}\right)$$
$$= c\log f + (1-c)\log(1-f)$$

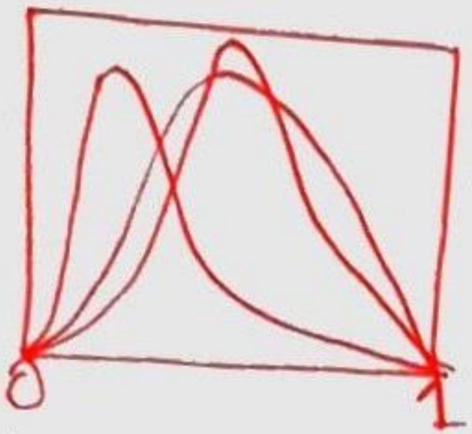
$$L(P(f|c)) = \sum_{i=1}^{n} l(P(c=c_i|f)) = \sum c_i \log f + (1-c_i)\log(1-f)$$

# PRIORS AND MAP

$P(\theta)$: Priors

$$\hat{\theta}_{MAP} = \underset{\hat{\theta}}{\operatorname{argmax}}\, p(\theta|x) = \underset{\theta}{\operatorname{argmax}}\, p(x|\theta)\,p(\theta)$$

$$= \underset{\theta}{\operatorname{argmax}} \left[\log p(x|\theta) + \log p(\theta)\right]$$




→ β-distribution

Since, we don't know anything about prior, so we assume it follows a "β distribution".

- WRONG PRIORS NEEDS LARGE AMOUNT OF DATA.

$$\hat{\xi}_{MAP} = \frac{n^{(HEAD)} + \alpha - 1}{n^{(HEAD)} + n^{(TAIL)} + \alpha + \beta - 2}$$

Probability of Head

BAYESIAN ESTIMATION