Bayesian Interpretation

$$J(w) = \frac{1}{2}\sum_{i=1}^{N}(t_i - y_i)^2$$

optimization problem

$$\arg\min_{w} J(w) = \arg\min_{w} \frac{1}{2}\sum_{i=1}^{N}(t_i - y_i)^2$$

$$= \arg\max_{w} \left\{ -\frac{1}{2}\sum_{i=1}^{N}(t_i - y_i)^2 \right\}$$

Exp(·) is monotonic

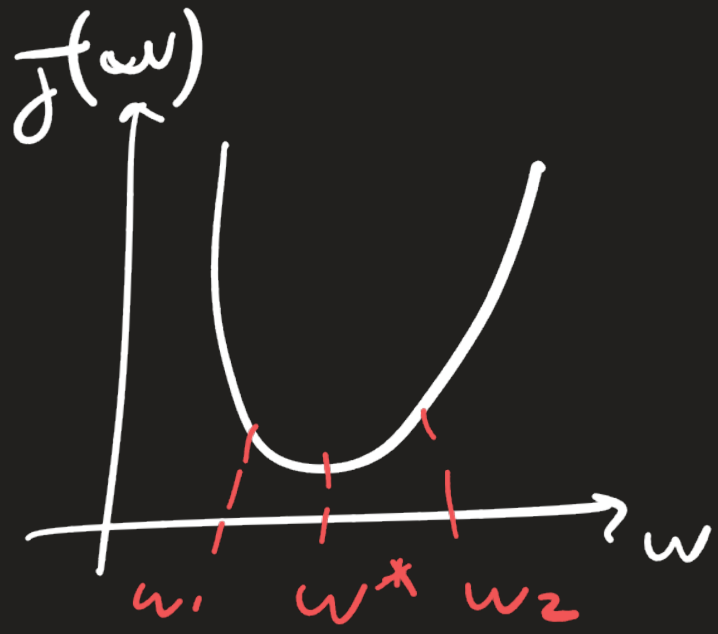
$$= \arg\max_{w} \mathrm{Exp}\left\{ -\frac{1}{2}\sum_{i=1}^{N}(t_i - y_i)^2 \right\}$$

reminder

$$e^{a+b+c} = e^{a} \cdot e^{b} \cdot e^{c}$$

$$= \arg\max_{w} \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} \underbrace{\mathrm{Exp}\left\{ -\frac{1}{2}\left(\overbrace{t_i - y_i}^{\varepsilon_i}\right)^2 \right\}}_{t_i \sim G(y_i, 1)}$$

$$= \arg\max_{w} \prod_{i=1}^{N} \underbrace{\frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\varepsilon_i^2 \right\}}_{\varepsilon_i \sim G(0, 1)}$$



Univariate Gaussian Distribution

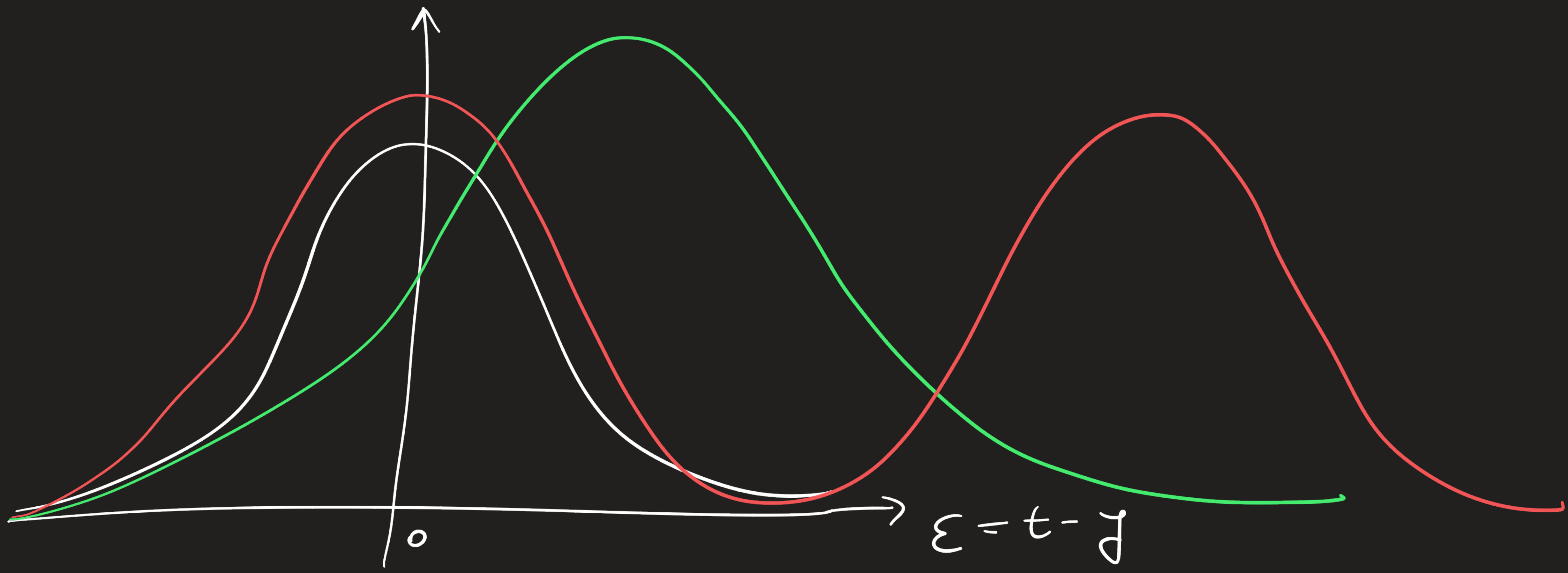
Random Variable : X

mean

variance

$$X \sim G(\mu, \sigma^2)$$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \operatorname{Exp}\left\{ -\frac{1}{2 \cdot \sigma^2} (x-\mu)^2 \right\}$$


0
ε = t - y

MSE + Ridge Regularizer

$$J(w) = \frac{1}{2}\sum_{i=1}^{N}(t_i - y_i)^2 + \frac{\lambda}{2}\sum_{j=0}^{M} w_j^2$$

$$\boxed{\underset{w}{\arg\min}\, J(w)} = \underset{w}{\arg\max} - J(w)$$

$$= \underset{w}{\arg\max}\, \exp(-J(w))$$

$$= \underset{w}{\arg\max}\, \exp\left\{\underbrace{-\frac{1}{2}\sum_{i=1}^{N}(t_i - y_i)^2}\underbrace{-\frac{\lambda}{2}\sum_{j=0}^{M} w_j^2}\right\}$$

$$= \underset{w}{\arg\max}\, \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}(t_i - y_i)^2\right\}\exp\left\{-\frac{\lambda}{2}\sum_{j=0}^{M} w_j^2\right\}$$

$$= \underset{w}{\arg\max}\left[\prod_{i=1}^{N}\underbrace{\underbrace{\exp\left\{-\frac{1}{2}(t_i - y_i)^2\right\}}_{t_i \sim G(y_i, 1)}}_{P(t_i|w)}\right]\left[\prod_{j=0}^{M}\underbrace{\underbrace{\exp\left\{-\frac{\lambda}{2}w_j^2\right\}}_{w_j \sim G(0, 1/\lambda)}}_{P(w_j)}\right]$$

$$= \underset{w}{\arg\max}\prod_{i=1}^{N}P(t_i|w)\prod_{j=0}^{M}P(w_j)$$

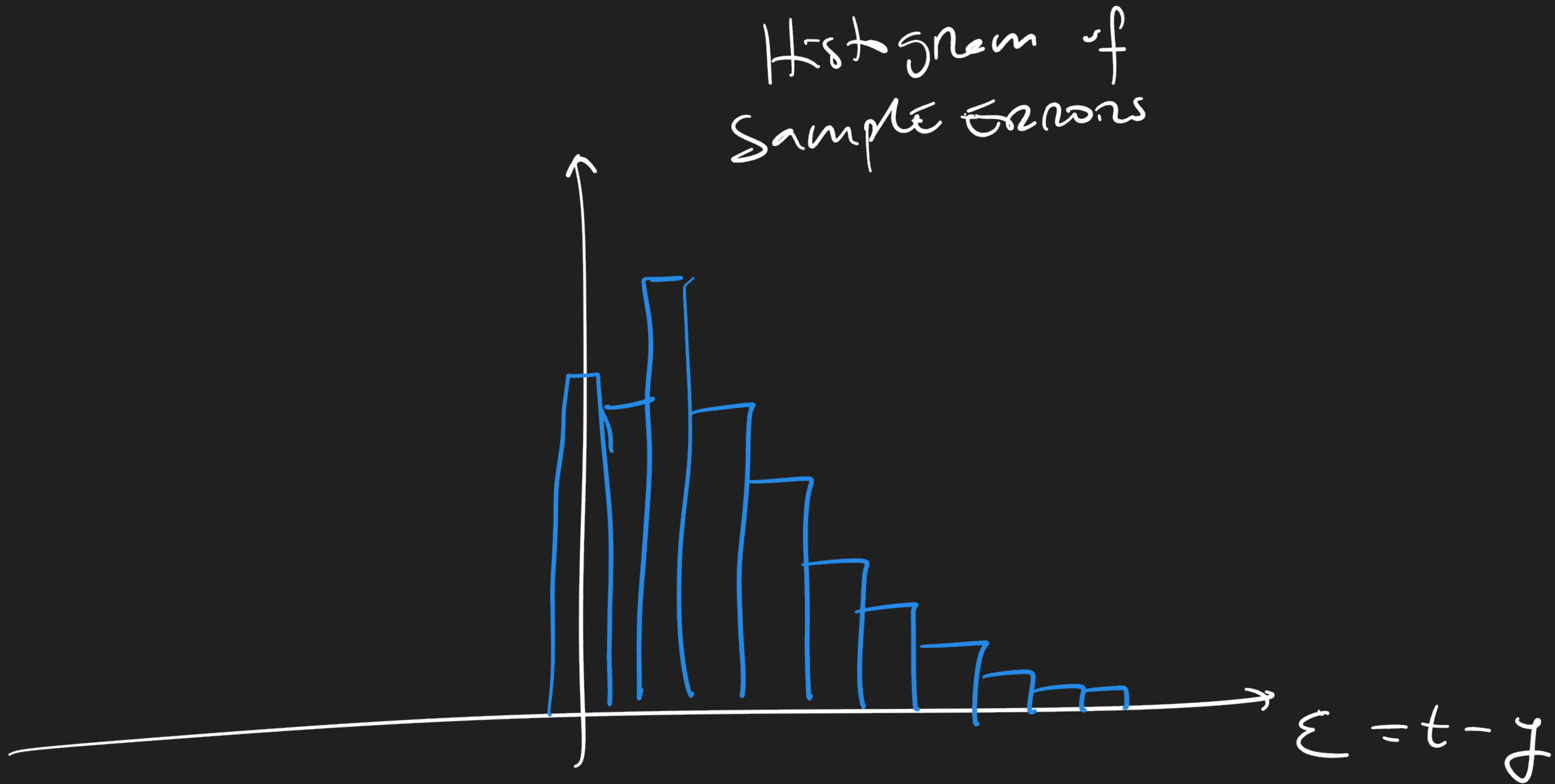
$$= \underset{w}{\arg\max}\,\underbrace{P(t_1|w)\cdot P(t_2|w)\ldots P(t_N|w)}_{P(t_1\cap t_2\cap\ldots\cap t_N|w)}\cdot\underbrace{P(w_0)\cdot P(w_1)\ldots P(w_j)}_{P(w_0\cap w_1\cap\ldots\cap w_j)}$$

$$= \underset{w}{\arg\max}\,\underbrace{P(t|w)}_{\text{Data likelihood}}\cdot\underbrace{P(w)}_{\text{prior probability}}$$

$$J(w) = \frac{1}{2} \sum_{i=1}^{N} (t_i - y_i)^2 + \frac{\lambda}{2} \sum_{j=0}^{M} |w_j|$$

$$\underset{w}{\arg\min}\ J(w) = \underset{w}{\arg\max}\ G(t \mid y, 1) \cdot \mathcal{L}(w \mid 0, 1/\lambda)$$

$$\text{Laplacian}(x \mid a, b) = \frac{1}{2b} \cdot \exp\left\{ -\frac{1}{2} \cdot |x - a| \right\}$$



$t \geq 0$

DATA: $x_i = \{\overset{\text{Tails}}{0}, \overset{\text{Heads}}{1}\}$, $\{x_i\}_{i=1}^{N}$

OBSERVATION: $\{1, 0, 1, 0, 0\}$

DATA-DRIVEN approach:

$$P(x = 1) = \frac{\#\text{ observed "1"}}{\#\text{ total samples}} = \frac{2}{5}$$

Maximum Likelihood Estimation (MLE)

prior-induced approach:

1) Start by "guessing" the candidates for the coin/model parameters.

{fair, 2-headed, coin flips heads w/ 30%}

2) Evaluate each hypothesis:

$$P(\text{fair} \mid \{1,0,1,0,0\}) = \frac{p(\{1,0,1,0,0\} \mid \text{fair}) \cdot p(\text{fair})}{p(\{1,0,1,0,0\})}$$

Every sample is independent

$$= \frac{P^2(1 \mid \text{fair}) \cdot P^3(0 \mid \text{fair}) \cdot P(\text{fair})}{P(\{1,0,1,0,0\})}$$

$$= \frac{(1/2)^2 \cdot (1/2)^3 \cdot 1/3}{-}$$

Law of Total Prob.

$P(\text{2-headed} \mid \text{obs.})$

$P(\text{biased} \mid \text{obs.})$

3) Determine which prior belief is more likely and use it to train the model. (Next lecture).