Likelihood.

$$D = \{16, 4, 64, 32\}$$

$$p(D|h) = \prod p(x|h)$$

$p(D|h \rightarrow \text{Evens})$

$= p(16|\text{Evens}) * \cdots$

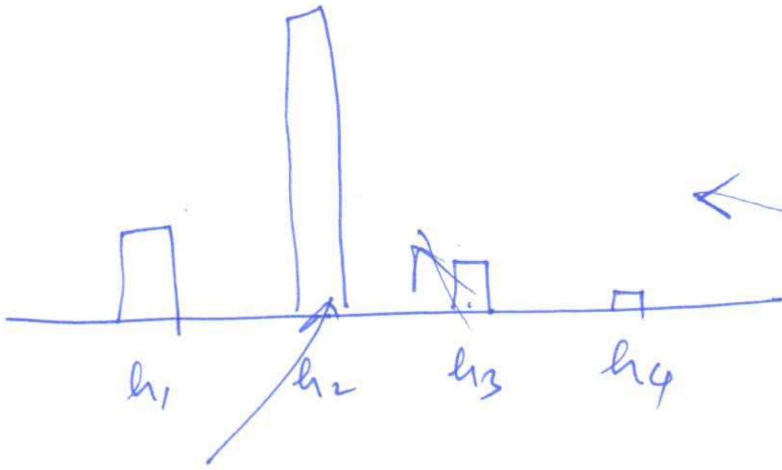
$= \frac{1}{50} * \frac{1}{50} * \frac{1}{50} * \frac{1}{50}$

$p(D|h \rightarrow \text{Squares})$

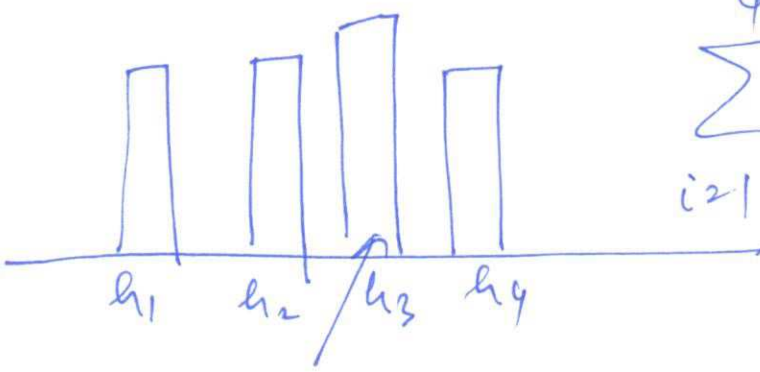
$= \frac{1}{9} * \frac{1}{9} * \frac{1}{9} * 0 = 0$

$p(D|h \rightarrow \text{Powers of } 2)$

$= \frac{1}{6} * \frac{1}{6} * \frac{1}{6} * \frac{1}{6} =$



$\leftarrow p(Y=c \mid c=h_2)$



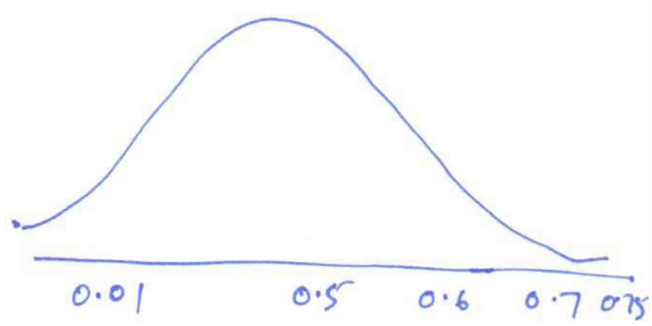
$$\sum_{i=1}^{4} p(Y=c \mid h=h_i)\, p(h_i \mid D)$$

$D = \{T, T, H, T\}$

$P(X^* = \text{heads} \mid \theta)$

$\theta \leftarrow$ probability of heads

$0 \leq \theta \leq 1$



MLE of θ. $D - \{T, T, H, T\}$

What is likelihood of D, given the data is generated from a Bernoulli distribution with θ-para.

$p(x=T|\theta) = (1-\theta)$

$p(x=T|\theta) = (1-\theta)$

$p(X=H|\theta) = \theta$

$p(X=T|\theta) = 1-\theta$

Lik. $= (1-\theta)^3\,\theta$

log.lik $= \log\theta + 3\log(1-\theta)$

Training data $\rightarrow$ N tosses

$N_1 \rightarrow$ # heads

$N-N_1 \rightarrow$ # tails.

lik $= p(D|\theta) = \theta^{N_1} \cancel{*} (1-\theta)^{N-N_1}$

$\frac{d}{d\theta}p(D|\theta) = 0$ [for highest likelihood]

$$\frac{d}{d\theta}\left[\theta^{N_1}(1-\theta)^{N-N_1}\right]$$

$$= N_1\,\theta^{N-1}(1-\theta)^{N-N_1} - (N-N_1)\,\theta^{N_1}(1-\theta)^{N-N_1-1}$$

$$= N_1\,\theta^{N-1}(1-\theta)^{N-N_1} - (N-N_1)\,\theta^{N_1}(1-\theta)^{N-N_1-1}$$

$$= \theta^{N-1}(1-\theta)^{N-N_1-1}\left[N_1(1-\theta) - (N-N_1)\,\theta\right]$$

$$\frac{d}{d\theta}p(D|\theta) = 0$$

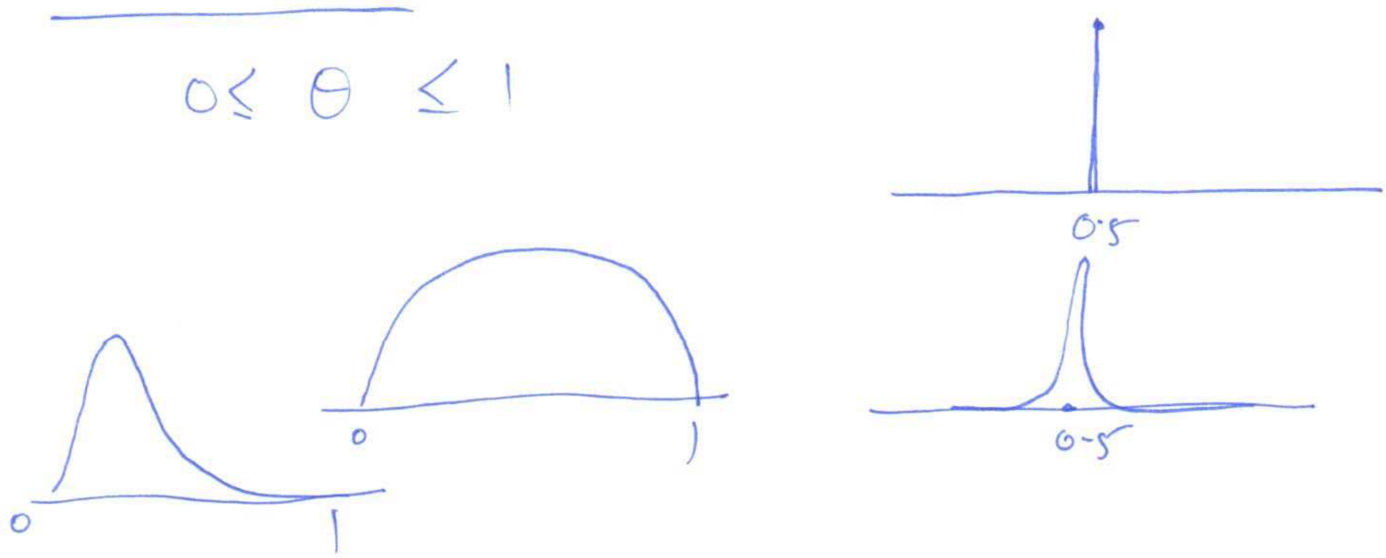
$$\Rightarrow \quad N_1(1-\theta) - (N-N_1)\,\theta = 0$$

$$\boxed{\theta_{MLE} = \frac{N_1}{N}}$$

$$p(x^* = \text{heads} \mid \theta_{MLE}) = \frac{N_1}{N}$$

A Prior on θ

$$0 \leq \theta \leq 1$$

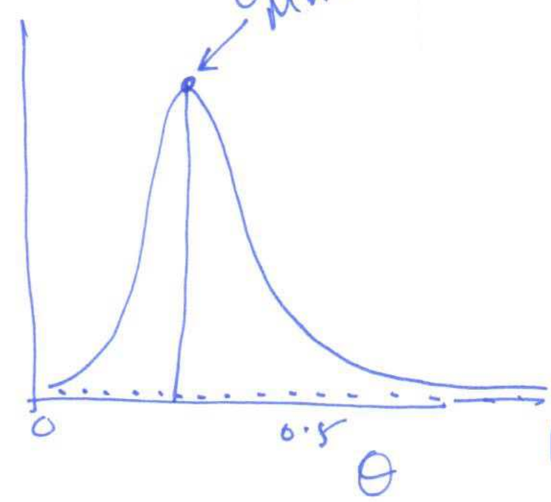


MLE. $\theta_{MLE} = \dfrac{N_1}{N_1 + N_0}$

$p(x^* = heads) = \theta_{MLE}$

MAP $\theta_{MAP} = \dfrac{a + N_1 - 1}{a + b + N_1 + N_0 - 2}$

$p(x^* = heads) = \theta_{MAP}$



Bayesian averaging.

$$p(x^* = heads) = \int_\theta p(x^* = heads \mid \theta)\, p(\theta \mid D)\, d\theta$$

$$= \boxed{\int \theta\, p(\theta \mid D)\, d\theta}$$ Expectation of θ

$$= \frac{N_1 + a}{N_1 + N_0 + a + b}$$

$$p(\theta|D) = \frac{p(D|\theta)\,p(\theta)}{\int_{\theta'} p(D|\theta')\,p(\theta')\,d\theta'}$$

Ignore

$$p(\theta|D) \propto p(D|\theta)\,p(\theta)$$

$$\theta^{N_1}(1-\theta)^{N-N_1} * \theta^{a-1}(1-\theta)^{b-1}$$

$$= \boxed{\theta^{N_1+a-1}(1-\theta)^{N+b-N_1-1}}$$

Posterior is also a Beta distribution

Conjugate priors

Beta $\longleftrightarrow$ Bernoulli

Prior: $p(\theta) = \mathrm{Beta}(\theta|a,b)$

$$\boxed{\mathbb{E}[\theta] = \frac{a}{a+b}}$$

Posterior $p(\theta|D) = \mathrm{Beta}(\theta|N_1+a, N_0+b)$ $(N-N_1)$

$$\mathbb{E}[\theta|D] = \frac{N_1+a}{N_1+N_0+a+b}$$

$x \in R^d$ Assume $D = ?$.

0.7	0.2	0.1
0.6	0.4	-0.2
3.1	2.7	-9.8

$X \rightarrow$ is generated by a MVN

$$\Rightarrow p(X|\mu, \Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

Estimate μ & Σ?

μ_{MLE} Σ_{MLE}

$\mu \rightarrow \mathbb{R}^d$

$\Sigma \rightarrow \mathbb{R}^{d \times d}$

$p(\mu | D)$

$p(\Sigma | D)$

μ_{MAP} Σ_{MAP}.

Pdf of a <u>MVN</u> $x \in \mathbb{R}^d$

$$p(x|\mu,\Sigma) = \mathcal{N}(x|\mu,\Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right]$$

$x_1, x_2, x_3 \ldots x_N$

$$\text{Lik} = \prod_{i=1}^{N} \left[\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)\right]\right]$$