Bayesian Learning

Friday April 2

$D = \{ \quad \}$

$\boxed{D = \{16\}}$

$D = \{1, 4, 16, 64\}$

$h_1 =$ all numbers between 1 & 100

$\boxed{h_2 = \text{all powers of } 4}$

Likelihood

$h =$ all ~~num~~ even numbers

$p(x=2 \mid h) = \frac{1}{\ }$

50

$$D = \{2, 8, 14, 12\}$$

$$L(D|h) = p(x=2|h) * p(x=8|h) * p(x=14|h) * p(x=12|h)$$

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

h: all powers of 4

$$D = \{1, 4, 16, 64\}$$

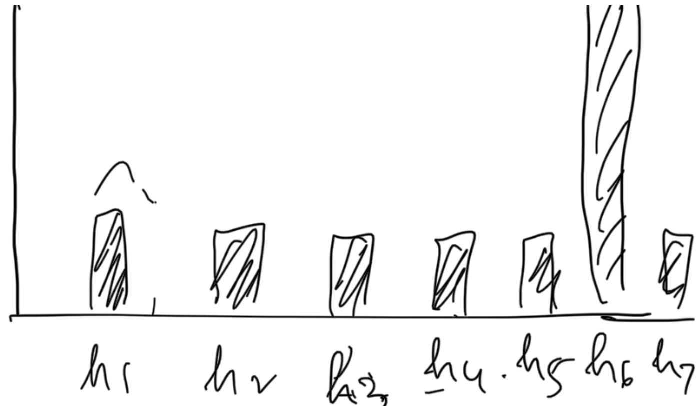
$$L(D|h) = \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4}$$

h: all odd numbers

$$D = \{1, 4, 16, 64\}$$

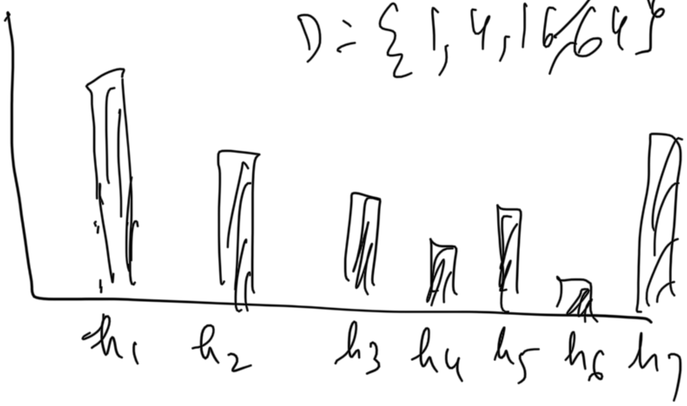
$$L(D|h) =$$

Prior

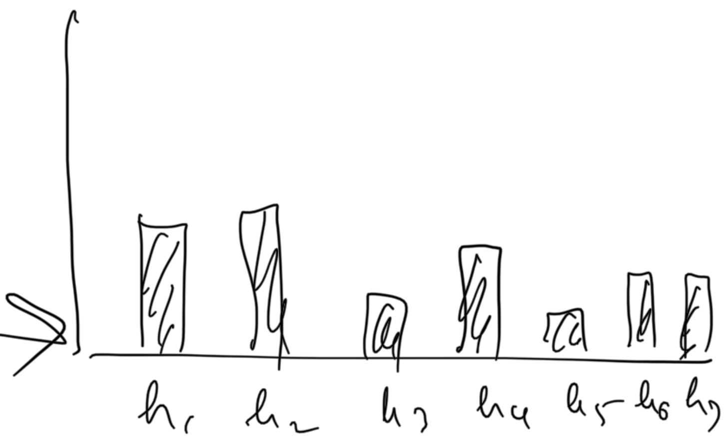


$D = \{1, 4, 16, 64\}$

Likelihood



Posterior distr



x^*

$$\cancel{p(x^* \mid D)}$$

$$P(Y = c \mid X = x^*, D)$$

Mon April 5

Bayesian learning

Given $\boxed{D = \{4, 1, 16\}}$

and a set of hypotheses:

h_1 - All numbers between 1 & 100

h_2 - All even numbers between 1 & 100

h_3 - All powers of 2 between 1 & 100

h_4 - All powers of 4 between 1 & 100

1. Find the correct hypothesis that generated D

2. Find the probability of $x^* = 64$ to be generated by the same hypothesis that generated D.

$D = \{4, 1, 16\}$

1. Prior

$P(h)$

$h^{prior} = h_1$

$$P(x^* = 64 \mid h^{prior}) = \frac{1}{100}$$

2. Likelihood

$$L(D \mid h_1) = P(4 \mid h_1) * P(1 \mid h_1) * P(16 \mid h_1)$$
$$= \frac{1}{100} * \frac{1}{100} * \frac{1}{100} = 10^{-6}$$

$$L(D \mid h_2) = P(4 \mid h_2) * P(1 \mid h_2) * P(16 \mid h_2)$$
$$= \frac{1}{50} * 0 * \frac{1}{50} = 0$$

$$L(D \mid h_3) = \frac{1}{7} * \frac{1}{7} * \frac{1}{7} = 0.002$$

$$L(D \mid h_4) = \frac{1}{4} * \frac{1}{4} * \frac{1}{4} = 0.015$$

$h^{MLE} = h_4$

$$P(x^* = 64 \mid h_4) = \frac{1}{4} = 0.25$$

3. MAP

$$P(h_1 \mid D) = \frac{P(D \mid h_1) * P(h_1)}{P(D \mid h_1) P(h_1) + P(D \mid h_2) P(h_2) + P(D \mid h_3) P(h_3)}$$

$+ P(D|h_4)\, P(h_4)$

$$P(h_2|D) = \frac{P(D|h_2) * P(h_2)}{P(D|h_1)\,P(h_1) + P(D|h_2)\,P(h_2) + P(D|h_3)\,P(h_3) + P(D|h_4)\,P(h_4)}$$

$P(h_3|D) =$

$P(h_4|D) =$

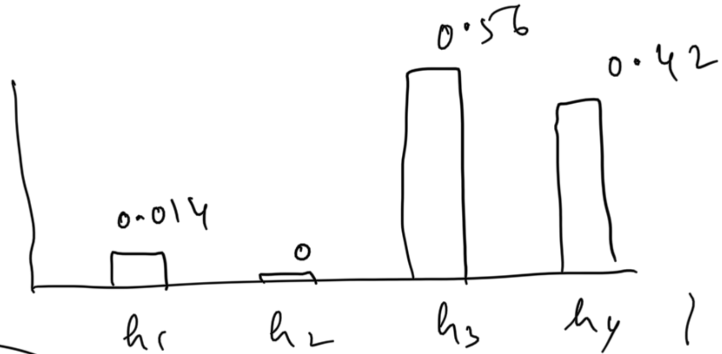
$$Z = 10^{-6} * 0.5 + 0 * 0.489 + 0.002 * 0.01 + 0.015 * 0.001$$

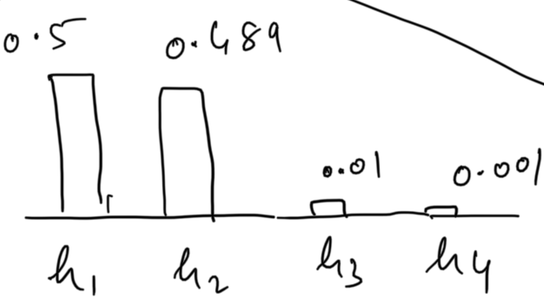
$$= 3.55e{-5}$$

$$P(h_1|D) = \frac{10^{-6} * 0.5}{3.55 * 10^{-5}}$$

$\vdots$

$P(h|D)$





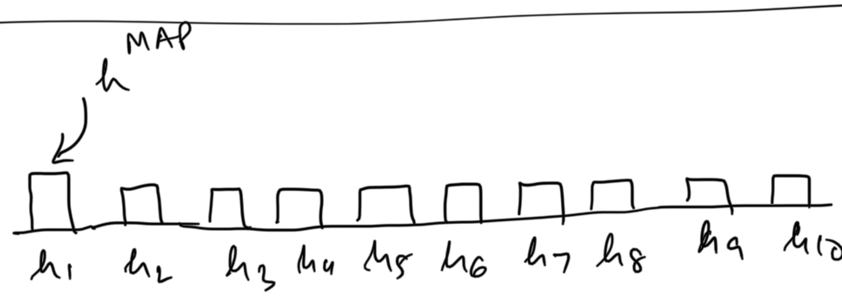
$p(h)$

$$h^{MAP} = h_3$$

$$p(x^* = 64 \mid h_3) = \frac{1}{7}$$

4.

$p(h|D)$



Bayesian Averaging.

$$\underline{p(x^* = 64 \mid D)} = p(x^* = 64 \mid h_1)\, \underline{P(h_1 \mid D)}$$
$$+ \, p(x^* = 64 \mid h_2)\, P(h_2 \mid D)$$
$$+ \, P(x^* = 64 \mid h_3)\, \underline{P(h_3 \mid D)}$$
$$+ \, P(x^* = 64 \mid h_4)\, \underline{P(h_4 \mid D)}$$

---

① Frequentists → MLE

② Bayesian → MAP

→ Bayesian Averaging

$$X \in \{0, 1\} \quad \{tail, head\}$$

$D := \{ \underbrace{0, 0, 0, 1, 1, 0, 0, 1, 1, \ldots}_{N} \}$

$= N_0$ 0's (fails)

and N_1 1's (heads)

$X \sim \text{Ber}(\theta)$

$0 \leq \theta \leq 1$

---

likelihood of D:

$\boxed{\text{i.i.d}}$

$p(D|\theta) = \cancel{\beta}\ \theta^{N_1} (1-\theta)^{N_0}$

independent & identically distributed

$\hat{\theta}_{MLE} = \underset{\theta}{\text{argmax}}\ \underline{\theta^{N_1}(1-\theta)^{N_0}}$

$$\frac{d}{d\theta} \left\{ \underline{\theta^{N_1}}\ \underline{(1-\theta)^{N_0}} \right\} = N_1 \theta^{N_1 - 1} (1-\theta)^{N_0} + N_0 \theta^{N_1} (1-\theta)^{N_0 - 1} (-1)$$

$$= \theta^{N_1 - 1} (1-\theta)^{N_0 - 1} \left[N_1 (1-\theta) - N_0 \theta \right]$$

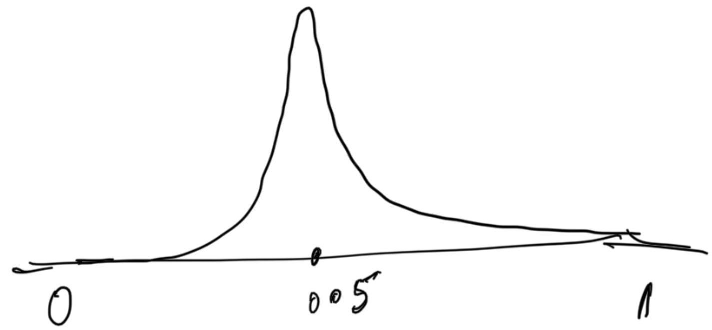
$$\boxed{\hat{\theta}_{MLE} = \frac{N_1}{(N_0 + N_1)}}$$

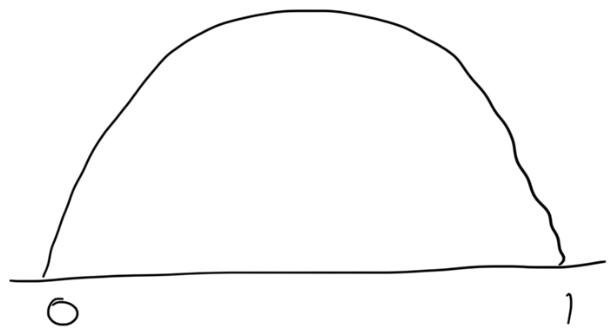
$$p(x^* = 1 | D) = p(x^* = 1 | \hat{\theta}_{MLE})$$
$$= \hat{\theta}_{MLE}$$

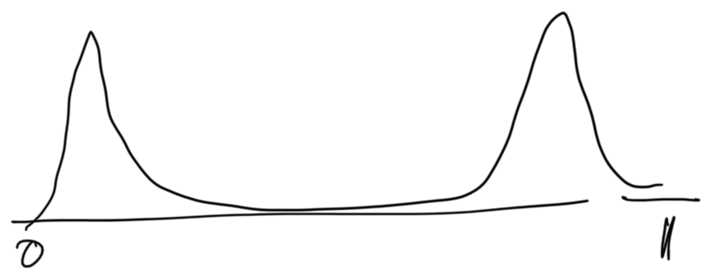
$0 \le \theta \le 1$

Prior for θ:

Beta( )







Beta (a, b)

θ^{prior} argmax Beta(a, b)

$$\boxed{\frac{a-1}{a+b-1}}$$

Wed April 7

Deen Office Hours at 2:45 PM

e.g: H H H H T

$$\hat{\theta}_{MLE} = \frac{4}{5} = 0.8$$

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

$$\text{var}[\theta] = \frac{ab}{(a+b)^2(a+b+1)}$$

$$\text{Mode} = \frac{a-1}{a+b-2}$$

$$p(\theta|D) = \frac{p(D|\theta)\,p(\theta)}{\int_0^1 p(D|\theta')\,p(\theta')\,d\theta'}$$

posterior pdf for θ

$$\theta^{N_1}(1-\theta)^{N_0}\frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}$$

$$= \frac{\theta \quad \cdots}{\int \cdots}$$

Constant w.r.t. θ

$$\boxed{p(\theta|D) \propto \theta^{N_1+a-1}(1-\theta)^{N_0+b-1}}$$

$$\boxed{\theta|D \sim \text{Beta}(\hat{a} = N_1 + a, \hat{b} = N_0 + b)}$$

Beta and Bernoulli are Conjugate ~~prior~~ pairs.

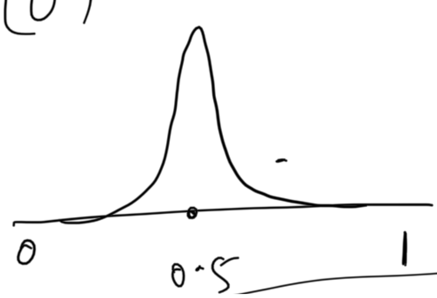
The posterior $\theta|D$ will also be a Beta distributed r.v.

$$\theta|D = \text{Beta}(a + N_1, b + N_0)$$

$$\boxed{\hat{\theta}_{MAP} = \frac{a + N_1 - 1}{a + b + N - 2}} \qquad N = N_0 + N_1$$

$p(\theta) \sim \text{Beta}(4, 4)$ $\qquad D = $ H, H, H, H, T

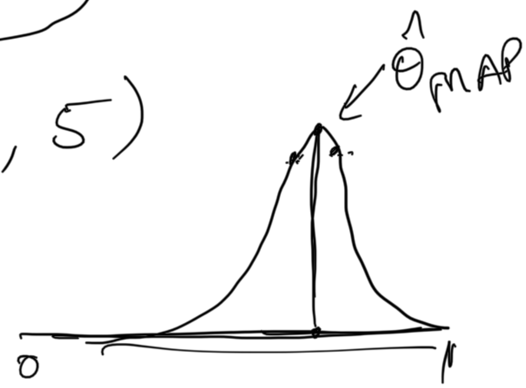
$N_1 = 4 \quad N_0 = 1$



$$\boxed{\hat{\theta}_{MLE} = \frac{4}{5} = 0.8}$$

$$\hat{\theta}_{prior} = \frac{4-1}{4+4-2} = \frac{3}{6} = 0.5$$

$$p(\theta|D) \sim \text{Beta}(8, 5)$$



$$P(x^*=1 \mid D) = \int_{\theta} P(x^*=1|\theta)\, p(\theta|D)\, d\theta$$

$$= \int_0^1 \theta\, p(\theta|D)\, d\theta$$

$$= \mathbb{E}[\theta|D]$$

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

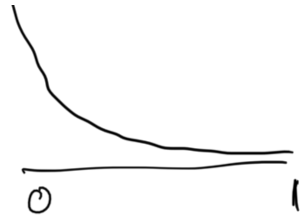
$$D = \{T, T, T, T, T\}$$

$N_1 = 0 \qquad N_0 = 5$

$$\hat{\theta}_{MLE} = 0$$

1

$$p(\theta) = \underline{\text{Beta}(2, 10)}$$



$$\hat{\theta}_{MAP} = \frac{a + N_1 - 1}{a + b + N - 2} = \frac{2 + 0 - 1}{2 + 10 + 5 - 2}$$

$$= \frac{1}{15}$$

If $N_1 = 5 \quad N_0 = 0$

$$\hat{\theta}_{MAP} = \frac{2 + 5 - 1}{2 + 10 + 5 - 2} = \frac{6}{15}$$

<u>MVN</u>

$\mathbf{x}$ is a vector of length D

μ – a vector $(D \times 1)$

Σ – $D \times D$ matrix

---

use MLE to estimate μ & Σ

$$L(D \mid \mu, \Sigma) = \prod_{i=1}^{N} p(\mathbf{x}_i \mid \mu, \Sigma)$$

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

$LL(D|\mu,\Sigma)$

$$= -\sum_{i=1}^{N} \log\left[(2\pi)^{D/2} |\Sigma|^{1/2}\right]$$

$$- \frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)$$

$$= -\frac{ND}{2}\log 2\pi - \frac{N}{2}\log|\Sigma|$$

$$- \frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)$$

$$\frac{\partial}{\partial\mu} LL(D|\mu,\Sigma) = 0$$

$$\frac{\partial}{\partial\Sigma} LL(D|\mu,\Sigma) = 0$$

Check class handouts + Matrix cookbook

Bayesian learning:

- → MLE
- → Prior
- → Posterior
- → MAP
- → Bayesian averaging

Bernoulli

Gaussian distribution