EECE-5644
Intro. to Machine Learning
Spring 2016
Homework #: 2

Submission Date: _2_ / _8_ / 2016

Name (Last, First):

Homework Submission Rules:

1. All submitted work should be legible. Do not write in small or in overly cursive characters.
2. Your submission of problem solutions must be in the given order, i.e., P1, P2, P3, etc. Do not submit in a random order.
3. Use this cover page as the first page for each homework submission. Homework submitted without a cover page will result in a 10% reduction in the overall score for that homework.
4. All plots must have axis labeled, titles, and legends if applicable.
5. Code specified in the problems must be submitted via the Blackboard site alongside a pdf of your homework solutions. The code must be contained in a zip file with the following file name format: hw_##_lastname.zip. Do not print the code
6. Be concise when writing your solutions and use both sides of the page.

(PLEASE DO NOT WRITE BELOW THIS LINE)

Problem-#	1	2	3	4	5	6	7	8	9	10	TOTAL
Points											

Elise Jortberg
EECE 5644
HW 2

1) 2.12

$P(\omega_{max}|x) \geq P(\omega_i|x)$ for $i = 1 \dots c$

a) Show $P(\omega_{max}|x) \geq 1/c$

$\int_0^\infty p(x)\,dx = 1 \longrightarrow \sum_{i=1}^{c} p(\omega_i|x) = 1$

upper limit $p(\omega_i|x) = p(\omega_{max}|x)$

$\therefore p(\omega_{max}|x) \geq \frac{1}{c}$ else would never add to 1

- if $p(\omega_{max}|x) < \frac{1}{c}$ then $\sum = \frac{1}{c} + \frac{1}{c+a_1} + \frac{1}{c+a_2} + \dots$ never converges to 1
- "worst case" scenario, all $p(\omega_i|x)$ equal then $p(\omega_{max}|x) = \frac{1}{c}$, so otherwise has to be $\geq 1/c$

b) Show $p(error) = 1 - \int p(\omega_{max}|x)p(x)\,dx$

$p(error) = 1 - p(correct)$

$= likelihood \cdot prior$

$= 1 - \int p(\omega_{max}|x)p(x)\,dx$

c) Show $p(error) \leq \frac{c-1}{c}$

$p(error) \leq 1 - \int \frac{1}{c} p(x)\,dx$

$\leq 1 - \frac{1}{c} \cancel{\int p(x)\,dx}$ 1

$\leq 1 - 1/c$

$\leq \frac{c-1}{c}$

d) The error in c) would occur when $p_1(x) = p_2(x)$ (distributions are the same). i.e. a gaussian where $\mu_1 = \mu_2$ and $\Sigma_1 = \Sigma_2$

2) 3.1

$$p(x|\theta) = \begin{cases} \theta e^{-\theta x} & x \geq 0 \\ 0 & \text{else} \end{cases}$$

a) see figures

b) Show $\hat{\theta}_{ML} = \frac{1}{\frac{1}{n}\sum_{k=1}^{n} x_k}$

$\nabla_\theta p(D|\theta) = 0$

$D = \{x_1, \ldots, x_n\}$

$\hat{\theta}_{ML} = \arg\max_\theta \ln(p(D|\theta))$

$\sum_{k=1}^{n} \ln(p(x_k|\theta)) = \sum_{k=1}^{n} \ln(\theta) - \theta x_k$

$\nabla_\theta \; n\ln\theta - \sum_{k=1}^{n} \theta x_k = 0$

$\frac{\partial}{\partial\theta}(n\ln\theta - \sum_{k=1}^{n} \theta x_k) = 0$

$\frac{n}{\theta} - \sum_{k=1}^{n} x_k = 0 \longrightarrow \frac{n}{\theta} = \sum_{k=1}^{n} x_k$

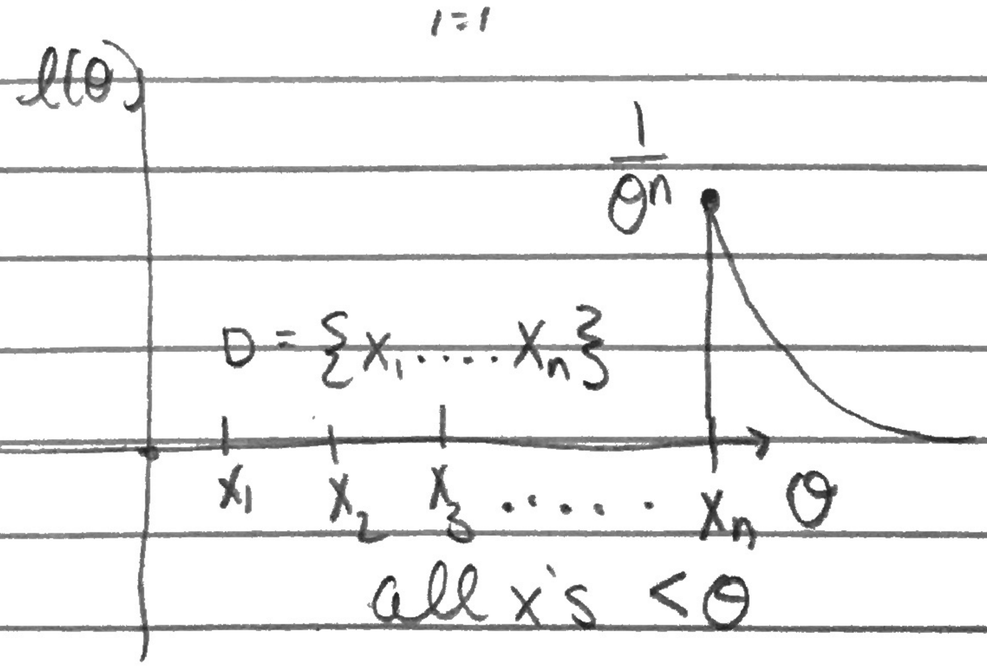
$\hat{\theta}_{ML} = \frac{n}{\sum_{k=1}^{n} x_k} = \frac{1}{\frac{1}{n}\sum_{k=1}^{n} x_k}$

3) 3.2

$$p(x|\theta) \sim U(0,\theta) = \begin{cases} 1/\theta & 0 \leq x \leq \theta \\ 0 & \text{else} \end{cases}$$

a) $l(\theta) = \prod_{i=1}^{n} p(x_i|\theta)$

$= \prod_{i=1}^{n} \frac{1}{\theta}$ where $x_i \leq \theta = \frac{1}{\theta^n}$



$\theta < \max(D)$ when $l(\theta) = 0$

$\theta \geq \max(D)$ when $l(\theta) = \frac{1}{\theta^n}$

$l(\theta) = \frac{1}{\theta^n}$ is maximized when θ is small

$\therefore \hat{\theta}_{ML} = \max(D) = x_n$

4) 3.8

$\begin{cases} z_{ik} = 1 & k^{th} \text{ sample} = \omega_i \\ z_{ik} = 0 & \text{else} \end{cases}$

$P(z_{i1}, \ldots, z_{in} \mid p(\omega_i)) = \prod_{k=1}^{n} P(\omega_i)^{z_{ik}} (1-P(\omega_i))^{1-z_{ik}}$

a) Show

$\begin{cases} p(\omega_i) = p(z_{ik}=1 \mid p(\omega_i)) & \text{if: } z_{ik} \\ 1-p(\omega_i) = p(z_{ik}=0 \mid p(\omega_i)) & \text{else: } 1-z_{ik} \end{cases}$

$p(z|\omega) = ?$ for all z_{ik}

in general: $p(x|\theta) = \prod_{k=1}^{n} p(x_k|\theta)$ (1)

$\nabla_\theta \ln(p(x_k|\theta)) = 0$ (2)

$p(z_{ik} \mid p(\omega_i)) = p(\omega_i)^{z_{ik}} (1-p(\omega_i))^{1-z_{ik}}$

$\therefore P(z_{i1} \ldots z_{in} \mid p(\omega_i)) = \prod_{k=1}^{n} p(\omega_i)^{z_{ik}} (1-p(\omega_i))^{1-z_{ik}}$

b) Show $\hat{P}_{ML}(\omega_i) = \frac{1}{n} \sum_{k=1}^{n} z_{ik}$

(2): $\frac{d}{\partial p(\omega_i)} \ln\left(\prod_{k=1}^{n} p(\omega_i)^{z_{ik}} (1-p(\omega_i))^{1-z_{ik}}\right)$

$\frac{d}{dp(\omega_i)} \sum_{k=1}^{n} z_{ik} \ln(p(\omega_i)) + (1-z_{ik}) \ln(1-p(\omega_i)) = 0$

$\sum_{k=1}^{n} z_{ik} \frac{1}{p(\omega_i)} + \frac{1}{1-p(\omega_i)} \sum_{k=1}^{n} (1-z_{ik}) = 0$

$\sum_{k=1}^{n} z_{ik} \frac{1}{p(\omega_i)} = \frac{1}{1-p(\omega_i)} \sum_{k=1}^{n} (1-z_{ik})$

$(1-p(\omega_i)) \sum_{k=1}^{n} z_{ik} = p(\omega_i) \sum_{k=1}^{n} (1-z_{ik})$

$\sum_{k=1}^{n} z_{ik} - p(\omega_i) \sum_{k=1}^{n} z_{ik} = p(\omega_i) \sum_{k=1}^{n} 1 = p(\omega_i) \sum_{k=1}^{n} z_{ik}$

$\sum_{k=1}^{n} z_{ik} - p(\omega_i) \sum_{k=1}^{n} z_{ik} = n\,p(\omega_i) - p(\omega_i) \sum_{k=1}^{n} z_{ik}$

$\sum_{k=1}^{n} z_{ik} (1-p(\omega_i)) = p(\omega_i) \left(n - \sum_{k=1}^{n} z_{ik}\right)$

$\frac{\sum_{k=1}^{n} z_{ik}}{n - \sum_{k=1}^{n} z_{ik}} = \frac{p(\omega_i)}{1-p(\omega_i)}$ →

$\therefore p(\omega_i) = \frac{1}{n} \sum_{k=1}^{n} z_{ik}$

This result means that we estimate/guess at $p(\omega_i)$ based on the input/training dataset to initialize the classifier/model. We don't base guess on prior/human knowledge.

5) 3.4

x = binary vector w/ distribution

$$P(x|\vec{\theta}) = \prod_{i=1}^{d} \theta_i^{x_i} (1-\theta_i)^{1-x_i}$$

where $\vec{\theta} = (\theta_1, \ldots \theta_d)^t$ unknown parameter vector

θ_i = probability $x_i = 1$

Show $\hat{\vec{\theta}}_{ML} = \frac{1}{n} \sum_{k=1}^{n} x_k$

in general: $p(x|D) = \int p(x|\theta)\, p(\theta|D)\, d\theta$

$D = \{x_1, \ldots x_n\}$ > multivariate

$\theta = \{\theta_1, \ldots \theta_d\}$

$p(D|\theta) = \prod_{k=1}^{n} p(x_k|\theta)$ becomes

$$p(D|\theta) = \prod_{k=1}^{n} \prod_{i=1}^{d} \theta_i^{x_{ki}} (1-\theta_i)^{1-x_{ki}}$$

$\nabla_\theta \ln(p(D|\theta)) = 0$

$$\nabla_{\theta_i}: \sum_{k=1}^{n} \sum_{i=1}^{d} x_{ki} \ln(\theta_i) + (1-x_{ki}) \ln(1-\theta_i) = 0$$

$$\sum_{k=1}^{n} \sum_{i=1}^{d} \frac{x_{ki}}{\theta_i} + (1-x_{ki}) \frac{1}{\theta_i - 1}$$

find each $\frac{\partial \ln p(x|\theta)}{\partial \theta_j} = 0$ based on arbitrary j

for singular case

→

$$\sum_{j=1}^{n}\sum_{i=1}^{d}\frac{\partial}{\partial\theta_k}\left[x_{ji}\ln\theta_i+(1-x_{ji})\ln(1-\theta_i)\right]$$

$$\frac{\partial}{\partial\theta_1}=x_{j2}\ln\theta_2=0$$

$$\frac{\partial}{\partial\theta_1}=x_{j1}\ln\theta_1=\frac{x_{j1}}{\theta_1}$$

substitute in so for one case other thetas $\to 0$

$\therefore$ for any i in vector $\theta=\{\theta_1,\ldots,\theta_d\}$

$$[\nabla_\theta l(\theta)]_i=\nabla_{\theta_i}l(\theta)$$

$$\frac{1}{\hat\theta_i}\sum_{k=1}^{n}x_{ki}=\frac{1}{1-\hat\theta_i}\sum_{k=1}^{n}(1-x_{ki})$$

Solve for $\hat\theta_i$

$$1-\hat\theta_i\sum_{k=1}^{n}x_{ki}=\hat\theta_i\sum_{k=1}^{n}(1-x_{ki})$$ (like problem 3.3)

$$\therefore\quad\hat\theta_i=\frac{1}{n}\sum_{k=1}^{n}x_{ki}$$

Since we solved for arbitrary i, this applies for all $i\in[1,d]$

$$\vec\theta_{ml}=\frac{1}{n}\sum_{k=1}^{n}\vec{x}_k$$

(e) 3.17

Derive Bayesian classifier for d dimensional multivariate Bernoulli case. $P(\vec{x}\mid D)=P(\vec{x}\mid\theta_i,\omega_i)$

$$P(\vec{x}\mid\vec\theta)=\prod_{i=1}^{d}\theta_i^{x_i}(1-\theta_i)^{1-x_i}$$

$$D=\{\vec{x}_1,\ldots,\vec{x}_n\}$$

a) if $\vec{s}=(s_1\ldots s_d)^t$ is sum of n samples show

$$P(D\mid\vec\theta)=\prod_{i=1}^{d}\theta_i^{s_i}(1-\theta_i)^{n-s_i}$$

17) $p(\vec{x}|\vec{\theta}) = \prod_{i=1}^{d} \theta_i^{x_i}(1-\theta_i)^{1-x_i}$

$D = \{\vec{x}_1 \dots \vec{x}_n\}$

a) $\vec{s} = (s_1 \dots s_d)^t$ sum of n samples

$p(D|\theta) = \prod_{k=1}^{n} p(\vec{x}_k|\vec{\theta})$

k has associated d, k's independent

$p(\vec{x}_k|\vec{\theta}) = \prod_{i=1}^{d} \theta_i^{x_{ki}}(1-\theta_i)^{1-x_{ki}}$

$p(D|\vec{\theta}) = \prod_{k=1}^{n}\prod_{i=1}^{d} \theta_i^{x_{ki}}(1-\theta_i)^{1-x_{ki}}$

for $\theta^{x_{ki}}$, $(1-\theta_i)^{1-x_{ki}}$ multiplying over k, $1 \to n$

in exponent $= \sum_{k=1}^{n}$

$P(D|\vec{\theta}) = \prod_{i=1}^{d} \theta^{\sum_{k=1}^{n} x_{ki}}(1-\theta_i)^{\sum_{k=1}^{n}(1-x_{ki})}$

$p(D|\vec{\theta}) = \prod_{i=1}^{d} \theta^{s_i}(1-\theta_i)^{n-s_i}$

b) assume uniform prior distribution for θ and identity $\int_0^1 \theta^m(1-\theta^n)\,d\theta = \frac{m!\,n!}{(m+n+1)!}$ solve $p(\vec{\theta}|D)$

$p(\vec{\theta}|D) = \frac{p(D|\vec{\theta})\,p(\vec{\theta})}{p(D)}$

uniform $\therefore p(\vec{\theta}) = 1$, need $p(D)$...

$p(D) = \int p(D|\vec{\theta})\,p(\vec{\theta})\,d\theta$

$= \int_{1\text{ to }d} \prod_{i=1}^{d} \theta^{s_i}(1-\theta_i)^{n-s_i}\,d\theta_{(1\dots d)}$

multiple integrals $1 \to d = \int_{1\text{ to }d}$, BINARY

$p(D) = \prod_{i=1}^{d} \int_0^1 \theta^{s_i}(1-\theta_i^{n-s_i})\,d\theta_i$ for arbitary case

identity: $m = s_i$, $n = n-s_i$

$p(D) \stackrel{!}{=} \prod_{i=1}^{d} \frac{s_i!\,(n-s_i)!}{(s_i+n-s_i+1)!} = \prod_{i=1}^{d} \frac{s_i!\,(n-s_i)!}{(n+1)!}$

plug back in to $p(\vec{\theta}|D)$ to solve $\rightarrow$

$$p(\vec{\theta}|D) = \frac{\prod_{i=1}^{d} \theta_i^{s_i}(1-\theta_i)^{n-s_i}}{\prod_{i=1}^{d} \frac{s_i!(n-s_i)!}{(n+1)!}}$$

$$= \prod_{i=1}^{d} \frac{(n+1)!\,\theta_i^{s_i}(1-\theta_i)^{n-s_i}}{s_i!(n-s_i)!}$$

c) plot $p(\vec{\theta}|D)$ for $d=1$, $n=1$ and both cases of s

$$p(\theta_{\substack{d=1\\n=1}}|D) = \frac{2!}{s_1!(1-s_1)!}\theta_1^{s_1}(1-\theta_1)^{1-s_1}$$

$$s_1 : \begin{cases} 0 \\ 1 \end{cases} \quad \text{for} \quad 0 \le \theta_1 \le 1$$

FIGURE IN MATLAB

d) Integrate $p(\vec{x}|\vec{\theta})p(\vec{\theta}|D)\,d\theta$ to find $p(\vec{x}|D)$

$$p(\vec{x}|D) = \int p(\vec{x}|\vec{\theta})p(\vec{\theta}|D)\,d\theta$$

$$= \int \prod_{i=1}^{d} \theta_i^{x_{ki}}(1-\theta_i)^{1-x_{ki}} \prod_{i=1}^{d} \frac{(n+1)!\,\theta_i^{s_i}(1-\theta_i)^{n-s_i}}{s_i!(n-s_i)!}\,d\theta$$

$$= \int \prod_{i=1}^{d} \theta_i^{x_i+s_i}(1-\theta_i)^{1-x_i+n-s_i} \frac{(n+1)!}{s_i!(n-s_i)!}\,d\theta$$

x_i = BINARY

no θ component

$$= \prod_{i=1}^{d} \frac{(n+1)!}{s_i!(n-s_i)!} \int_0^1 \theta_i^{x_i+s_i}(1-\theta_i)^{1-x_i+n-s_i}\,d\theta$$

use identity $m = x_i+s_i$, $n = 1-x_i+n-s_i$

$$= \prod_{i=1}^{d} \frac{(n+1)!}{s_i!(n-s_i)!} \frac{(x_i+s_i)!(1-x_i+n-s_i)!}{(x_i+s_i+1-x_i+n-s_i+1)}$$

$$p(\vec{x}|D) = \prod_{i=1}^{d} \frac{(n+1)!}{s_i!(n-s_i)!} \frac{(x_i+s_i)!(1-x_i+n-s_i)!}{(n+2)!}$$

$$\frac{(n+1)!}{(n+2)!} = \frac{1}{n+2}$$

$$p(\vec{x}|D) = \frac{(x_i+s_i)!(1-x_i+n-s_i)!}{s_i!(n-s_i)!(n+2)}$$

$\longrightarrow$

if $x_i = 1$

$$p(x|D) = \frac{(1+s_i)!\,(1+n-s_i)!}{s_i!\,(n-s_i)!\,(n+2)} = \frac{s_i+1}{n+2}$$

if $x_i = 0$, $1-(x_i = 1)$

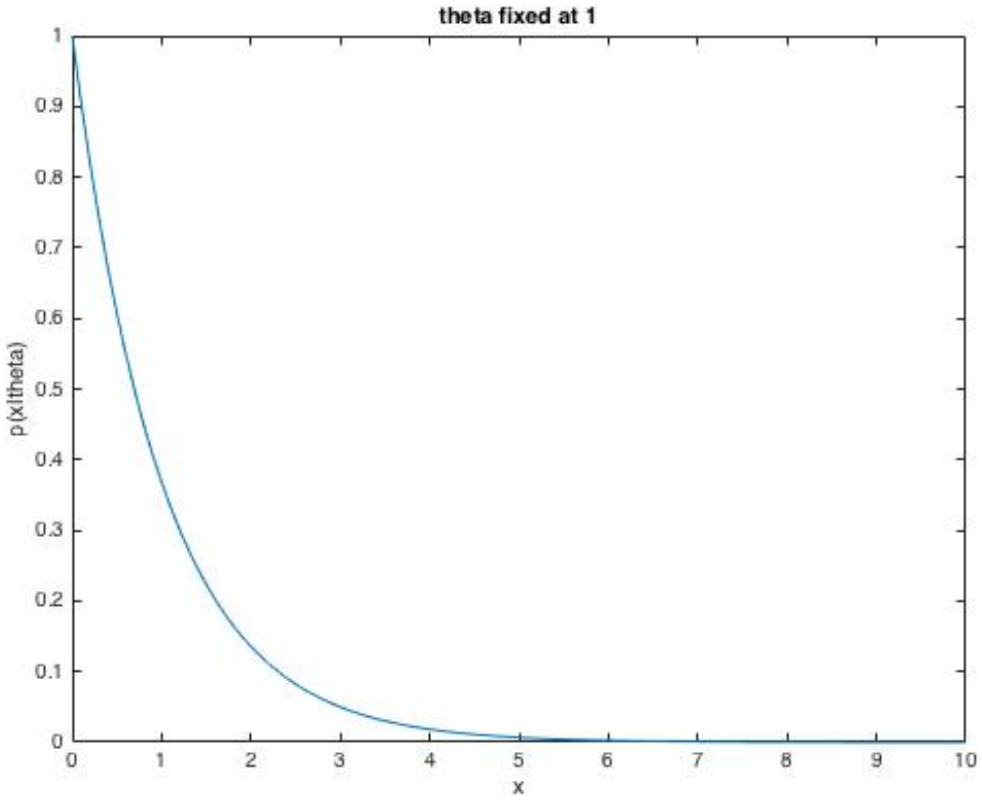
$$p(x|D) = 1 - \frac{s_i+1}{n+2}$$

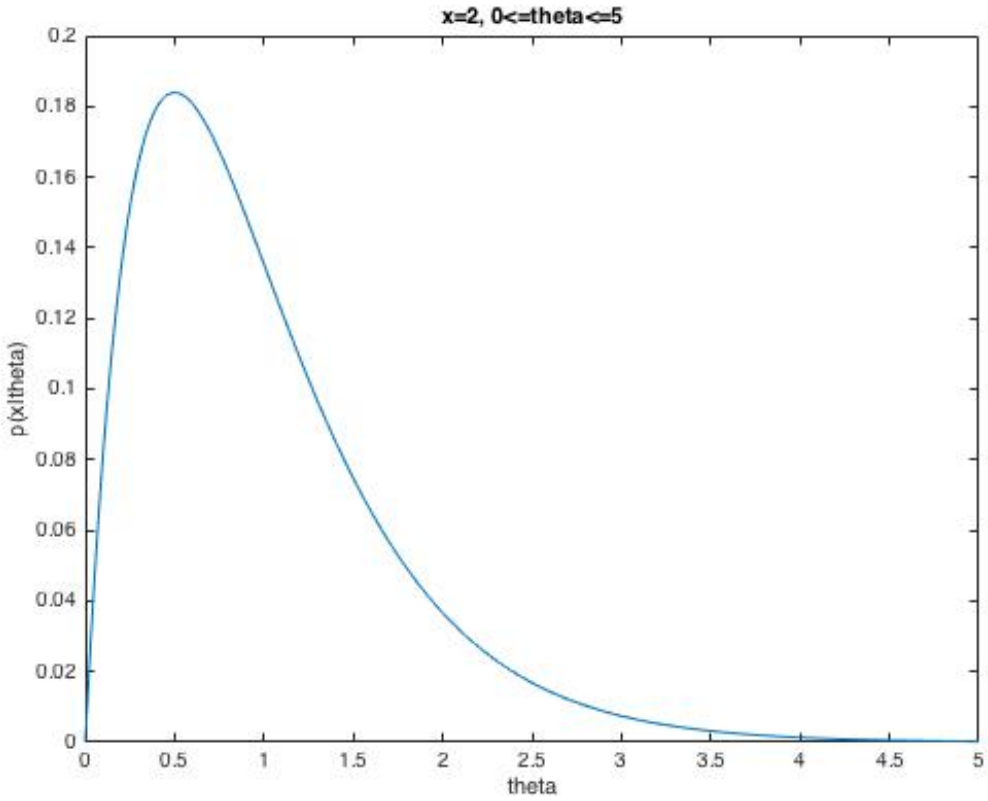
$$p(\vec{x}|D) = \left(\frac{s_i+1}{n+2}\right)^{x_i}\left(1-\frac{s_i+1}{n+2}\right)^{1-x_i}$$

e) obtain $p(\vec{x}|D)$ by using $\hat{\theta}$ for $\vec{\theta}$ in $p(\vec{x}|\vec{\theta})$ what is effective Bayesian estimate for $\vec{\theta}$?

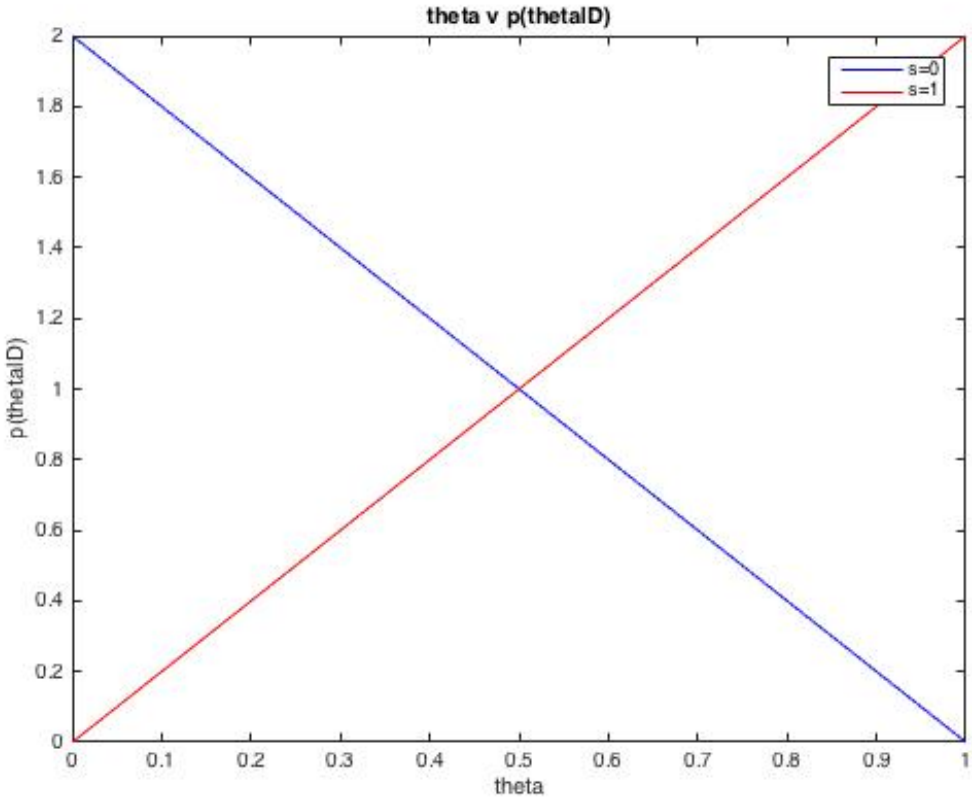
$p(\vec{x}|D)$ boils down to $\hat{\theta}_i = \frac{s_i+1}{n+2}$ to predict $\hat{\theta}_i$ for x_i case (not using $(1-\theta_i)^{1-x_{ki}}$ term)

Problem 3.1:


theta fixed at 1
p(x|theta)
x
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
1
2
3
4
5
6
7
8
9
10


x=2, 0<=theta<=5
p(x|theta)
theta
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5

Problem 3.17


theta v p(thetaID)
s=0
s=1
p(thetaID)
theta
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0.1
0.3
0.5
0.7
0.9