1

HW - 1

Q1

$P(\omega_1) = P(\omega_2)$

$p(x|\omega_i) \sim N(\mu_i, \sigma_i^2)$

$\mu_1 = 0$

$\sigma_1^2 = 1$

$\mu_2 = \mu$

$\sigma_2^2 = \sigma^2$

a) Bayes optimal decision boundary

conditional risk

$$R(\alpha_i|x) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j) P(\omega_j|x)$$

To find x^* :

$$R(\alpha_1|x) = R(\alpha_2|x)$$

with 1/0 loss we can get

$$R(\alpha_1|x) = 1 - P(\omega_1|x)$$
$$= P(\omega_2|x)$$

$$R(\alpha_2|x) = 1 - P(\omega_2|x)$$
$$= P(\omega_1|x)$$

hence $P(\omega_2|x) = P(\omega_1|x)$

$$= P(x|\omega_1) \times \cancel{P(\omega_1)} = P(x|\omega_2) \times \cancel{P(\omega_2)}$$

$$P(x|\omega_1) = P(x|\omega_2)$$

0000000

M Tu W Th F S Su

DD MM YYYY

a) $N(\mu_1, \alpha_1^2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x)^2}{2}}$

$N(\mu_2, \alpha_2^2) = \frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{(x-\mu)^2}{2\alpha^2}}$

taking ln on both sides

$$\ln(P(x/\omega_1)) = \ln P((x/\omega_2))$$

$$-\frac{1}{2}\ln(2\pi) - \frac{x^2}{2} = -\frac{1}{2}\ln(2\pi\alpha^2) - \frac{(x-\mu)^2}{2\alpha^2}$$

$$-\frac{x^2}{2} + \frac{(x-\mu)^2}{2\alpha^2} = -\frac{1}{2}\ln(2\pi\alpha^2) + \frac{1}{2}\ln(2\pi)$$

$$\frac{-x^2\alpha^2 + (x-\mu)^2}{2\alpha^2} = \frac{1}{2}\left[\ln(2\pi) - \ln(2\pi\alpha^2)\right]$$

$$\frac{-x^2\alpha^2 + (x-\mu)^2}{2\alpha^2} = \frac{1}{2}\ln\left(\frac{\cancel{2\pi}}{\cancel{2\pi}\alpha^2}\right)$$

$$\boxed{-x^2\alpha^2 + (x-\mu)^2 = \alpha^2 \ln\left(\frac{1}{\alpha^2}\right)}$$

Scanned with CamScanner

0000000

③

M Tu W Th F S Su

DD MM YYYY

(b) $\mu = 1$ $\quad \sigma^2 = 2$

$$-2x^2 + (x-1)^2 = 2\ln\left(\frac{1}{2}\right)$$

$$-2x^2 + x^2 - 2x + 1 = 2\ln\left(\frac{1}{2}\right)$$

$$-x^2 - 2x + 1 = -1.386294$$

$$x^2 + 2x - 1 = 1.386294$$

$$x^2 + 2x - 2.386294$$

solutions of this quadratic equation are

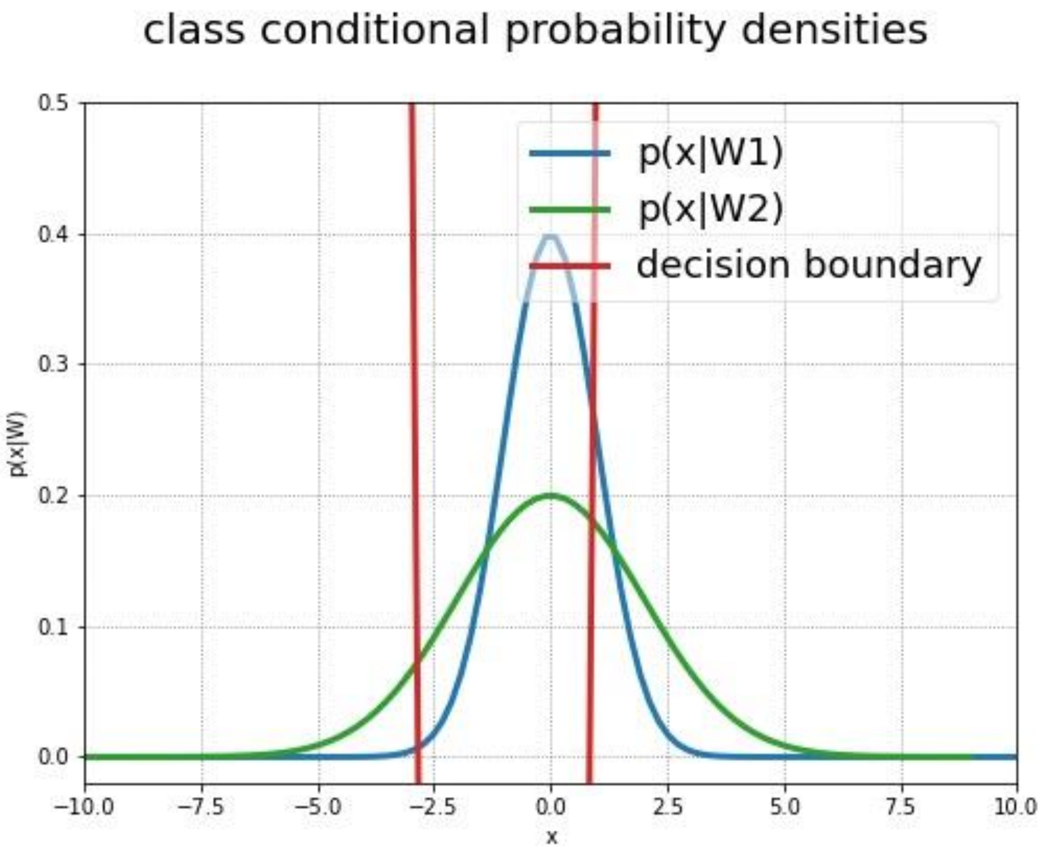
-2.8401

0.8401

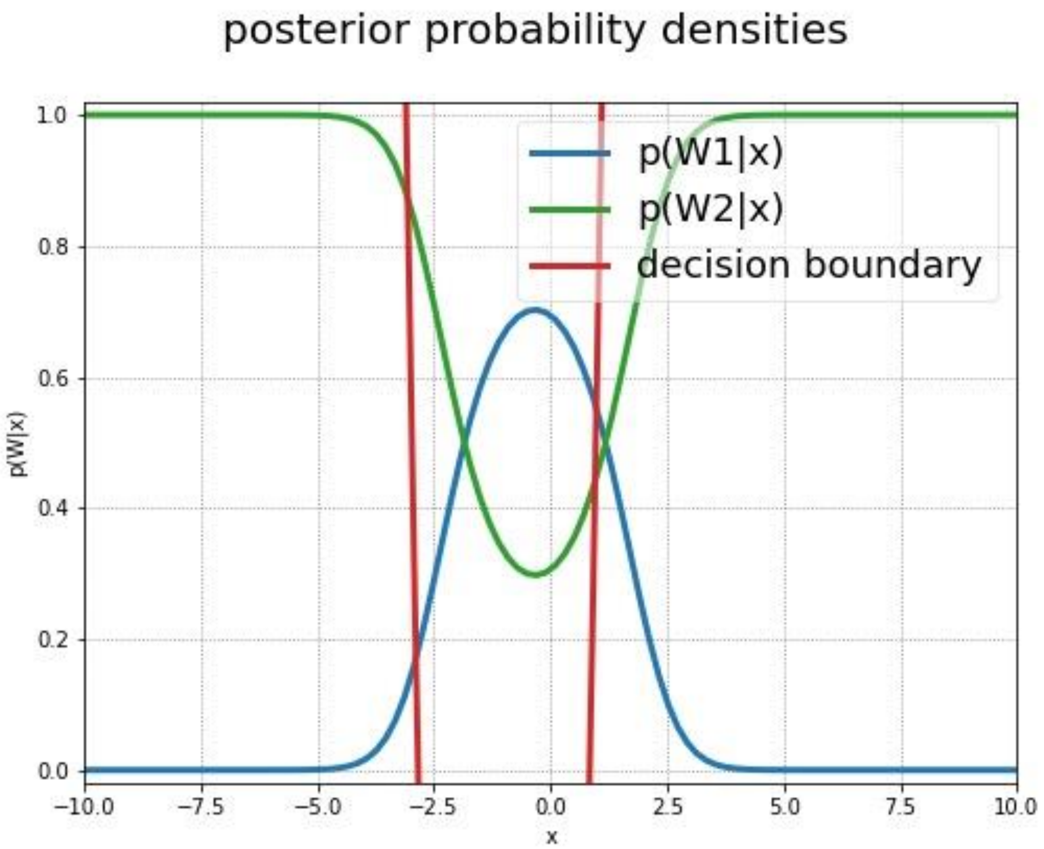
$$N(\mu_1, \sigma_1^2) = \frac{1}{\sqrt{2\pi}} e^{\frac{-(x)^2}{2}}$$

$$N(\mu_2, \sigma_2^2) = \frac{1}{\sqrt{4\pi}} e^{\frac{-(x-1)^2}{4}}$$

0000000



class conditional probability densities
p(x|W1)
p(x|W2)
decision boundary
p(x|W)
x



posterior probability densities
p(W1|x)
p(W2|x)
decision boundary
p(W|x)
x

(c) Bayes error rate

$$\int_{-\infty}^{-2.8401} p(\omega_1|x)\,p(x)\,dx$$

$$+ \int_{-2.8401}^{0.8401} p(\omega_2|x)\,p(x)\,dx$$

$$+ \int_{0.8401}^{\infty} p(\omega_1|x)\,p(x)\,dx$$

$$= \int_{-\infty}^{-2.8401} p(x|\omega_1)\,p(\omega_1)\,dx + \int_{-2.8401}^{0.8401} p(x|\omega_2)\,p(\omega_2)\,dx$$

$$+ \int_{0.8401}^{\infty} p(x|\omega_1)\,p(\omega_1)\,dx$$

$$= \frac{1}{2}\times\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-2.8401} e^{-\frac{x^2}{2}}\,dx + \frac{1}{2}\times\frac{1}{\sqrt{4\pi}}\int_{-2.8401}^{0.8401} e^{-\left(\frac{x-1}{4}\right)^2}\,dx$$

$$+ \frac{1}{2}\times\frac{1}{\sqrt{2\pi}}\int_{0.8401}^{\infty} e^{-\frac{x^2}{2}}\,dx$$

$$= 0.00113 + 0.22658 + 0.1002$$

0000000

M Tu W Th F S Su

D D M M Y Y Y Y

$$= \underline{\underline{0.327179}}$$

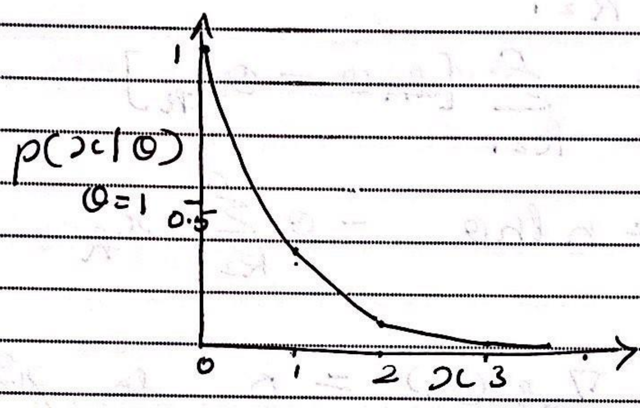
Scanned with CamScanner

Q2

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

a) $\theta = 1$

$p(x|\theta)$ vs x



$p(x|\theta)$ vs θ

θ $(0 \leq \theta \leq 5)$

$x = 2$

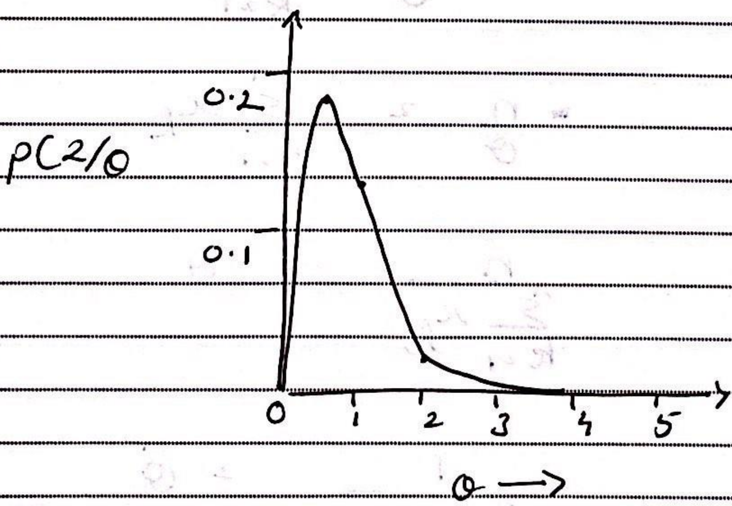
$0.5 e^{-\frac{1}{2}\times 2}$ $0.5 \times e^{-1}$

e^{-2}

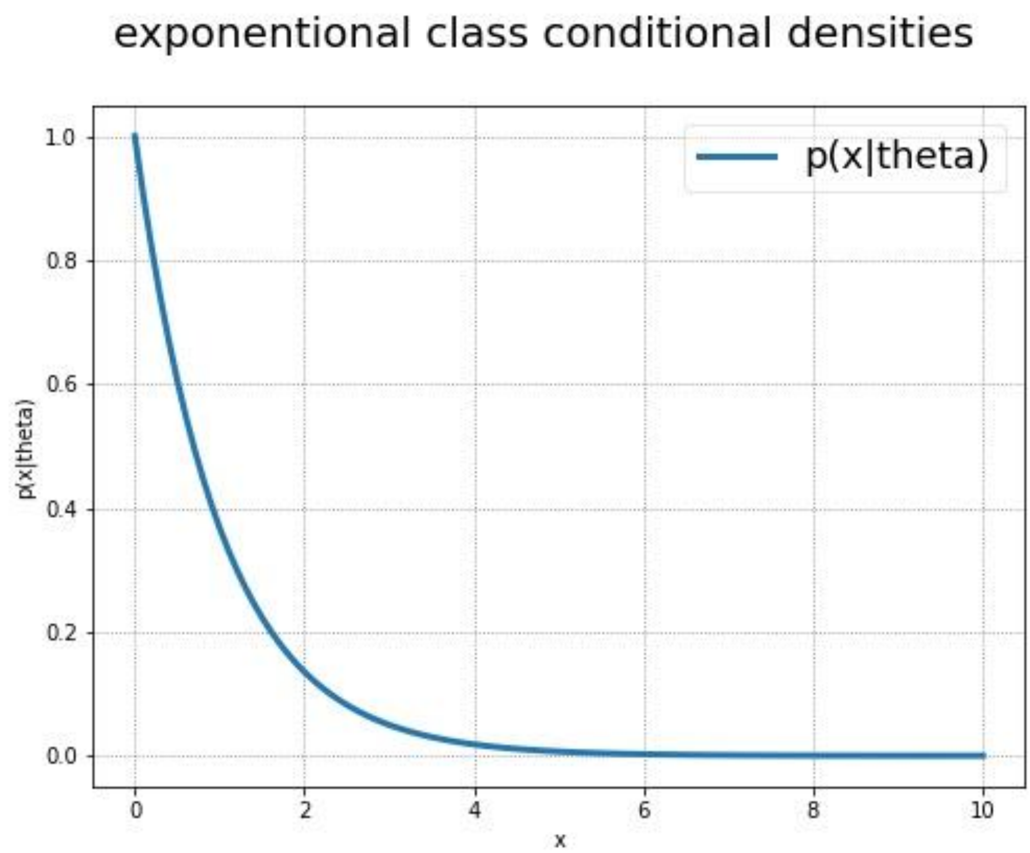
$2 e^{-2\times 2}$

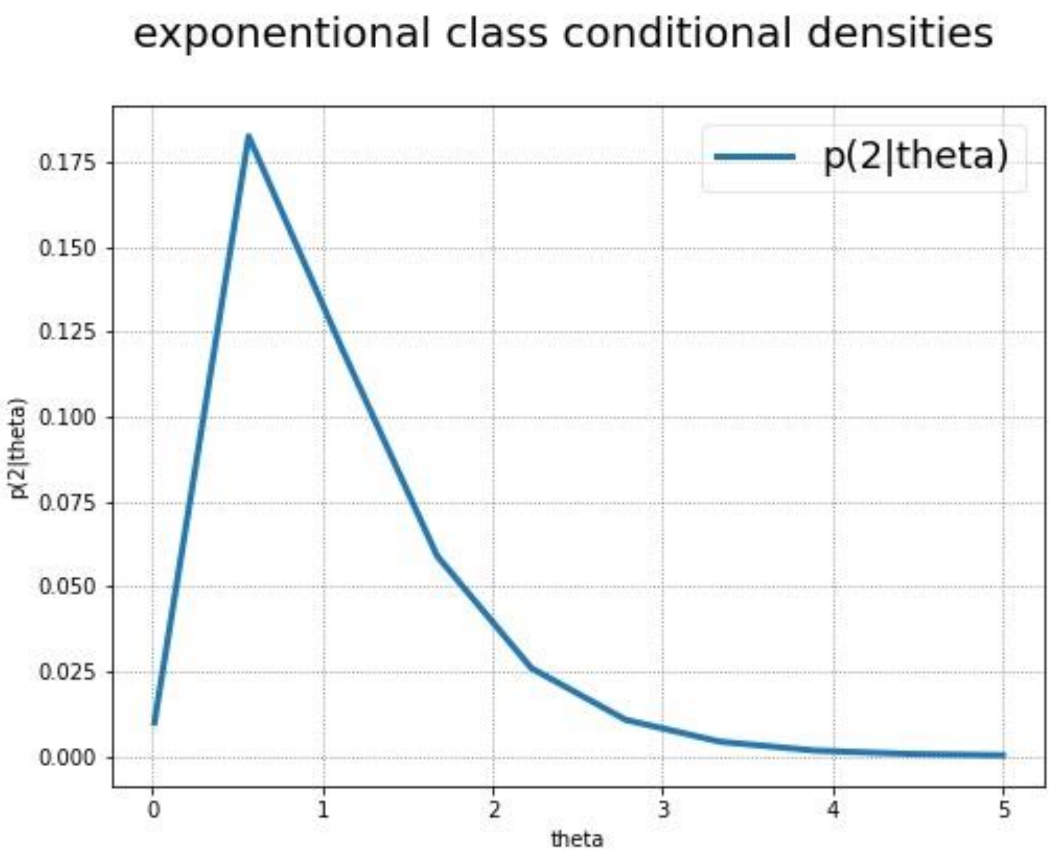
$3 e^{-3\times 2}$

$4 e^{-2}$



$p(x=2/\theta)$ is maximised when θ has a value less than 1.0


exponentional class conditional densities
p(x|theta)
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
x


exponentional class conditional densities
p(2|theta)
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0
1
2
3
4
5
theta

(7)

(b) $x_1 \ldots\ldots x_n$
maximum likelihood estimation for θ

log likelihood function is

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

$$= \sum_{k=1}^{n} [\ln\theta - \theta x_k]$$

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

we solve $\nabla_\theta l(\theta) = 0$ to find $\hat{\theta}$

$$\nabla_\theta l(\theta) = \frac{d}{d\theta}\left[n\log\theta - \theta\sum_{k=1}^{n} x_k\right]$$

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

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

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

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

⑧

Q 3 $p(x/\theta) = \begin{cases} 1/\theta & 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases}$

a) lets use an indicator function $I(\cdot)$ whose value is equal 1 if the logical value of its arguments is TRUE & 0 otherwise

Writing the likelihood function using $I(\cdot)$:

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

$$= \prod_{k=1}^{n} \frac{1}{\theta} I(0 \le x_k \le \theta)$$

$$= \frac{1}{\theta^n} I(\theta \geq \max_k x_k) \times I(\min_k x_k \geq 0)$$

from the above equation we observe that $\frac{1}{\theta^n}$ decreases monotonically as θ increases. And also $I(\theta \geq \max_k x_k)$ is 0 if θ is less than the maximum value of x_k. Therefore our likelihood function is maximized at

$$\underline{\hat{\theta} = \max_k x_k}$$

Q4) (a) To prove :

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

given :

$$z_{ik} = \begin{cases} 1 & \text{if state of nature of } k^{th} \text{ sample } \text{~~element~~} \text{ is } \omega_i \\ 0 & \text{otherwise} \end{cases}$$

The samples are drawn by successive independent selection of a state of nature ω_i with probability $P(\omega_i)$

$$P[z_{ik} = 1 / P(\omega_i)] = P(\omega_i)$$
$$P[z_{ik} = 0 / P(\omega_i)] = 1 - P(\omega_i)$$

The above equations can be unified as

$$P(z_{ik} / P(\omega_i)) = [P(\omega_i)]^{z_{ik}} [1 - P(\omega_i)]^{1-z_{ik}}$$

going by IID assumptions, we have

$$P(z_{i1} \ldots z_{in} / P(\omega_i)) = \prod_{k=1}^{n} P(z_{ik} / P(\omega_i))$$
$$= \prod_{k=1}^{n} [P(\omega_i)]^{z_{ik}} [1 - P(\omega_i)]^{1-z_{ik}}$$

(10)

(b) The log likelihood as a function of $P(\omega_i)$ is

$$l(P(\omega_i)) = \ln P(z_{i1}, \ldots, z_{in} \mid P(\omega_i))$$

$$= \ln \prod_{k=1}^{n} [P(\omega_i)]^{z_{ik}} [1 - P(\omega_i)]^{1-z_{ik}}$$

$$= \sum_{k=1}^{n} \left[z_{ik} \ln P(\omega_i) + (1 - z_{ik}) \ln (1 - P(\omega_i)) \right]$$

Therefore maximum-likelihood must satisfy

$$\nabla_{P(\omega_i)} l(P(\omega_i)) = \frac{1}{P(\omega_i)} \sum_{k=1}^{n} z_{ik} - \frac{1}{1 - P(\omega_i)} \sum_{k=1}^{n} (1 - z_{ik}) = 0$$

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

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

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

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

ie, the estimate of the probability of category ω_i is the probability of obtaining its indicatory value in the training data, just as we would expect

Q.5) x - d dimensional binary multivariate Bernoulli distribution.

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

$$\theta = (\theta_1 \ldots\ldots \theta_d)^t$$

$$P(x_i/\theta_i) = \begin{cases} \theta_i & x_i = 1 \\ 1-\theta_i & x_i = 0 \end{cases}$$

To ~~prove~~ prove:

max likelihood estimate for

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

---

likelihood for n samples is

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

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

12

To find max likelihood we set
$$\nabla_\theta l(\theta) = 0$$

since $\theta = (\theta_1 \dots \theta_d)^t$

we get
$$[\nabla_\theta l(\theta)]_i = \nabla_{\theta_i} l(\theta)$$
$$= \frac{1}{\theta_i} \sum_{k=1}^{n} x_{ki} - \frac{1}{(1-\theta_i)} \sum_{k=1}^{n} (1-x_{ki})$$
$$= 0$$

This implies that for any i
$$\frac{1}{\hat{\theta}_i} \sum_{k=1}^{n} x_{ki} = \frac{1}{(1-\theta_i)} \sum_{k=1}^{n} (1-x_{ki})$$

$$(1-\hat{\theta}_i) \sum_{k=1}^{n} x_{ki} = \hat{\theta}_i \left[n - \sum_{k=1}^{n} x_{ki} \right]$$

$$\sum_{k=1}^{n} x_{ki} - \cancel{\hat{\theta}_i \sum_{k=1}^{n} x_{ki}} = n\hat{\theta}_i - \cancel{\hat{\theta}_i \sum_{k=1}^{n} x_{ki}}$$

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

since this solution is valid for all $i = 1 \dots d$, we can write in vector form as

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

$\therefore$ we can interpret this equation as $\rightarrow$ the maximum likelihood value for θ is the sample mean

Q6) $P(x/D) = P(x/D_i; \omega_i)$
$P(x/\theta) = \prod_{i=1}^{d} \theta_i^{x_i} (1-\theta)^{1-x_i}$

$D = \{x_1 \dots x_n\}$

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

To prove : $P(D/\theta) = \prod_{i=1}^{d} \theta_i^{s_i} (1-\theta_i)^{n-s_i}$

lets denote $x_k = (x_{k1} \dots x_{kd})^t$

for $k = (1 \dots n)$

then $s_i = \sum_{k=1}^{n} x_{ki}$ where $i = 1 \dots d$

(4)

likelihood $= P(D/\theta)$

$$= P(x_1 \dots x_n/\theta)$$

$$= \prod_{k=1}^{n} P(x_k/\theta)$$

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

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

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

(b) uniform prior distribution for θ

$$\text{G} \int_0^1 \theta^m (1-\theta)^n \, d\theta = \frac{m!\, n!}{(m+n+1)!}$$

To prove :

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

as mentioned in the question that θ takes a uniform distribution hence we assume $p(\theta) = 1$ for $0 \le \theta_i \le 1$ where $i = 1 \dots d$

By Bayes Theorem :

$$P(\theta/D) = \frac{P(D/\theta) \times P(\theta)}{P(D)}$$

from (a) we know $P(D/\theta) = \prod_{i=1}^{d} \theta_i^{s_i} (1-\theta_i)^{n-s_i}$

probability density of D is

$$p(D) = \int P(D/\theta) p(\theta) d(\theta)$$

$$= \int \prod_{i=1}^{d} \theta_i^{s_i} (1-\theta_i)^{n-s_i} d\theta$$

$$= \int_0^1 \cdots\cdots \int_0^1 \prod_{i=1}^{d} \theta_i^{s_i} (1-\theta_i)^{n-s_i} d\theta_1 d\theta_2 \ldots d\theta_d$$

$$= \prod_{i=1}^{d} \int_0^1 \theta_i^{s_i} (1-\theta_i)^{n-s_i} d\theta$$

By using the property

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

By consolidating these results we get

$$p(\theta/D) = \frac{P(D/\theta)\, P(\theta)}{P(D)}$$

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

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

Scanned with CamScanner

(c) $d = 1 \quad n = 1$

two resulting possibilities for s_1

$$P(\theta_1/D) = \frac{2!}{s_1!\,(n-s_1)!}\,\theta_1^{s_1}(1-\theta_1)^{n-s_1}$$

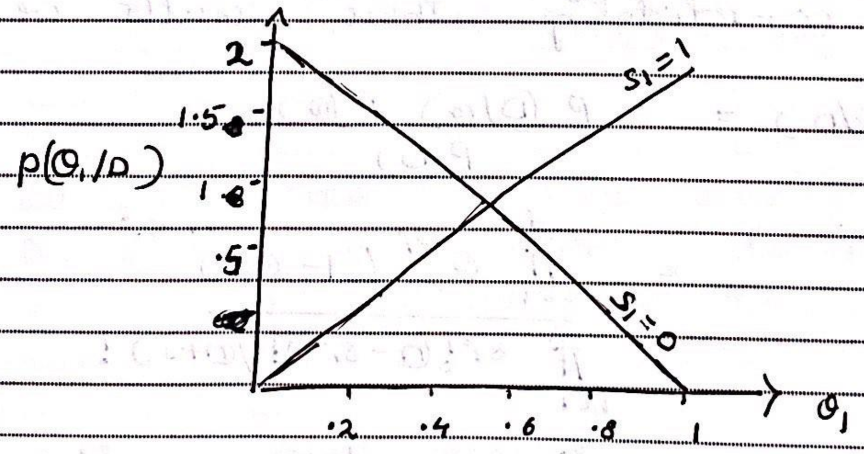
$$= \frac{2}{s_1!\,(1-s_1)!}\,\theta_1^{s_1}(1-\theta_1)^{1-s_1}$$

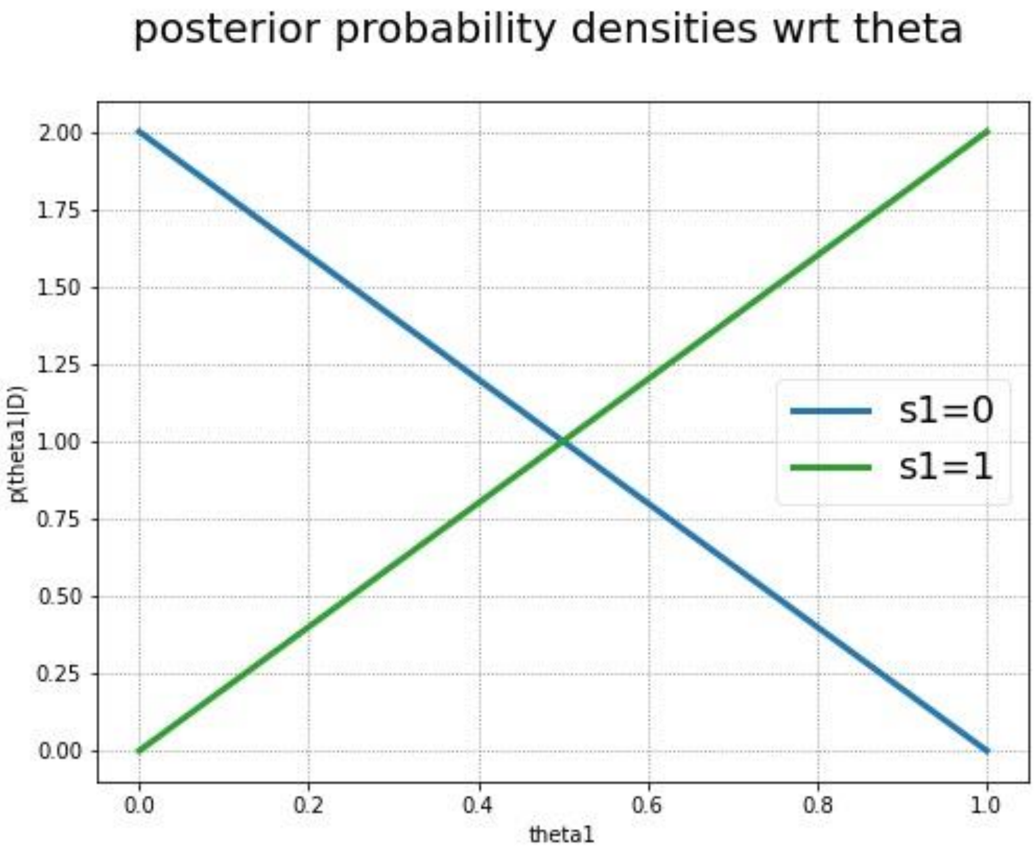
note : s_1 takes two discrete values as it is a bernoulli distribution those are 0 & 1

Thus densities are of the form

$s_1 = 0 \rightarrow p(\theta_1/D) = 2(1-\theta_1)$
$s_1 = 1 \rightarrow p(\theta_1/D) = 2(\theta_1)$

for θ_1 that range from 0 to 1 the densities are shown below




posterior probability densities wrt theta
p(theta1|D)
theta1
s1=0
s1=1
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.0
0.2
0.4
0.6
0.8
1.0

d) $P(x|D) = \prod_{i=1}^{d} \left[\frac{s_i+1}{(n+2)}\right]^{x_i} \left(1 - \frac{s_i+1}{n+2}\right)^{1-x_i}$

$$P(x|D) = \int_0^1 P(x|\theta)\, P(\theta|D)$$

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

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

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

$$= \prod_{i=1}^{d} \frac{(n+1)!}{s_i!\,(n-s_i)!} \times \frac{(s_i+x_i)!\,(1-x_i+n-s_i)!}{(n+2)!}$$

$$= \prod_{i=1}^{d} \frac{\cancel{(n+1)!}}{s_i!\,(n-s_i)!} \times \frac{(s_i+x_i)!\,(1-x_i+n-s_i)!}{\cancel{(n+1)!}\,(n+2)}$$

taking $x = 0$

$$= \frac{\cancel{s_i!}\,(n-s_i+1)!}{\cancel{s_i!}\,(n-s_i)!\,(n+2)} = \frac{(n-s_i+1)}{(n+2)}$$

$$= \left(1 - \frac{s_i+1}{n+2}\right)$$

(18)

taking $x = 1$

$$= \frac{(s_i+1)!\,\cancel{(n-s_i)!}}{(s_i)!\,\cancel{(n-s_i)!}\,(n+2)}$$

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

$$P(x|D) = \prod_{i=1}^{d} \left(\frac{s_i+1}{n+2}\right)^{x_i} \left(1 - \frac{s_i+1}{n+2}\right)^{1-x_i}$$

e) substituting $\hat{\theta}$ for θ in $P(x|\theta)$

Then we can obtain $P(D|\theta)$

$$= \prod_{i=1}^{d} \hat{\theta}_i^{s_i} (1-\hat{\theta}_i)^{n-s_i}$$

which is the kernel of a beta distribution with parameters

$$B(\alpha = s_i+1, \beta = n - s_i + 1)$$

The bayes estimator θ with squared error is simply the posterior mean (mean of a beta distribution with above parameters

$$\theta_{bayes} = E(\theta|D)$$
$$= \frac{\alpha}{\alpha+\beta}$$

Scanned with CamScanner

$$= \frac{s_i + 1}{s_i + 1 + n - s_i + 1}$$

$$= \frac{s_i + 1}{n+2}$$

$$= \frac{\sum_i X_i + 1}{n+2}$$