MACHINE LEARNING

ELİF HANGÜL

2.7) Bill tosses a bent coin N times, obtaining a sequence of heads and tails. We assume that the coin has a probability f_H of coming up heads; we do not know f_H. If n_H heads have occured in N tosses, what is the probability distribution of f_H? What is the probability that the $N+1$ th outcome will be a head?

2.8) Assuming a uniform prior on f_H, $P(f_H)=1$, solve the problem posed in example 2.7. Sketch the posterior distribution of f_H and compute the probability that the $N+1$ th outcome will be a head, for

(a) $N=3$ and $n_H=0$

(b) $N=3$ and $n_H=2$

(c) $N=10$ and $n_H=3$

(d) $N=300$ and $n_H=29$.

You will find the beta integral useful:

$$\int_0^1 dp_a\, p_a^{F_a}(1-p_a)^{F_b} = \frac{\Gamma(F_a+1)\Gamma(F_b+1)}{\Gamma(F_a+F_b+2)} = \frac{F_a!\,F_b!}{(F_a+F_b+1)!}$$

Answer:

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

- Prior : $P(f_H)=1$
- Evidence : $P(n_H \mid N) = \int_0^1 df_H\, f_H^{n_H}(1-f_H)^{N-n_H}$

→ From the beta integral we will have

$$P(n_H \mid N) = \frac{n_H!\,(N-n_H)!}{(n_H+N-n_H+1)!} = \frac{n_H!\,(N-n_H)!}{(N+1)!}$$

• The likelihood : $P(n_H | f_H, N) = f_H^{n_H} (1-f_H)^{N-n_H}$

→ So the posterior probability will be

$$P(f_H | n_H, N) = \frac{f_H^{n_H} (1-f_H)^{N-n_H} \times 1}{\frac{n_H! (N-n_H)!}{(N+1)!}}$$

→ To find the probability that the N+1 th toss will be a head, we integrate over f_H. With the help of sum rule we get

$$P(h | n_H, N) = \int d_{f_H} P(h|f_H) \, P(f_H | n_H, N)$$

$$= \int_0^1 d_{f_H} P(h|f_H) \frac{f^{n_H} (1-f_H)^{N-n_H}}{\frac{n_H! (N-n_H)!}{(N+1)!}}$$

→ Since $P(h|f_H) = f_H$

$$P(h | n_H, N) = \int_0^1 d_{f_H} \frac{f^{n_H+1} (1-f_H)^{N-n_H}}{\frac{n_H! (N-n_H)!}{(N+1)!}}$$

$$= \frac{(N+1)!}{n_H! (N-n_H)!} \int_0^1 d_{f_H} \, f^{n_H+1} (1-f_H)^{N-n_H}$$

→ Using the beta integral again

$$P(h | n_H, N) = \frac{(N+1)!}{n_H! (N-n_H)!} \cdot \frac{(n_H+1)! \, (N-n_H)!}{(N+2)!}$$

$$= \frac{n_H + 1}{N+2}$$

→ Sketch of posterior distribution of f_H

posterior density $\propto$ Likelihood $\times$ Prior density.

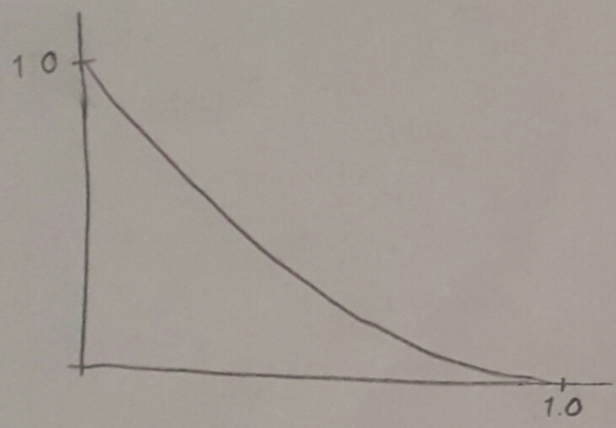
Since prior of f_H is 1 and likelihood is $f_H^{n_h}(1-f_H)^{N-n_h}$

• posterior distribution of f_H is $f_H^{n_h}(1-f_H)^{N-n_h}$

→ Now let's sketch for four options. For all sketches y = Posterior distribution outcome of given f_H, x is f_H

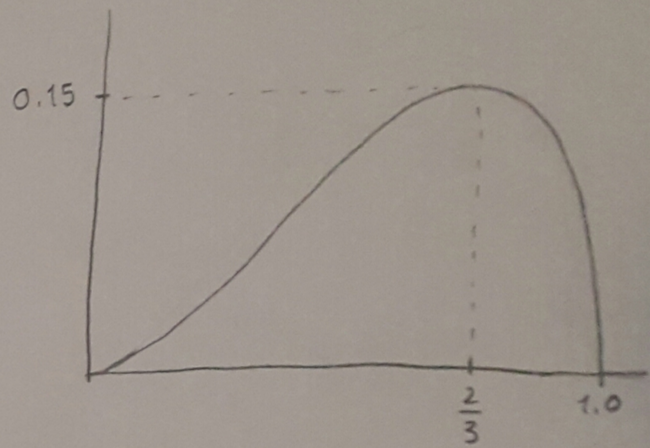
a) $N=3$ $n_h=0$

Sketch $f_H^{0}(1-f_H)^{3}$



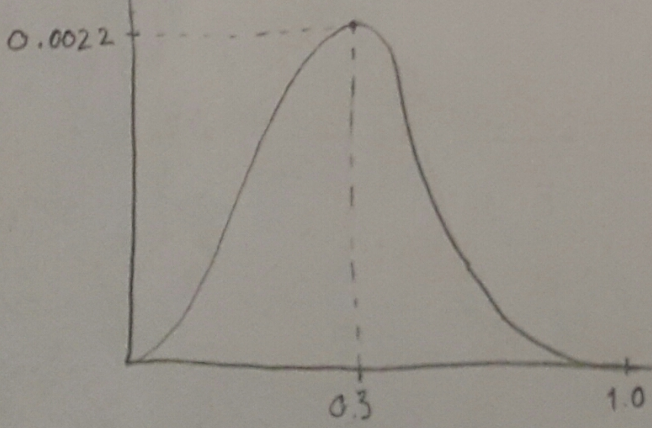
b) $N=3$ $n_h=2$

Sketch $f_H^{2}(1-f_H)^{1}$



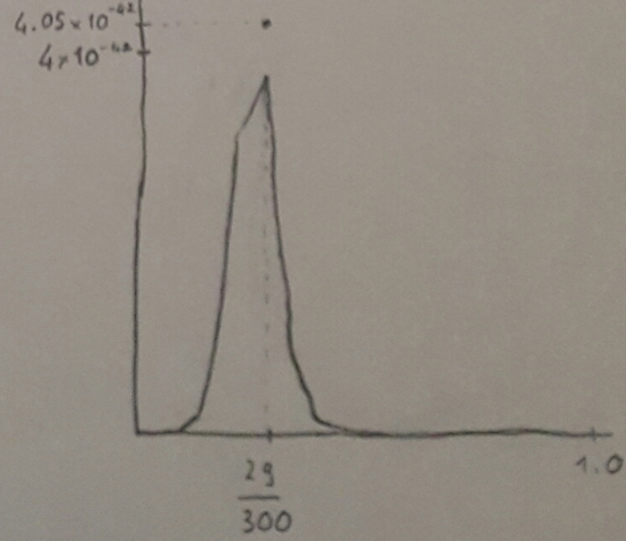
c) $N=10$ $n_h=3$

Sketch $f_H^{3}(1-f_H)^{7}$



d) $N=300$ $n_H=29$

Sketch $f_H^{29}(1-f_H)^{271}$



→ Put the values inplace

(a) $N = 3$ and $n_H = 0$

$$\frac{n_H + 1}{N + 2} = \frac{0 + 1}{3 + 2} = \frac{1}{5}$$

(b) $N = 3$ and $n_H = 2$

$$\frac{n_H + 1}{N + 2} = \frac{2 + 1}{3 + 2} = \frac{3}{5}$$

(c) $N = 10$ and $n_H = 3$

$$\frac{n_H + 1}{N + 2} = \frac{3 + 1}{10 + 2} = \frac{4}{12} = \frac{1}{3}$$

(d) $N = 300$ and $n_H = 29$

$$\frac{n_H + 1}{N + 2} = \frac{29 + 1}{300 + 2} = \frac{30}{302} = \frac{15}{151}$$

2.10) Urn A contains three balls: one black and two white; urn B contains three balls: two black and one white. One of the urns is selected at random and one ball is drawn. The ball is black. What is the probability that the selected urn is urn A?

Answer: Let E_A, E_B denote the events of selecting urns A and B respectively.

$P(E_A) = P(E_B) = \frac{1}{2}$

→ $P(E_A | B)$ is asked. For that let's find $P(B | E_A)$ and $P(B | E_B)$ which denote the probability of a black ball selection from urns A and B respectively.

$P(B | E_A) = \frac{1}{3}$, $P(B | E_B) = \frac{2}{3}$

→ By Bayes Theorem

$$P(E_A|B) = \frac{P(E_A)\,P(B|E_A)}{P(E_A)\,P(B|E_A) + P(E_B)\,P(B|E_B)}$$

$$= \frac{\frac{1}{2}\cdot\frac{1}{3}}{\frac{1}{2}\cdot\frac{1}{3} + \frac{1}{2}\cdot\frac{2}{3}} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}$$,,

2.11) Urn A contains five balls : one black, two white, one green and one pink; urn B contains five hundred balls : two hundred black, one hundred white, 50 yellow, 40 cyan, 30 sienna, 25 silver, 20 gold and 10 purple. One of the urns is selected at random and one ball is drawn. The ball is black. What is the possibility that the urn is urn A?

Answer: Same steps in solution for 2.10 are applied.

→ Let E_A, E_B denote the events of selecting urns A and B respectively.

$P(E_A) = P(E_B) = \frac{1}{2}$

→ $P(E_A|B)$ is asked. $P(B|E_A)$ and $P(B|E_B)$ denote the probability of a black ball selection from urns A and B respectively.

$P(B|E_A) = \frac{1}{5}$, $P(B|E_B) = \frac{2}{5}$

→ By Bayes Theorem

$$P(E_A|B) = \frac{P(E_A)\,P(B|E_A)}{P(E_A)\,P(B|E_A) + P(E_B)\,P(B|E_B)}$$

$$= \frac{\frac{1}{2}\cdot\frac{1}{5}}{\frac{1}{2}\cdot\frac{1}{5} + \frac{1}{2}\cdot\frac{2}{5}} = \frac{\frac{1}{10}}{\frac{3}{10}} = \frac{1}{3}$$,,

3.5) Sketch the posterior probability $P(p_a | s = aba, F = 3)$.

What is the most probable value of p_a? What is the mean value of p_a under this distribution?

Answer the same questions for the posterior probability $P(p_a | s = bbb, F = 3)$

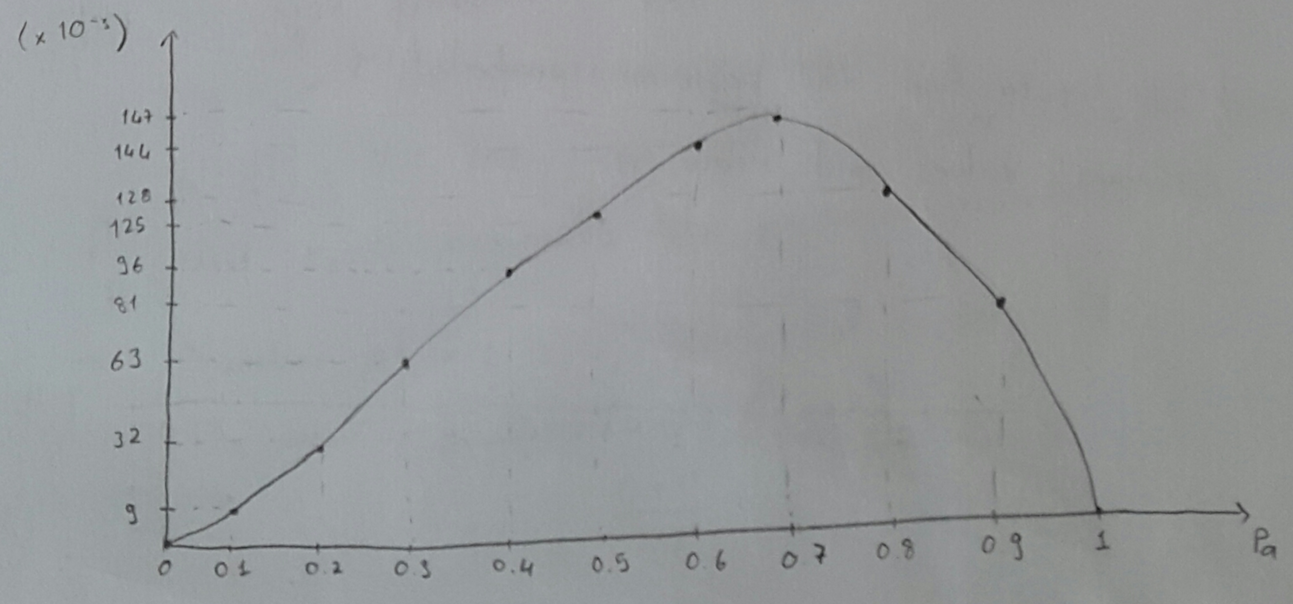
Answer: Let's start with $P(p_a | s = aba, F = 3)$

The sequence probability is $p_a \cdot p_b \cdot p_a$. Since $p_b = (1 - p_a)$ we can write the posterior probability as $p_a \cdot (1 - p_a) \cdot p_a$.

$P(p_a | s = aba, F = 3) = p_a^2 (1 - p_a)$

Let's try to find the posterior probability for different values and draw it.

p_a	Posterior Probability
0	0
0.1	0.009
0.2	0.032
0.3	0.063
0.4	0.096
0.5	0.125
0.6	0.144
0.7	0.147
0.8	0.128
0.9	0.081
1	0



→ The most probable value of p_a is the value that maximizes the posterior probability.
To find this point derivative of the function is needed.

- $p_a^2(1-p_a) = p_a^2 - p_a^3$

- $\frac{d}{dp_a}\left(p_a^2 - p_a^3\right) = 2p_a - 3p_a^2$

→ The value that makes the derivative equal to 0 is the value we are looking for.

$2p_a - 3p_a^2 = 0$

$2p_a = 3p_a^2$

$p_a = \frac{2}{3}$ //

→ Let a denote the number of a's in the sequence and n denote the number of times the coin is tossed.

The mean is : $\bar{p} = (a+1)/(n+2)$

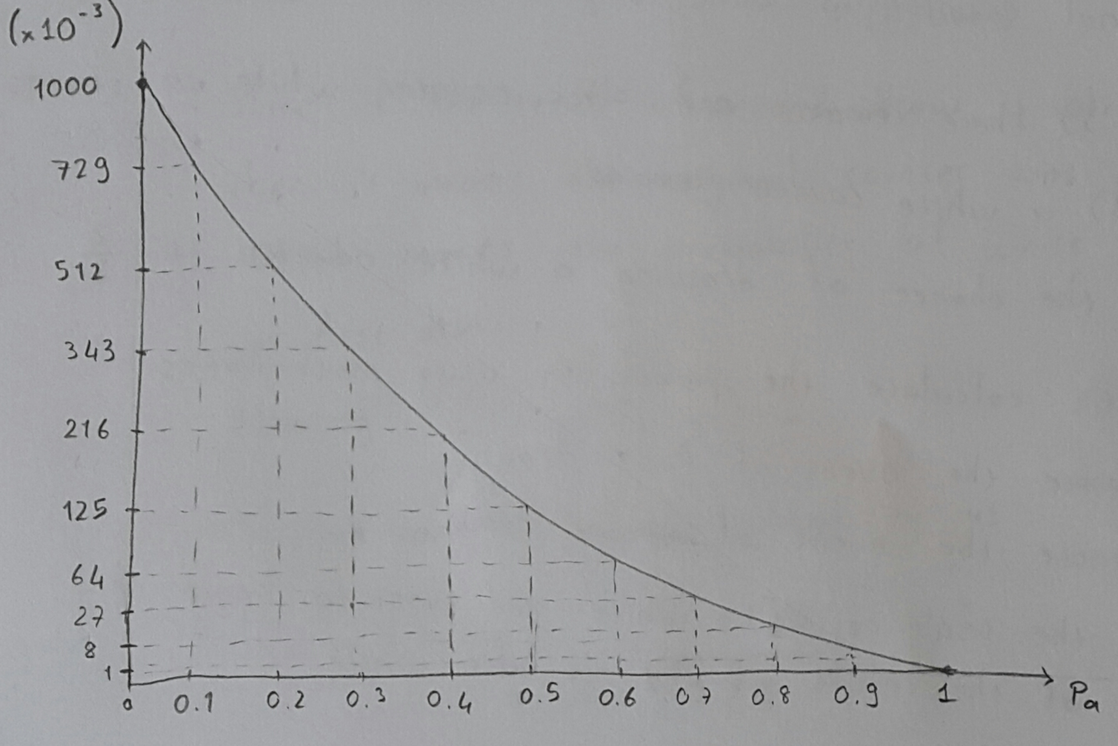
- In this case $a = 2$ and $n = 3$
- So the mean value of p_a is $\frac{2+1}{3+2} = \frac{3}{5}$ //

γ Now let's consider the posterior probability $P(p_a | s = bbb, F = 3)$.

- $P(p_a | s = bbb, F = 3) = (1-p_a).(1-p_a).(1-p_a)$
  $= (1-p_a)^3$

Let's try to find the posterior probability for different values and draw it.

P_a	Posterior Probability
0	1
0.1	0.729
0.2	0.512
0.3	0.343
0.4	0.216
0.5	0.125
0.6	0.064
0.7	0.027
0.8	0.008
0.9	0.001
1	0



→ The most probable value of P_a is the value that maximizes the posterior probability. $(1-p_a)^3$. We can guess and clearly see from the graph that $P_a = 0$ maximizes the posterior probability.

→ Number of a's is 0. Number of tosses is 3.

The mean is: $\overline{P_a} = (a+1)/(n+2)$

$$= \frac{0+1}{3+2} = \frac{1}{5}$$

3.12) A bag contains one counter, known to be either white or black. A white counter is put in, the bag is shaken, and a counter is drawn out, which proves to be white. What is now the chance of drawing a white counter? [Notice that the state of the bag, after the operations, is exactly identical to its state before]

Answer: At the end three states could be written for the bag.

1) • Original counter is black, new white counter is selected.

2) • Original counter is white, new white counter is selected.

3) • Original counter is white, original white counter is selected.

In state 1) there remains one black counter while in states 2) and 3) a white counter remains.

Therefore the chance of drawing a white counter is $\frac{2}{3}$ //

→ Now let's calculate the probability using mathematics.

Let E_1 denote the event of first draw.

Let E_2 denote the event of second draw.

We know the result of E_1 is white and want to know the probability of the result of E_2 is also white.

w: white

• First, calculation of $P(E_1 = w)$

$P(E_1 = w) = P[\text{original is white}] + P[\text{Original is black, white is selected}]$

→ The probability of the original black or white is $\frac{1}{2}$.

→ The probability of while original is black and the newly added white counter is selected $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

($\frac{1}{2}$ for the probability of the original counter is black

$\frac{1}{2}$ for selecting the white counter between one black and one white counter. It makes $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$)

→ According to these probabilities

$$P(E_1 = w) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$$

- Secondly, calculation of $P(E_2 = w)$

The only case when we draw a white counter on the second draw is when there was a white counter in the bag originally.

So $P(E_2 = w) = \frac{1}{2}$

→ Now let's write the equation.

$$P(E_2 = w \mid E_1 = w) = \frac{P(E_1 = w \mid E_2 = w) \cdot P(E_2 = w)}{P(E_1 = w]}$$

- Since if the result of second draw is white (E_2 is white) then it means the original counter was also white. Which makes the probability of white counter selection at first draw is 1.

Meaning $P(E_1 = w \mid E_2 = w) = 1$.

⇒ So we can write the probability as :

$$P(E_2 = w \mid E_1 = w) = \frac{1 \cdot 1/2}{3/4}$$

$$= \frac{2}{3}$$ //

3.14) In a game, two coins are tossed. If either of the coins comes up heads, you have won a prize. To claim the prize, you must point to one of your coins that is head and say "look, that coin's a head, I've won." You watch Fred play the game. He tosses the two coins, and he points to a coin and says "look, that coin's a head, I've won." What is the probability that the other coin is a head?

Answer: Let's start with writing the possible outcomes of coin toss.

	Coin 1	Coin 2
1)	Head	Head
2)	Head	Tail
3)	Tail	Head
4)	Tail	Tail

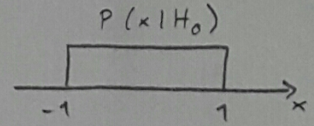
Since we know Fred points to a head, we can eliminate the 4th option. (At least one coin is head, eliminate no head option)

Only 1st option accounts for both coins get head and in total there are 3 options.

So the probability of the other coin is a head $\frac{1}{3}$ //

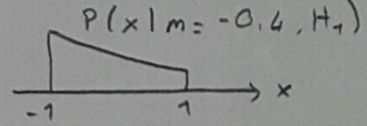
28.1) Random variables x come independently from a probability distribution $P(x)$. According to model H_0, $P(x)$ is a uniform distribution.

$P(x \mid H_0) = \frac{1}{2} \quad x \in (-1, 1)$



According to model H_1, $P(x)$ is a nonuniform distribution with an unknown parameter $m \in (-1, 1)$:

$P(x \mid m, H_1) = \frac{1}{2}(1 + mx) \quad x \in (-1, 1)$



Given the data $D = \{0.3, 0.5, 0.7, 0.8, 0.9\}$, what is the evidence for H_0 and H_1?

Answer: Let's start with defining evidence.
$P(D \mid H)$ is called the evidence of the model.
The ratio in evidence of data between two hypothesis gives the Bayes factor
If Bayes factor > 1 then model favors H_0, otherwise it favors H_1 $\left(\text{Bayes factor} = \frac{P(D \mid H_0)}{P(D \mid H_1)}\right)$

- To calculate the $P(D \mid H)$ value we need to enter our data points into the given distribution and product the results.
- Since $P(x \mid H_0) = \frac{1}{2}$ for every x in $(-1, 1)$, for all five data points we will have $\frac{1}{2}$.

$\rightarrow P(D \mid H_0) = \left(\frac{1}{2}\right)^5 = \frac{1}{32} = 0.03125$

- $P(D|H_1)$ is calculated from $P(x|m, H_1)$ distribution. This distribution has a m parameter.

Let's put our data points as x values.

$$P(D|H_1) = \left(\frac{1}{2}\right)^5 (1+0.3m)\cdot(1+0.5m)\cdot(1+0.7m)\cdot(1+0.8m)\cdot(1+0.9m)$$

To come up with a value let's integrate over the polynomial and give -1 as lower bound and +1 as upper bount.

$$P(D|H_1) = \frac{1}{32}\int_{-1}^{1}(1+0.3m).(1+0.5m)(1+0.7m).(1+0.8m)(1+0.9m)\,dm$$

$$= 0.15403$$

- So the evidence for H_0

$P(D|H_0) = 0.03125$

and

the evidence for H_1

$P(D|H_1) = 0.1540$

=> Now let's look into ratio of evidence of H_0 and H_1 to see which hypothesis fits better for this data.

$$\frac{P(D|H_0)}{P(D|H_1)} = \frac{0.3125}{0.1540} \approx 2.03$$

- Which means H_0 is a better hypothesis than H_1 for the given data

=> Now let's try one more taking $m = -0.4$ as given in the example graph.

$P(x|m, H_1)$ distribution.

- $P(D|H_1)$ is calculated from

This distribution has a m parameter and according to graph m is -0.4 for $x \in (-1, 1)$. Since all data (x) is in this interval we will take m as -0.4

$$P(x \mid -0.4, H_1) = \frac{1}{2}(1 - 0.4x)$$

→ Now we insert our x values into distribution

- $x = 0.3 \rightarrow \frac{1}{2}(1 - 0.4 \times 0.3) = 0.44$
- $x = 0.5 \rightarrow \frac{1}{2}(1 - 0.4 \times 0.5) = 0.4$
- $x = 0.7 \rightarrow \frac{1}{2}(1 - 0.4 \times 0.7) = 0.36$
- $x = 0.8 \rightarrow \frac{1}{2}(1 - 0.4 \times 0.8) = 0.34$
- $x = 0.9 \rightarrow \frac{1}{2}(1 - 0.4 \times 0.9) = 0.32$

→ Now to calculate $P(D \mid H_1)$:

$$P(D \mid H_1) = (0.44 \times 0.4 \times 0.36 \times 0.34 \times 0.32)$$

$$= 0.006893568$$

γ So the evidence for H_0

$P(D \mid H_0) = 0.03125$

and

the evidence for H_1

$P(D \mid H_1) = 0.006893568$

⇒ Now let's look into ratio of evidence of H_0 and H_1 to see which hypothesis fits better for this data.

$$\frac{P(D \mid H_0)}{P(D \mid H_1)} = \frac{0.03125}{0.006893568} = 4.53321$$

- Which means H_0 is a better hypothesis than H_1 for the given data