

Introduction to Bayesian Inference

Osamu Maruyama

<http://www.design.kyushu-u.ac.jp/~maruyama/summer-school-2019.pdf>

Contents

1. Bayes' theorem
2. Example: diagnosis
3. Bayes updating

Statistical inference



You lost a book yesterday.

A = “The book I lost yesterday should be on the desk of my house.”

A is 80%!

$$P(A) = 0.8.$$

Degree of belief.

Statistical inference

You guess today's dinner.

A = "The dinner today might be mapo dofu."

A is 10%!

$P(\text{dinner today} = \text{mapo})$



Bayesian inference

Technique of statistical inference using a posterior probability distribution

$$P(\theta|D)$$

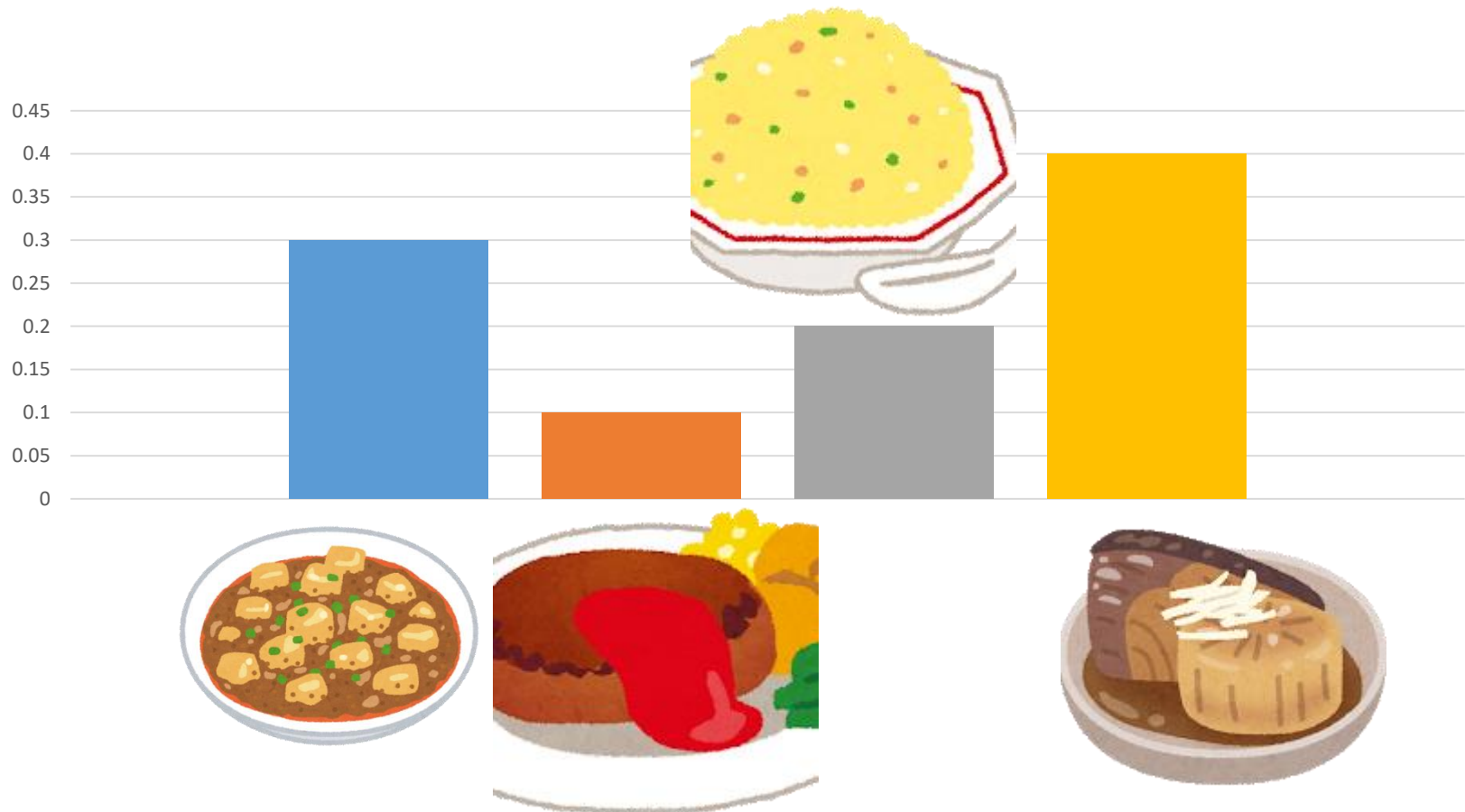
θ : random variable representing parameter, hypothesis

: **target objects to be estimated**

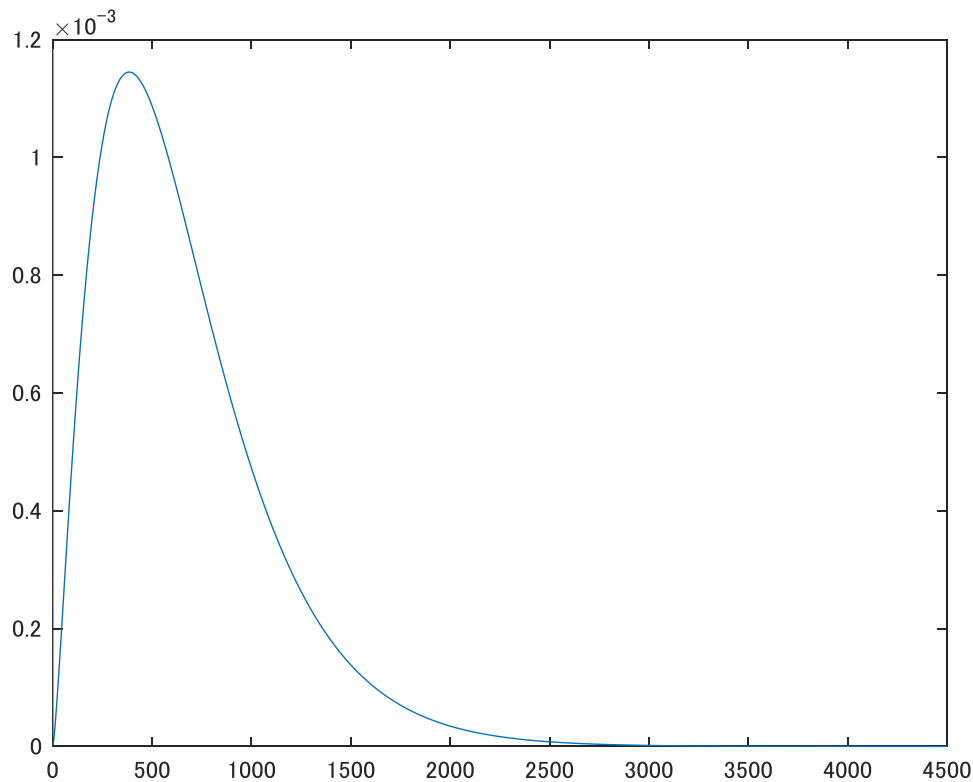
D : random variable representing observed data

: Observed data

Discrete random variable



Continuous random variable



Gamma distribution:

$$P(x|a, b) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-\frac{x}{b}}$$

A probability is a function satisfying

$$0 \leq P(x) \leq 1$$

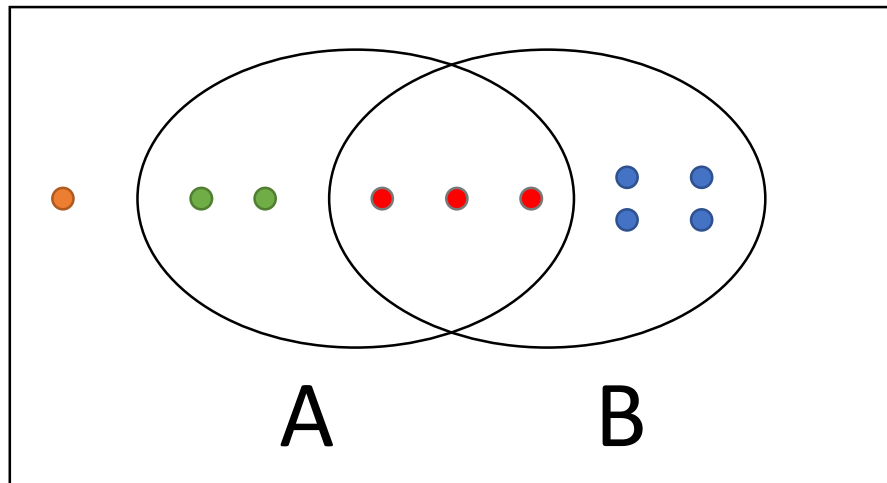
$$\sum_x P(x) = 1$$

Conditional probability

$$P(B|A) \quad (= P_A(B))$$

$$= \frac{P(A \cap B)}{P(A)}$$

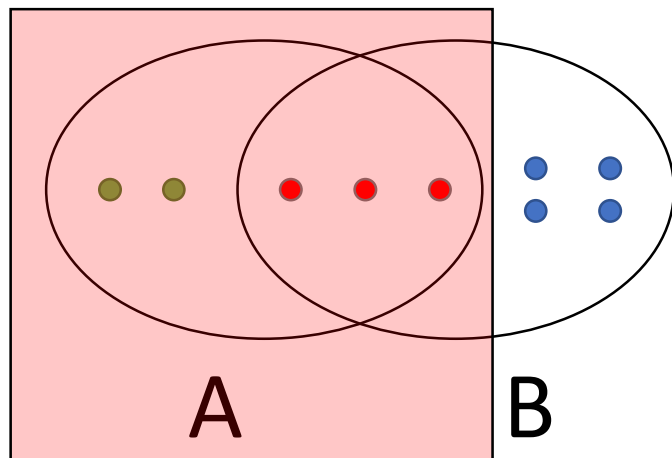
$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$



$$P(A) = \frac{5}{10}$$

$$P(B) = \frac{7}{10}$$

$$P(A \cap B) = \frac{3}{10}$$



$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{3}{10} \cdot \frac{10}{5} = \frac{3}{5}$$

← Compare the direct calculation of $P(B|A)$

Conditional probability

$$P(B|A)$$

$$(\quad = P_A(B))$$

$$= \frac{P(A \cap B)}{P(A)}$$

Multiplication theorem

$$P(A \cap B)$$

$$= P(B|A)P(A)$$

Bayes' theorem

Recall multiplication theorem:

$$P(A \cap B) = P(B|A)P(A)$$

$$\therefore P(A \cap B) = P(B \cap A) = P(A|B)P(B)$$

The two R.H.Ss are equal:

$$P(A|B)P(B) = P(B|A)P(A).$$

$$\therefore P(B|A) = \frac{P(A|B)P(B)}{P(A)}.$$

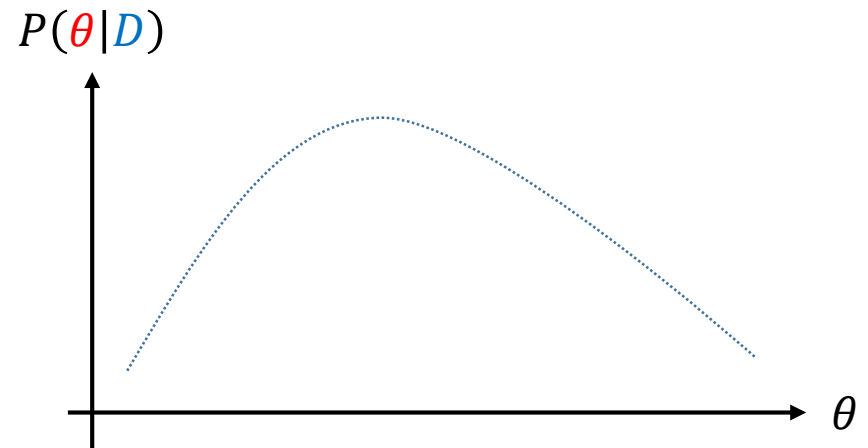
Key concepts of Bayes' theorem

Bayes' theorem

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

θ : parameter, hypothesis (representing causes)
: **target object to be estimated**

D : observed data



Bayes' theorem

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

θ : parameter, hypothesis (cause)

D : observed data

- $P(\theta|D)$
 - posterior probability (事後確率) .
 - interpretation: probability of cause θ when event D happens.
- $P(D|\theta)$
 - likelihood function (尤度関数)
 - interpretation: likelihood of D under θ .
- $P(\theta)$
 - prior probability (事前確率) .
 - interpretation: a general degree of belief in θ .

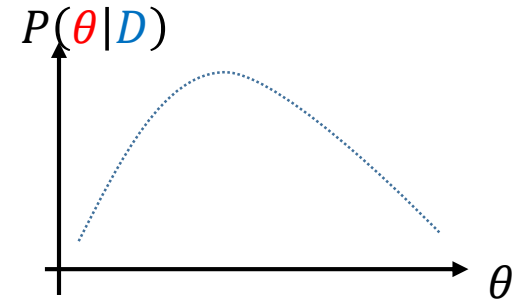
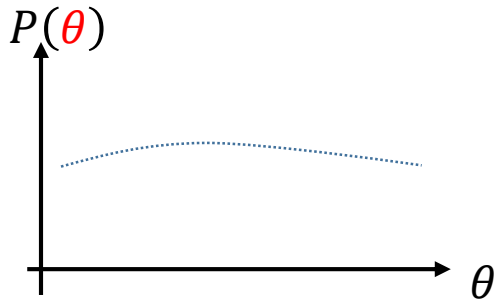
Usefulness of Bayes' theorem

Current knowledge:

$$P(\theta)$$

Updated knowledge:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$



Example: Diagnosis

Example: Diagnosis

You are diagnosed positive.

Question 1:

Do you believe that you are affected?

D i a g n o s i s	negative 10%	negative 80%	
	positive 90%		
		positive 20%	
Affected 罹患 0.001%		Unaffected 非罹患 99.999%	

Example: Diagnosis

You are diagnosed positive.

Question 2:

What's the probability that you are affected?

Luckily we have statistical data:

A = Affected

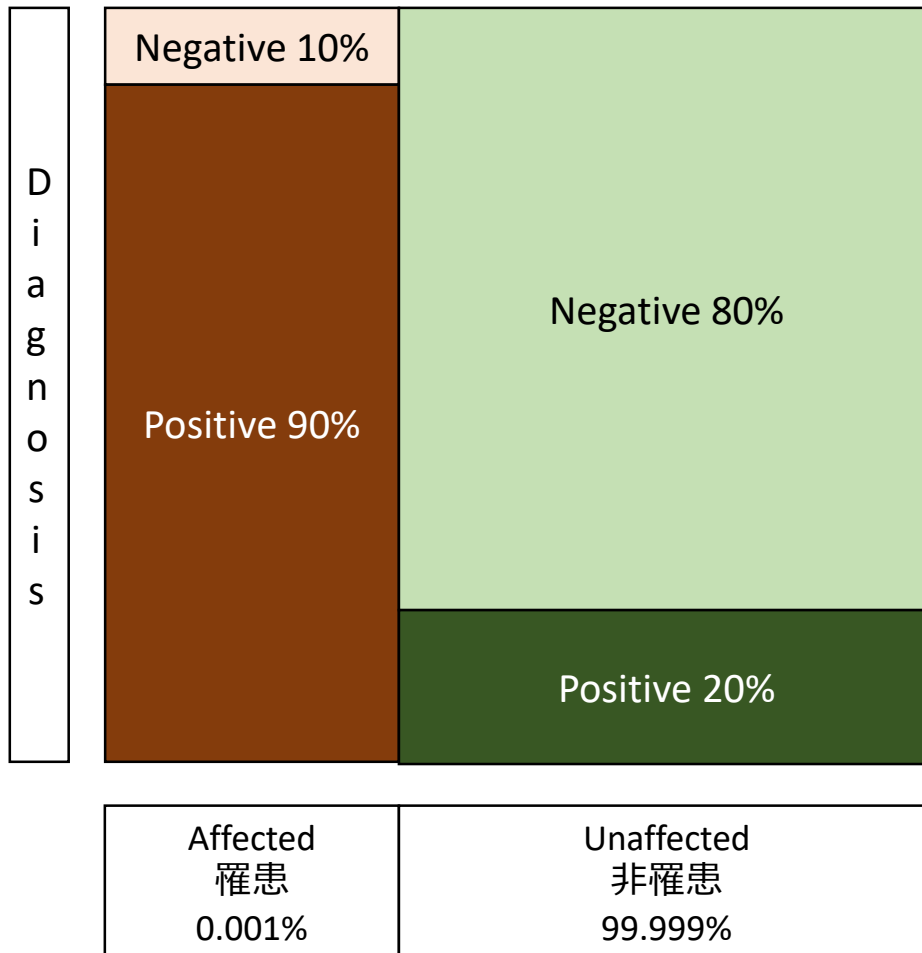
U = Unaffected

P = Positive

N = Negative

$D = \{P, N\}$: random variable for diagnose

$R = \{A, U\}$: random variable for real state



$$P(R = A) = 0.00001$$

$$P(R = U) = 0.99999$$

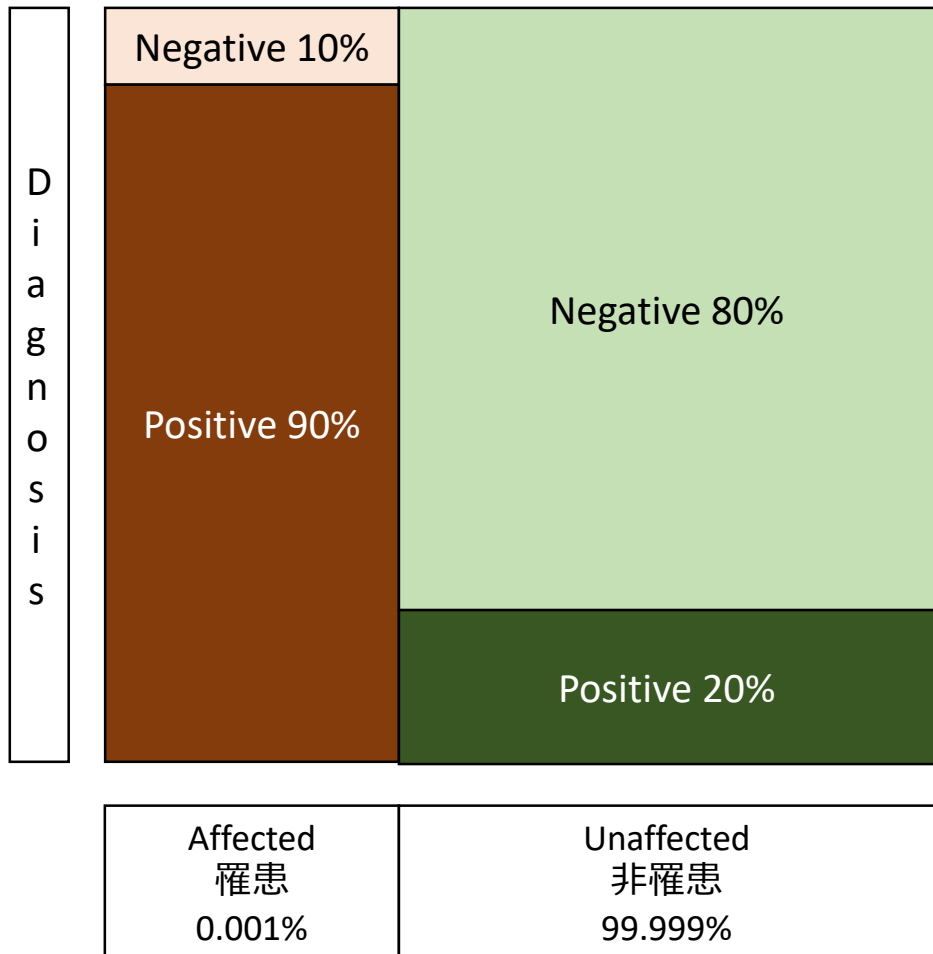
$$P(D = P|R = A) = 0.9$$

$$P(D = N|R = A) = 0.1$$

$$P(D = P|R = U) = 0.2$$

$$P(D = N|R = U) = 0.8$$

Example: Diagnosis



A = Affected

U = Unaffected

P = Positive

N = Negative

$D = \{P, N\}$: random variable for diagnose

$R = \{A, U\}$: random variable for real state

$$P(R = A) = 0.00001$$

$$P(R = U) = 0.99999$$

$$P(D = P|R = A) = 0.9$$

$$P(D = N|R = A) = 0.1$$

$$P(D = P|R = U) = 0.2$$

$$P(D = N|R = U) = 0.8$$

Which is larger?

$$P(R = A|D = P)$$

$$P(R = U|D = P)$$

$$\begin{aligned}P(R = A) &= 0.00001 \\P(R = U) &= 0.99999 \\P(D = P|R = A) &= 0.9 \\P(D = N|R = A) &= 0.1 \\P(D = P|R = U) &= 0.2 \\P(D = N|R = U) &= 0.8\end{aligned}$$

Using Bayes' theorem

$$\begin{aligned}P(R = A|D = P) &= \frac{P(D = P|R = A)P(R = A)}{P(D = P)} \\&= \frac{0.9 \cdot 0.00001}{P(D = P)} = \frac{0.000009}{P(D = P)}\end{aligned}$$

$$\begin{aligned}P(R = U|D = P) &= \frac{P(D = P|R = U)P(R = U)}{P(D = P)} \\&= \frac{0.2 \cdot 0.99999}{P(D = P)} \approx \frac{0.2}{P(D = P)}\end{aligned}$$

Bayesian updating



- θ : probability of getting the head of a coin when it is flipped.
- Evaluate the value of θ as posterior probabilities!
- Likelihood function:
 - Bernoulli (ベルヌーイ) distribution with parameter θ
 - H : Head
 - T : Tail
 - $p(H|\theta) = \theta$
 - $p(T|\theta) = 1 - \theta$

Initial step of Bayesian updating



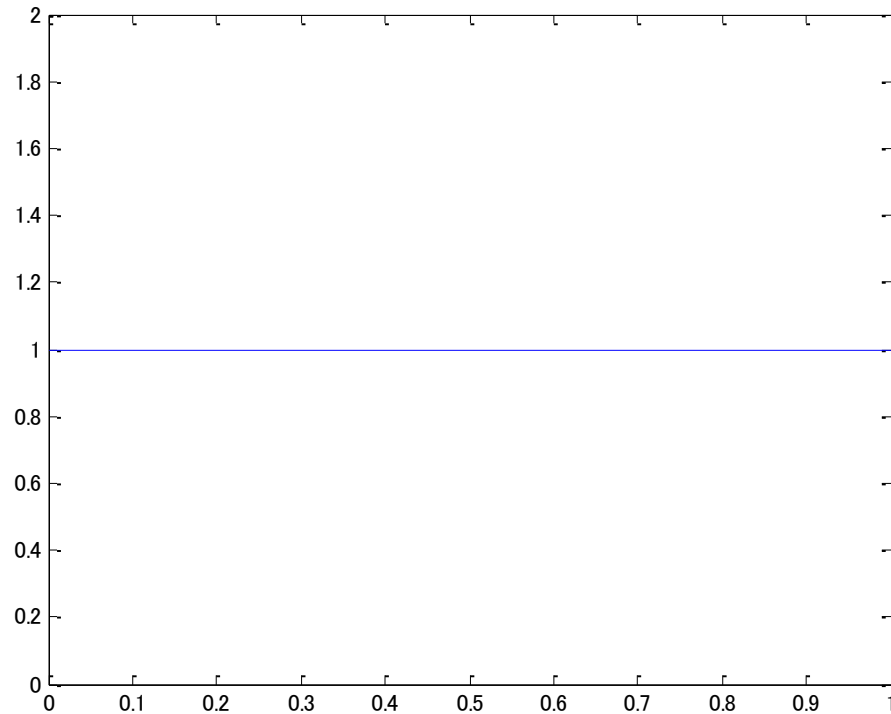
- No prior knowledge, assume a prior
- It is assumed $p(\theta) = \text{constant}$.

$$\therefore p(\theta) = 1$$

p is a uniform distribution.

Current (initial) prior distribution

$p(\theta) = 1$



This means we have no information on θ .

Suppose we have event
 D_1 : the head appeared

Find the posterior distribution $p(\theta|D_1)$
 $p(\theta|D_1) \propto p(H|\theta)p(\theta) = \theta \times 1 = \theta$

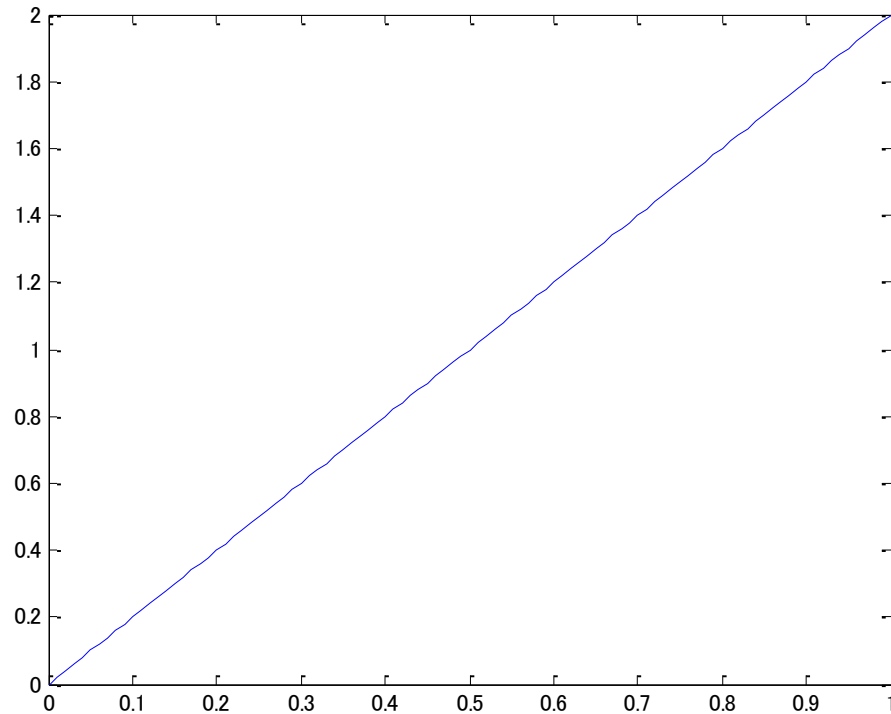
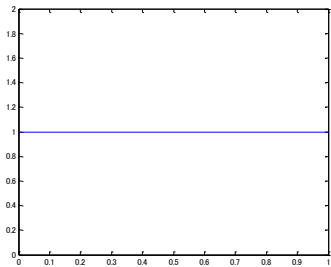
Normalization:

$$\int_0^1 p(H|\theta)p(\theta) d\theta = \int_0^1 \theta d\theta = \left[\frac{1}{2} \theta^2 \right]_0^1 = \frac{1}{2}$$

We have

$$p(\theta|D_1) = 2\theta.$$

Posterior distribution $p(\theta|D_1) = 2\theta$



Reflecting D_1 : the head appeared,
the higher θ is, the higher the probability is

Suppose we have event
 D_2 : the head appeared

$$\begin{aligned} & p(\theta|D_2, D_1) \\ & \propto p(D_2|\theta)p(\theta|D_1) \\ & = p(H|\theta)p(\theta|D_1) \\ & = \theta \times 2\theta = 2\theta^2 \end{aligned}$$

The latest posterior distribution, $p(\theta|D_1)$ is used as the prior distribution in this step because it is the best knowledge of θ .

Normalization:

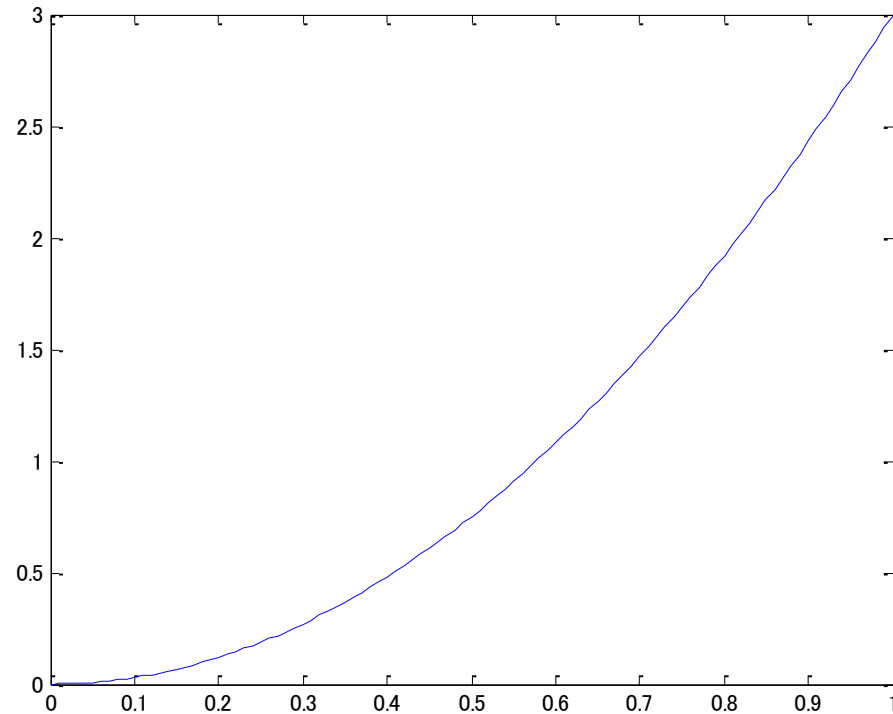
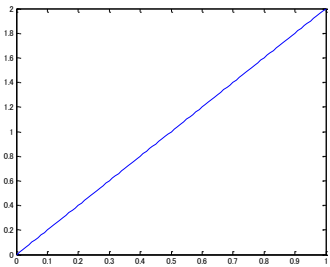
$$\int_0^1 2\theta^2 d\theta = \left[\frac{2}{3} \theta^3 \right]_0^1 = \frac{2}{3}$$

Thus, we have

$$= \frac{p(\theta|D_2, D_1)}{p(D_2|D_1)} = 3\theta^2$$

Posterior distribution

$$p(\theta|D_2, D_1) = 3\theta^2$$



This graph looks reasonable because we had $D_1 = D_2 = \text{head}$.

Suppose we have event
 D_3 : the tail appeared

Find the posterior $p(\theta|D_3, D_2, D_1)$

$$\begin{aligned} & p(\theta|D_3, D_2, D_1) \\ & \propto p(D_3|\theta)p(\theta|D_2, D_1) \\ & = p(T|\theta)p(\theta|D_2, D_1) \\ & = (1 - \theta) \times 3\theta^2 \end{aligned}$$

Normalization:

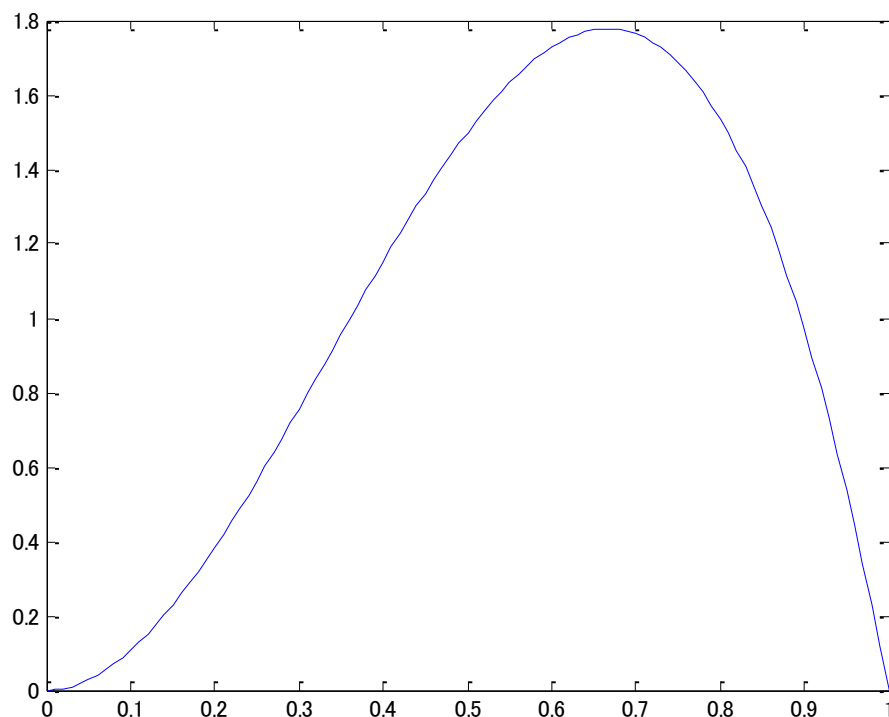
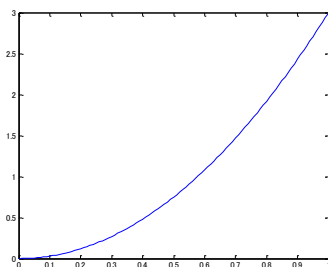
$$\int_0^1 (1 - \theta) \times 3\theta^2 d\theta = \left[\theta^3 - \frac{3}{4}\theta^4 \right]_0^1 = \frac{1}{4}$$

We have

$$p(\theta|D_3, D_2, D_1) = 12(1 - \theta) \times \theta^2.$$

Posterior distribution

$$p(\theta|D_3, D_2, D_1) = 12(1 - \theta) \times \theta^2$$



Let $f(\theta) = 12(1 - \theta)\theta^2 = 12(\theta^2 - \theta^3)$.

$$f'(\theta) = 12(2\theta - 3\theta^2) = 12\theta(2 - 3\theta).$$

$$f'(\theta) = 0 \Leftrightarrow \theta = 0, \frac{2}{3}$$

The fact that the maximum value of the posterior is given at $\theta = \frac{2}{3}$ coincides with the current observations: H, H, T.

Suppose we have event
D4: the tail appeared

Find the posterior $p(\theta|D_4, D_3, D_2, D_1)$

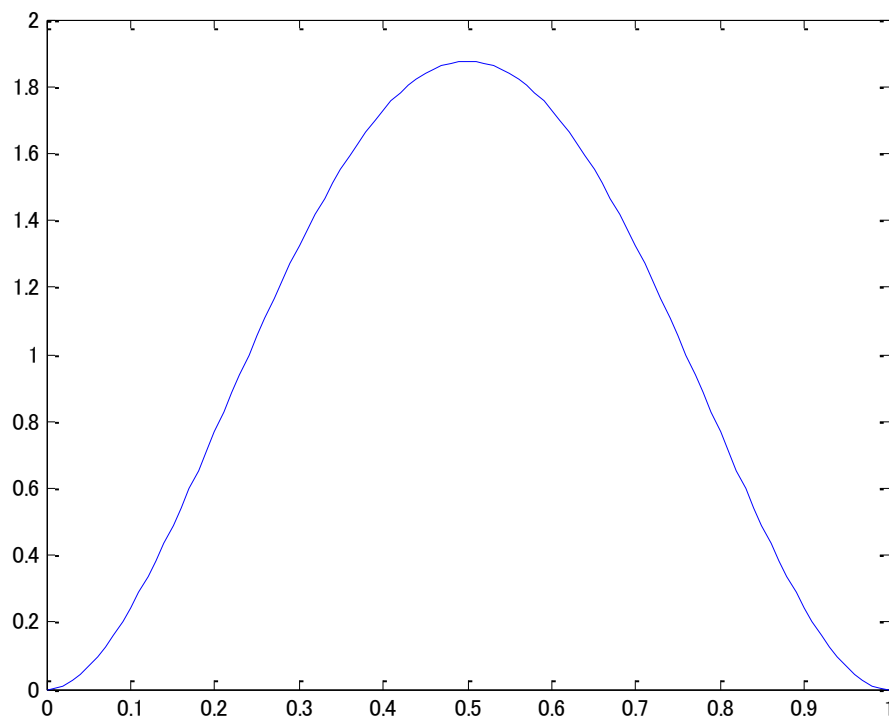
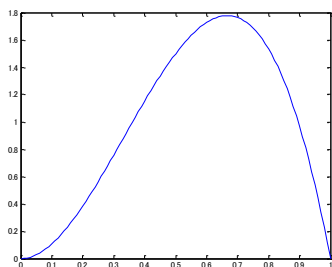
$$\begin{aligned} & p(\theta|D_4, D_3, D_2, D_1) \\ & \propto p(D_4|\theta)p(\theta|D_3, D_2, D_1) \\ & = p(T|\theta)p(\theta|D_3, D_2, D_1) \\ & = (1 - \theta) \times 12(1 - \theta) \times \theta^2 \end{aligned}$$

After normalizing it, we have

$$p(\theta|D_4, D_3, D_2, D_1) = 30(1 - \theta)^2 \times \theta^2.$$

Posterior distribution

$$p(\theta|D_4, D_3, D_2, D_1) = 30(1 - \theta)^2 \times \theta^2$$

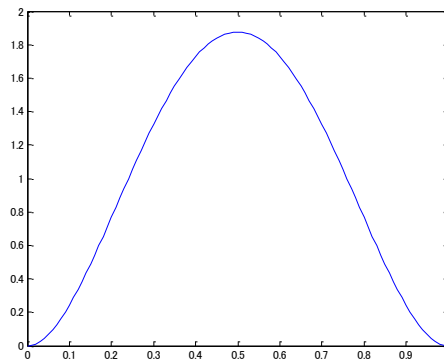
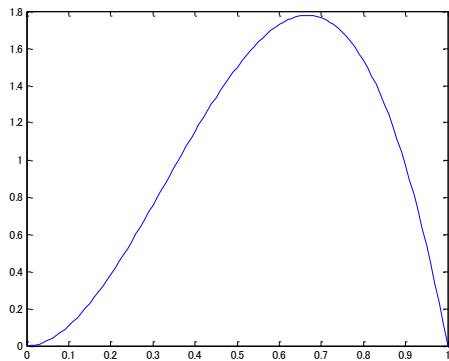
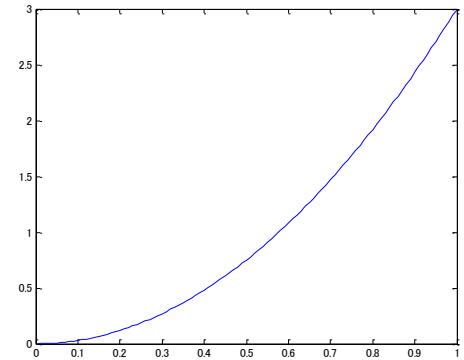
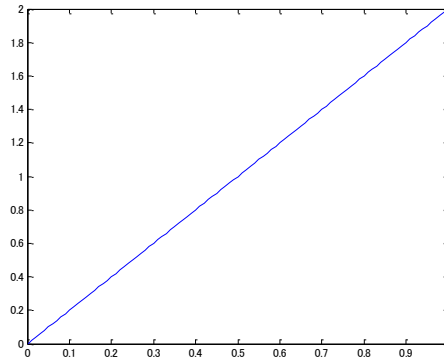
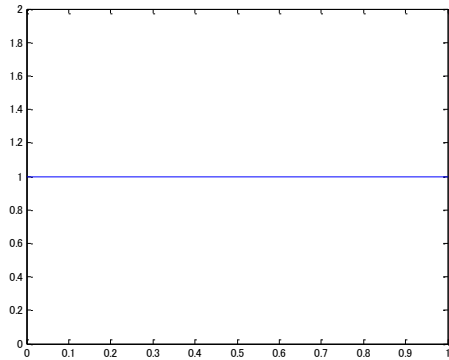


$$\text{Let } f(\theta) = 30(1 - \theta)^2 \times \theta^2 = 30(\theta^2 - 2\theta^3 + \theta^4)$$

$$f'(\theta) = 30(2\theta - 6\theta^2 + 4\theta^3) = 60\theta(1 - 3\theta + 2\theta^2) = 120\theta\left(\theta - \frac{1}{2}\right)(\theta - 1).$$

$$f'(\theta) = 0 \Leftrightarrow \theta = 0, \frac{1}{2}, 1.$$

Data: H H T T



We can integrate old and new data by Bayes' theorem. It is difficult in the traditional statistics.