Statistics and data analysis I

Week 5
"2D data (2)"

Takashi Sano and Hirotada Honda

Lecture plan

Week1: Introduction of the course and some mathematical preliminaries
Week2: Overview of statistics, One dimensional data(1): frequency and histogram
Week3: One dimensional data(2): basic statistical measures
Week4: Two dimensional data(1): scatter plot and contingency table
Week5: Two dimensional data(2): correlation coefficients, simple linear regression and concepts of Probability /
Probability(1):randomness and probability, sample space and probabilistic events
Week6:Probability(2): definition of probability, additive theorem, conditional probability and independency
Week7:Review and exam(i)
Week8: Random variable(1): random variable and expectation
Week9: Random variable(2): Chebyshev's inequality, Probability distribution(1):binomial and Poisson distributions
Week10: Probability distribution(2): normal and exponential distributions
Week11: From descriptive statistics to inferential statistics -z-table and confidwncw interval-
Week12: Hypothesis test(1) -Introduction, and distributions of test statistic (t-distribution)-
Week13: Hypothesis test(2) -Test for mean-
Week14: Hypothesis test(3) -Test for difference of mean-
Week15: Review and exam(2)

※ Might be changed!

Agenda

1. Simple regression
2. Probability(1):Randomness and probability, sample space and event
3. Bayes' theorem

Simple regression

Simple regression

- Regression line : an approximate line that passes nearby the observed 2-d data on the scatter plot.
- Derived by the least square method.

  - Example : Relationship between age and blood pressure.

x: age	35	45	55	65	75
y: average blood pressure [mmHg]	114	124	143	158	166

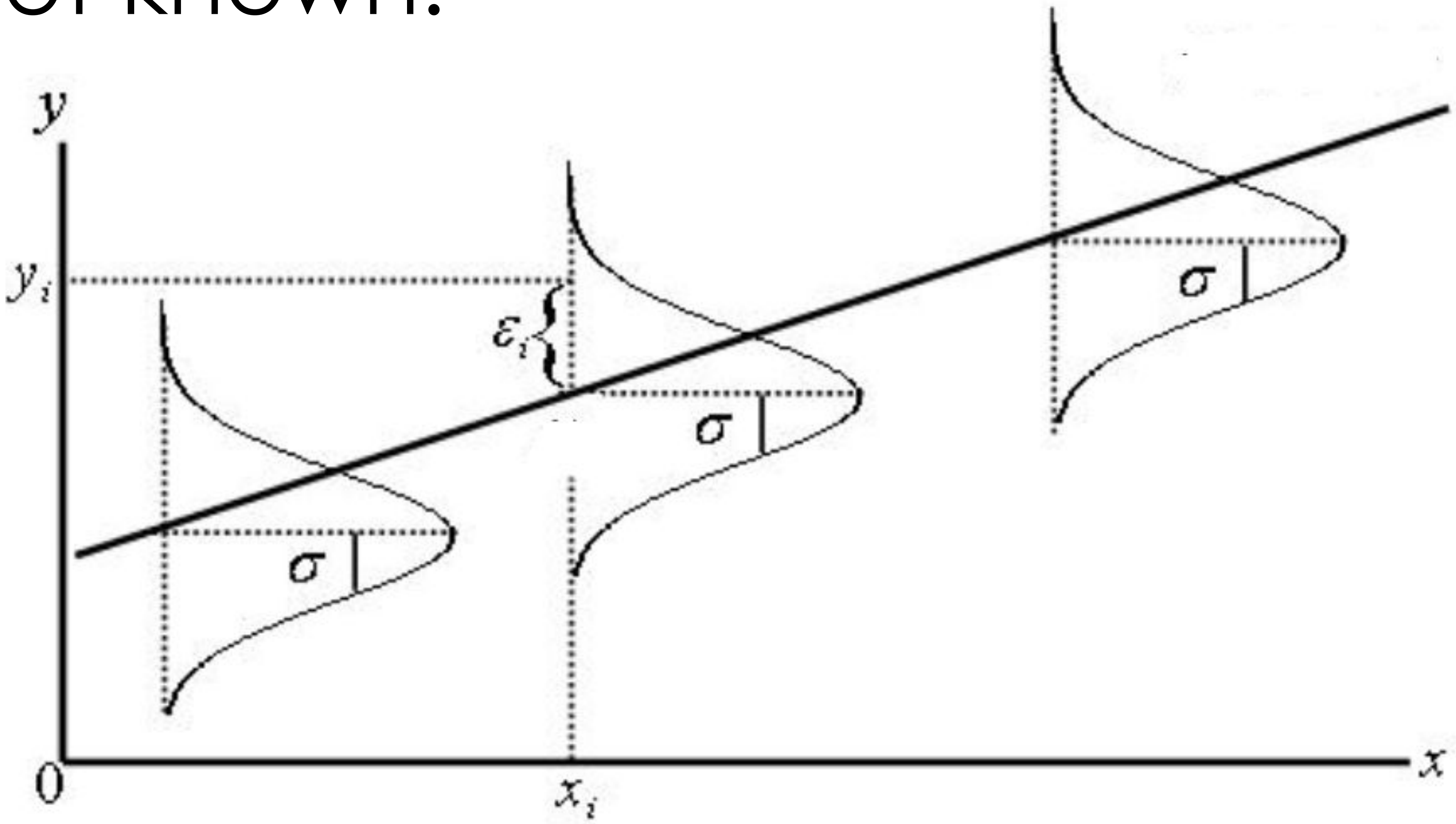
x: independent variable, y: dependent variable

$y = ax + b$ The slope a is called as "regression coefficient"

Model of SR

$$y_i = ax_i + b + \varepsilon_i \ (i = 1, 2, \ldots, n)$$

- σ is not known.



Regression line obtained by least square

- Derive a line that has a minimum distance from observed data.
- If it is denoted as y=ax+b, and each plot as (x_i,y_i), the sum of the distance is represented as

$$L = \sum_{i=1}^{N} \{y_i - (ax_i + b)\}^2$$

  - Then, determine *a* and *b* so that *L* defined above attains its minimum.

Estimation of regression coefficient

- In order to find the regression coefficients, you have to partially differentiate L with respect to a and b.
- The points at which the partial derivatives vanish are the estimated values of regression coefficients.

$$\frac{\partial L}{\partial a} = \sum_i \frac{\partial}{\partial a}(y_i - a - bx_i)^2 = -2\sum_i (y_i - a - bx_i) = 0$$

$$\frac{\partial L}{\partial b} = \sum_i \frac{\partial}{\partial b}(y_i - a - bx_i)^2 = -2\sum_i (y_i - a - bx_i)x_i = 0$$

$$\begin{cases} \sum_i y_i - na - b\sum_i x_i = 0 \\ \sum_i x_i y_i - a\sum_i x_i - b\sum_i x_i^2 = 0 \end{cases}$$

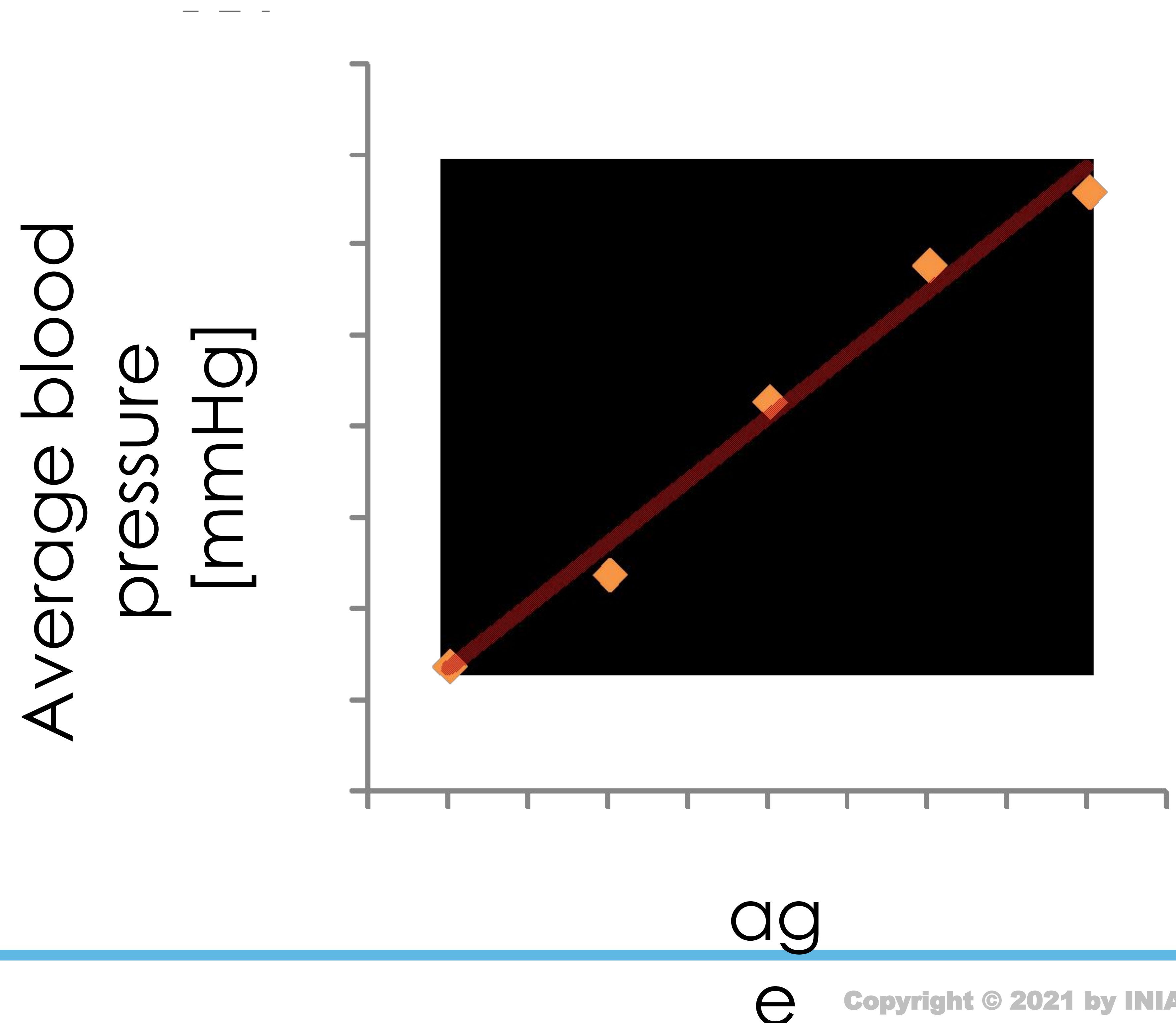
Estimation of regression coefficient

$$a = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \frac{s_{xy}}{s_x^2},$$

- By solving these, you obtain:

$$b = \bar{y} - a\bar{x}$$

Example of regression line



Regression line

$y = 1.38x + 65.1$

Correlation coefficient

$r_{xy} = 0.9918$

Coefficient of determination

$r^2 = 0.9837$

A measure of the goodness of the line ranges in [0,1] (closer to 1, better it is.)

Review so far

- Least square
- The regression line always passes through the mean of x and y.

SR (=Simple regression) by python

Ex.

”BloodPressure.csv”
“BloodPressure.ipyn

Find the regression eq.

x : Ages	35	45	55	65	75
y : Average blood pressure[mmHg]	114	124	143	158	166

Step①: Loading data and making scatter plot

Cell1

```
import pandas as pd
import statsmodels.api as sm

df=pd.read_csv('BloodPressure.csv',sep=',')
df.head()
```

Load the data.

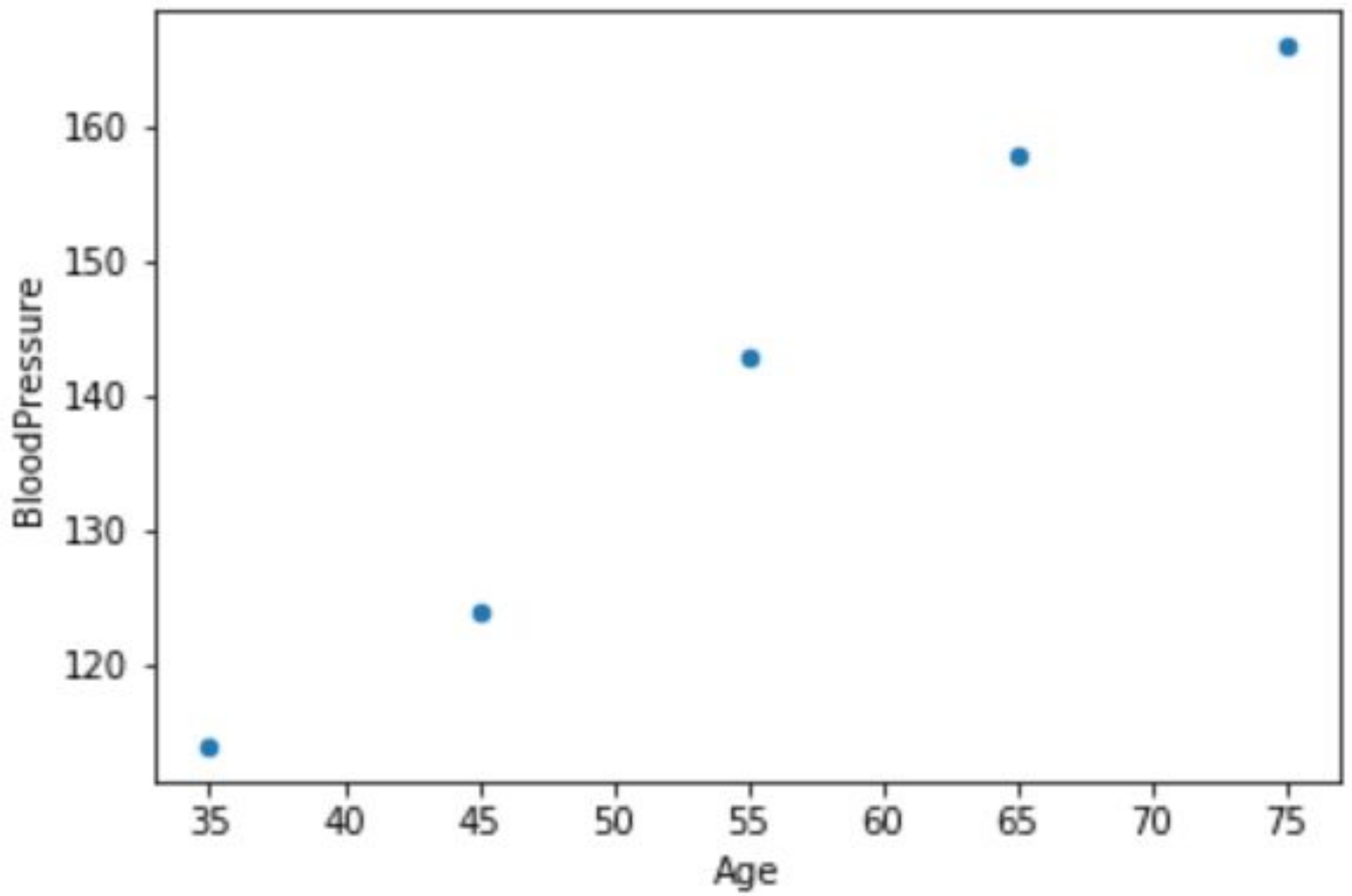
	Age	BloodPressure
0	35	114
1	45	124
2	55	143
3	65	158
4	75	166

Step②: Variables and scatter plot

Cell2

```
x=df[['Age']]
y=df[['BloodPressure']]
df.plot(kind='scatter',x='Age',y='BloodPressure')
```

<matplotlib.axes._subplots.AxesSubplot at 0x19cab099160>



Specify the explanatory Variable (or independent variable) [x] and the dependent variable [y]

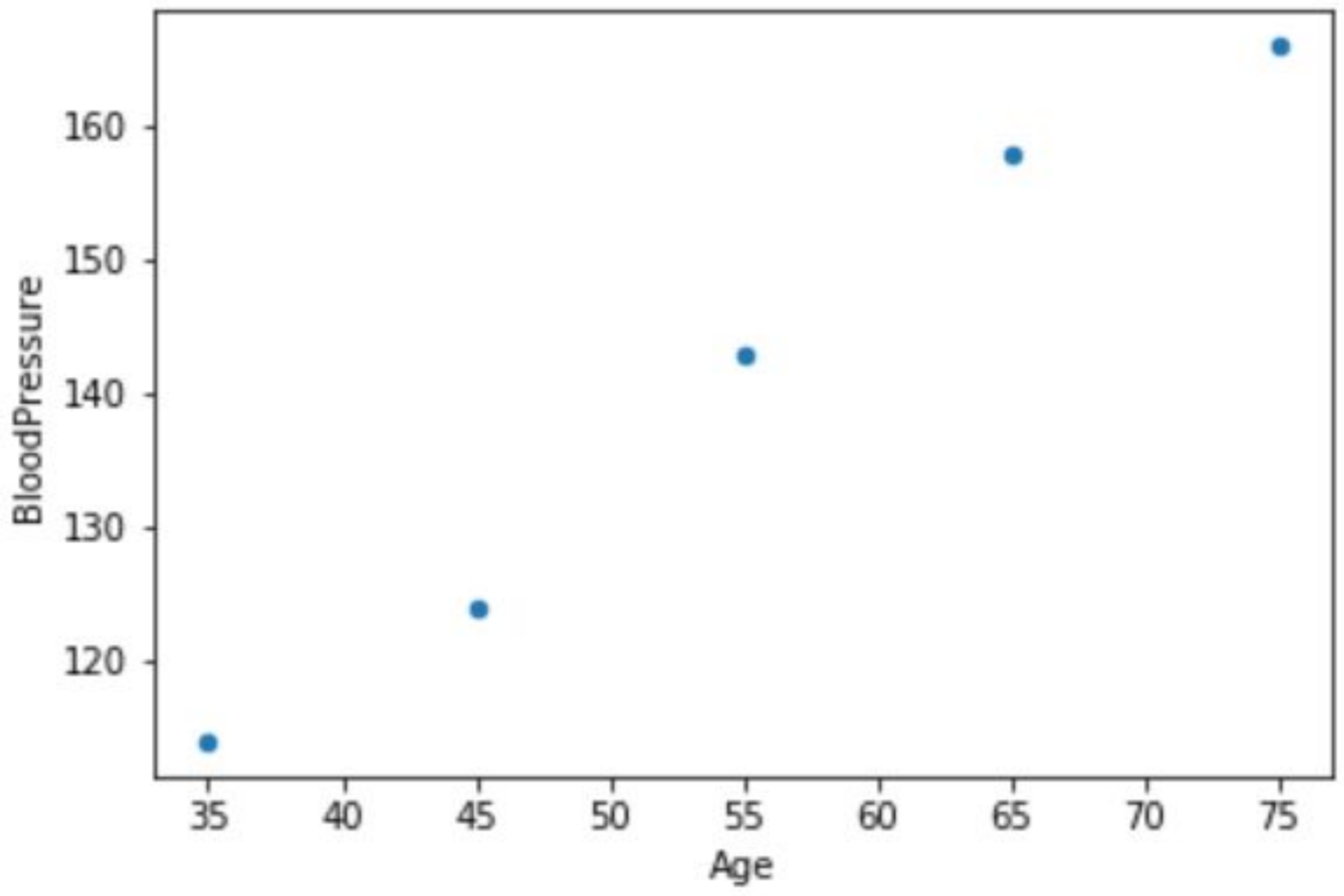
Step②: Variables and scatter plot

Cell2

```
x=df[['Age']]
y=df[['BloodPressure']]
df.plot(kind='scatter',x='Age',y='BloodPressure')
```

Scatter plot

```
&lt;matplotlib.axes._subplots.AxesSubplot at 0x19cab099160&gt;
```



Step③: Simple regression (SR)

Cell3

```
mod=sm.OLS(y,sm.add_constant(x))
res=mod.fit()
print(res.summary())
```

SR and present the result.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          BloodPressure   R-squared:                       0.984
Model:                            OLS   Adj. R-squared:                  0.978
Method:                 Least Squares   F-statistic:                     180.8
Date:                Sun, 12 May 2019   Prob (F-statistic):           0.000889
Time:                        17:33:39   Log-Likelihood:                -11.704
No. Observations:                   5   AIC:                             27.41
Df Residuals:                       3   BIC:                             26.63
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         65.1000      5.828     11.169      0.002      46.551      83.649
Age            1.3800      0.103     13.446      0.001       1.053       1.707
==============================================================================
Omnibus:                          nan   Durbin-Watson:                   2.423
Prob(Omnibus):                    nan   Jarque-Bera (JB):                0.572
Skew:                          -0.118   Prob(JB):                        0.751
Kurtosis:                       1.360   Cond. No.                         228.
==============================================================================
```

Step③: Simple regression (SR)

Cell3

```
mod=sm.OLS(y,sm.add_constant(x))
res=mod.fit()
print(res.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          BloodPressure   R-squared:                       0.984
Model:                            OLS   Adj. R-squared:                  0.978
Method:                 Least Squares   F-statistic:                     180.8
Date:                Sun, 12 May 2019   Prob (F-statistic):           0.000889
Time:                        17:33:39   Log-Likelihood:                -11.704
No. Observations:                   5   AIC:                             27.41
Df Residuals:                       3   BIC:                             26.63
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         65.1000      5.828     11.169      0.002      46.551      83.649
Age            1.3800      0.103     13.446      0.001       1.053       1.707
==============================================================================
Omnibus:                          nan   Durbin-Watson:                   2.423
Prob(Omnibus):                    nan   Jarque-Bera (JB):                0.572
Skew:                          -0.118   Prob(JB):                        0.751
Kurtosis:                       1.360   Cond. No.                         228.
==============================================================================
```

R-squared
(=determination coef.)
0.98

Step③:Simple regression (SR)

Cell3

```
mod=sm.OLS(y,sm.add_constant(x))
res=mod.fit()
print(res.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          BloodPressure   R-squared:                       0.984
Model:                            OLS   Adj. R-squared:                  0.978
Method:                 Least Squares   F-statistic:                     180.8
Date:                Sun, 12 May 2019   Prob (F-statistic):           0.000889
Time:                        17:33:39   Log-Likelihood:                -11.704
No. Observations:                   5   AIC:                             27.41
Df Residuals:                       3   BIC:                             26.63
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         65.1000      5.828     11.169      0.002      46.551      83.649
Age            1.3800      0.103     13.446      0.001       1.053       1.707
==============================================================================
Omnibus:                          nan   Durbin-Watson:                   2.423
Prob(Omnibus):                    nan   Jarque-Bera (JB):                0.572
Skew:                          -0.118   Prob(JB):                        0.751
Kurtosis:                       1.360   Cond. No.                         228.
==============================================================================
```

The regression eq. is
Y=1.38x + 65.10

Multiple regression

- In case the dependent variable y is explained by multiple independent variables (x1,x2,…), it is called as the multiple regression (MR, hereafter).

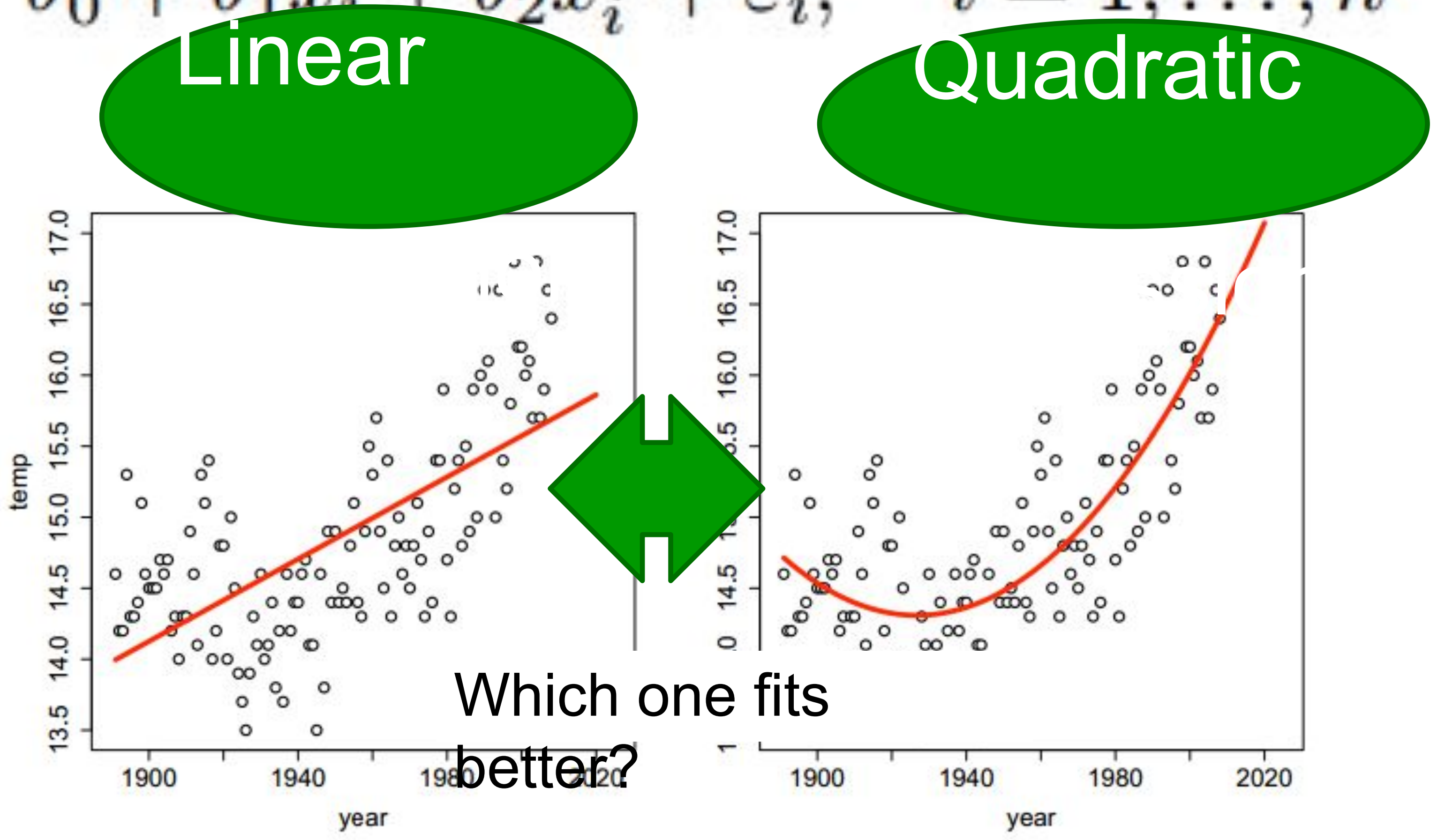
- For instance, if the independent variables are two, the observed data are represented by an approximated plane in the 3-d region, $y = a_1x_1 + a_2x_2 + b$.
  - Let us explain the physical age as a function of the blood pressure and lung capacity.
  - Physical age = 0.428×maximum blood pressure [mmHg] -0.0077×lung capacity[ml] + 11.8

- In MR, correlation coefficient is replaced by the multiple correlation coefficient.

Other regressions

- In case the data is not clearly subjected to linear function, we use other regressions. For instance, we sometimes use the polynomial regression.
  - Ex) average temperature in Nagoya.

$$Y_i = \theta_0 + \theta_1 x_i + \theta_2 x_i^2 + \varepsilon_i, \quad i = 1, \ldots, n$$



2. Probability(1) : Randomness, probability, sample space and event

Randomness

- Randomness : We can’t predict what to happen next in the deterministic manner.
  - If you are tossing a coin, you can’t predict head or tail.
  - But totally, there is a certain rule of the randomness.
- Probability theory deals, not the randomness itself, but the *rule of the randomness*.
  - If you become familiar with the probability theory, you can’t predict “head or tail” in the coin toss.
  - But you will be able to state that head or tail appear with the same probability.

Event, sample point, and sample space

- In statistics, we say what happens as "*event*."
- The result of each event is called as "*sample point*", and is denoted as ω.
- The universal set of the set of the sample points is called as "*sample space*", and is denoted as Ω.
- Each event is defined as a subset of the sample space.
  - The sample space itself is an event.
  - A situation that includes no sample point is also regarded as an event. It is called as a *null event*, and denoted as ϕ.
  - An event A, that consists of sample points $\omega_1, \omega_2, \ldots, \omega_n$, is denoted as $A = \{\omega_1, \omega_2, \ldots, \omega_n\}$.

Examples

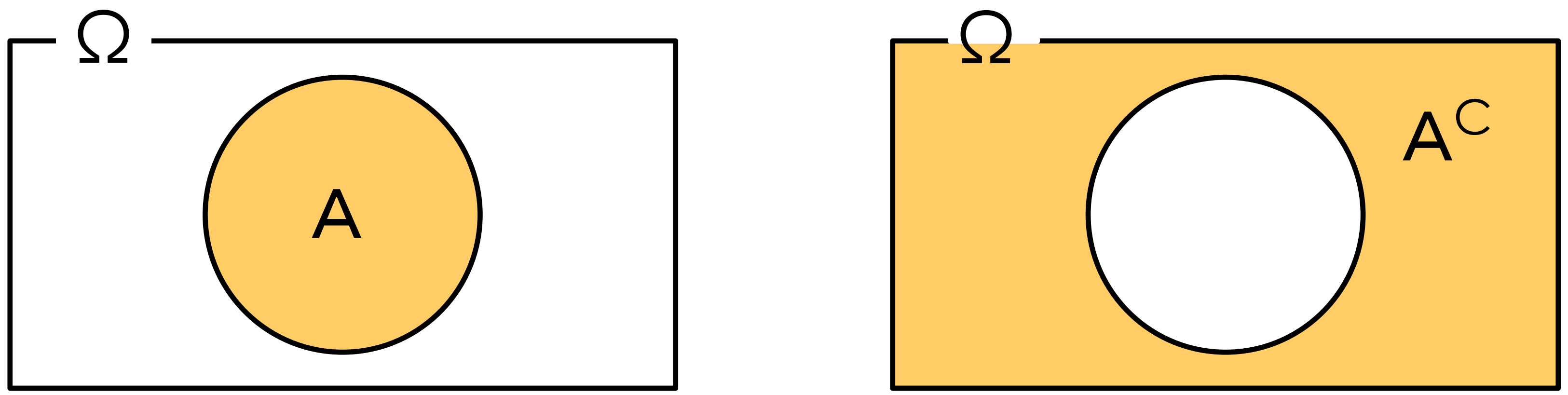
- Let us denote one side of a coin as “1” and another by “0” . After the tossing it twice, the sample space of the result is

$$\Omega = \{(0,0),\ (0,1),\ (1,0),\ (1,1)\}$$

- When we roll a dice once,
  - The sample space Ω is $\Omega = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$.
  - An event A that the number was odd, is $A = \{1,\ 3,\ 5\}$.
  - An event B, that the number was not less than 4, is $B = \{4,\ 5,\ 6\}$.

Venn diagram

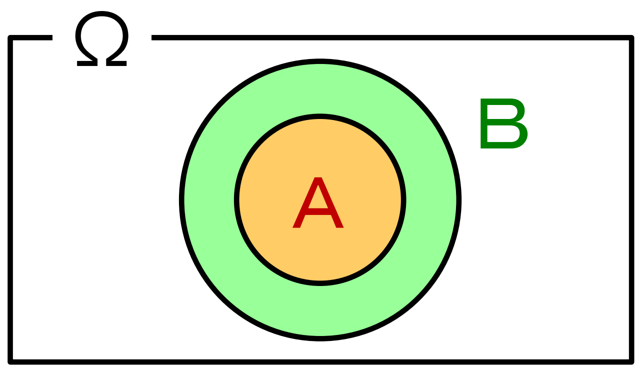
- Venn diagram: useful for understanding the relationship between events.
  - The sample space Ω is denoted as a rectangle.
  - Other events are denoted as circles in Ω.



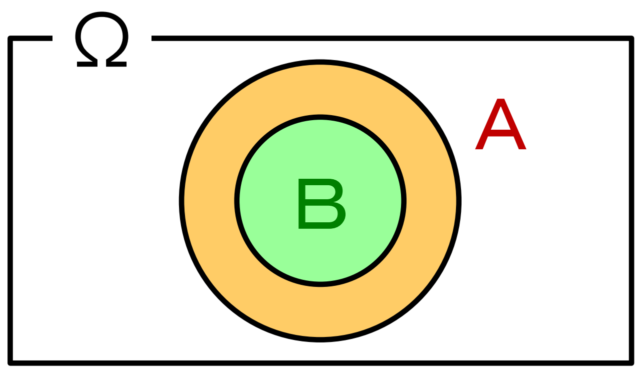
- An event, that an event *A* does not occur, is also an event. It is called as a *complement event*, and is denoted as A^C.
- The complement of Ω is ϕ and vice versa.

Venn diagram

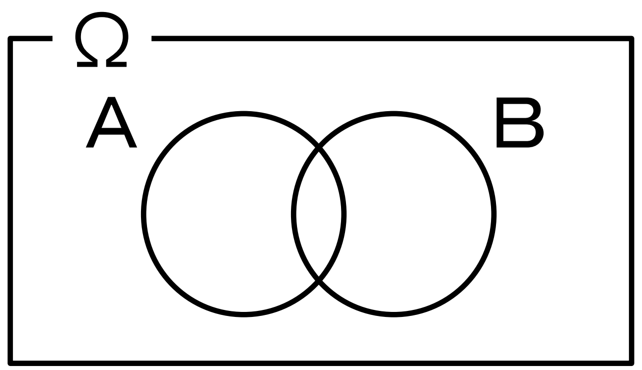
- The relationship between two events *A* and *B*.

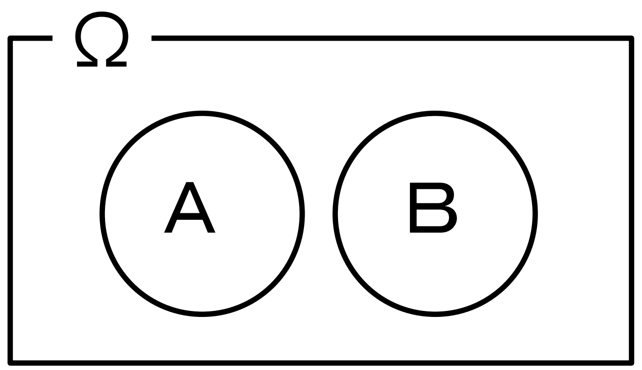


$A \subset B$ (Ais a subset of B)



$A \supset B$ (B is a subset of A)

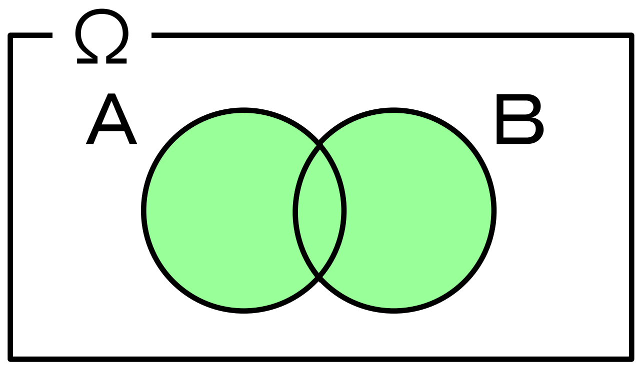




Exclusive events

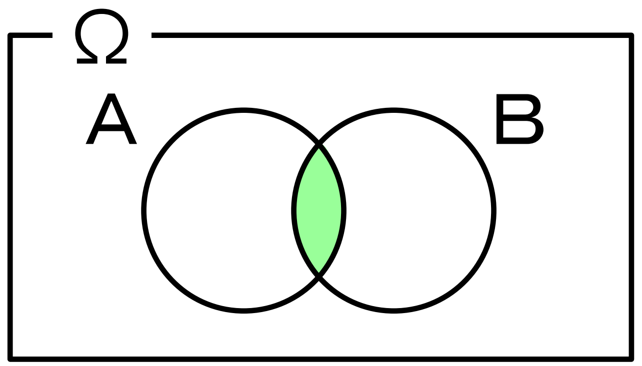
Venn diagram

- The sum and product events of two events *A* and *B*.



$A \cup B$

A or B



$A \cap B$

A and B

- De Morgan's laws

$(A \cup B)^C = A$

$(A \cap B)^C = A^C$

Numbers of permutation and combination

- n！: Factorial of n: $n\times(n-1)\times\ldots\times 2\times 1$
  - We define $0!=1$.
- The number of permutation to take *r* samples from the population of *n* elements is:
  - $_nP_r = n!/(n-r)!$
- The number of combination to make a sample of *r* elements from the population of *n* elements is
  - $_nC_r = {_nP_r}/r! = n!/\{r!(n-r)!\}$
- The number of making *k* groups each of which consists of *r* elements from the population of n(=*rk*) elements is:
  - $\underbrace{{_nC_r} * {_{n-r}C_r} * {_{n-2r}C_r} * \ldots * {_{n-(k-1)r}C_r} / {_kP_k}}_{k\text{-group}}$

Take a notice！

Python codes

```
import math
math.factorial(5)
```

5
!

```
120
```

```
import math

def permutations_count(n, r):
    return math.factorial(n) // math.factorial(n - r)

print(permutations_count(5, 2))
```

${}_5P_2$

```
20
```

```
import math

def combinations_count(n, r):
    return math.factorial(n) // math.factorial(n - r) // math.factorial(r)

print(combinations_count(5, 2))
```

${}_5C_2$

```
10
```

2-1. Probability; its definition

Probability

- Likeliness of occurrence of events
  - The probability of an event A's occurrence is denoted as P(A).

- [Reference] There have been various trials for the definition of *probability*.
- Laplace's definition
- Frequency theory
- Kolmogorov's definition
- …

Definition of probability

- The probability of an event *A* , denoted as P(A), is a function of sets satisfying:
  - (i) $0 \leqq P(A) \leqq 1$ for all events A,
  - (ii) $P(\Omega) = 1$,
  - (iii) and $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$
    for arbitrary exclusive sets A_1, A_2 ....

Supplementary issues

So far, you have learned *probability* as a ratio of the number of ways of the event you are interested in out of those of all possible events.
In general, however, we should think of the probability of infinite events; c.f. infinitely many tossing of a die.

Dealing with "infinitely many trials" or "sample space involving infinite elements" mathematically causes some difficulties.

Supplementary issues

Feature (iii) in the former definition of probability should be defined with a set of infinitely many sets. This is an essential difference from what you have learned until middle school (or high school); Seems abstract, but enables you to consider probability on infinite sample spaces. Actually, it's based on *measure theory* and *integral calculus* (in the sense of Lesbegue). In general, probability cannot be determined on all subsets of the sample space (i.e., ' the set A''s are nor arbitrary); we can define probability on families of sets having some 'good' features. We omit the detail here.

Addition theorem

- When two sets A and B are exclusive, i.e. $A \bigcap B = \phi$

P(A∪B)＝P(A)＋P(B)

- Suppose you are rolling a dice. Let us denote an event that the pip of "1" is shown as *A*, and an event that the the pip of "2" is shown, as *B*. Then,P(A∪B)＝1/6＋1/6＝1/3.
- In case the product of events A and B is not null, i.e., $A \bigcap B \neq \phi$

P(A∪B)＝P(A)＋P(B)ーP(A∩B)

- Let us denote an event that the pip of an odd is shown as *A*, and an event that the the pip not greater than 3 is shown, as *B*. Then, P(A∪B)＝1/2＋1/2ー1/3＝2/3

2-2. Conditional probability and independence

Conditional probability

- Let us assume that an event *B* has occurred. Then, the probability of an event *A* under this assumption, is called as the *conditional probability of A given B.* It is denoted as $P(A \mid B)$.

  - It is defined as:

$$P(A \mid B) = P(A \cap B) / P(B)$$

  - Rewriting above, we have

$$P(A \cap B) = P(A \mid B) \cdot P(B)$$

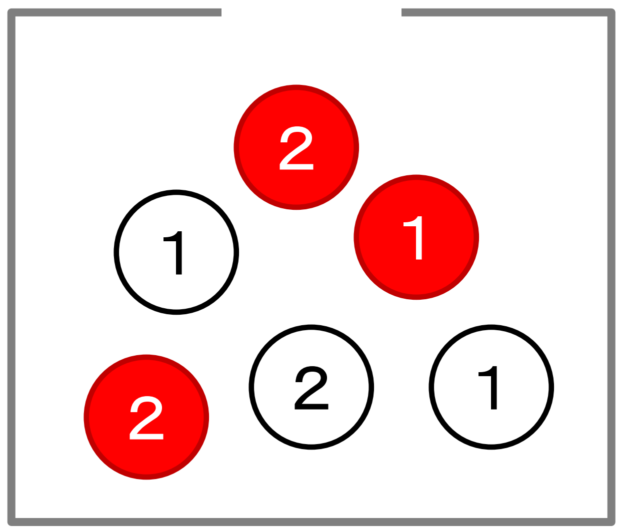
←Multiplication theorem

  - Because we imposed no assumptions on A and B, we also have

$$P(A \cap B) = P(B \mid A) \cdot P(A)$$

Example

- Suppose you are taking a ball out of a box, in which there are three red balls and three white balls.
- As the figure below shows, a number “1” or “2” is marked on each ball.
- Now, find the probability that the number is “1” given that you have taken a white ball out of the box.



Event-A:Pip 1 appears Event-B : White ball

$$P(A \mid B) = P(A \cap B) / P(B)$$

$$= 2/6 \,/\, \tfrac{1}{2} = 2/3$$

Note that the value is higher than the case you don’t know the color of the ball.

Independence

- When the probability of an event A doesn't depend on that of an event B, i.e., when $P(A) = P(A \mid B)$ holds,

we say that *A is independent of B*.

 In this case, we have

$$P(A \cap B) = P(A \mid B) \cdot P(B)$$
$$= P(A) \cdot P(B)$$

Example

- Suppose you throw a dice twice. Then, find the probability that the pip of one are shown twice.
  - Let A be an event that the pip of one is shown in the first throw. Then, $P(A)=1/6$.
  - Let B be an event that the pip of one is shown in the second throw. Then, $P(B)=1/6$.
  - Since these two events are independent, we have

$$P(A \cap B) = P(A) \cdot P(B) = 1/36$$

Example

- In case you roll five dice at once, find the probability that three of them show the pip of one.
  - For each dice, the probability of showing one is 1/6, and the probability of other pips is 5/6.
  - The number of ways that just three dices show the pip of one, and others show pips other than one, is ${}_5C_3$.
  - Thus, the desired probability is

$$, {}_5C_3 \cdot (1/6) \cdot (1/6) \cdot (1/6) \cdot (5/6) \cdot (5/6) \fallingdotseq 0.032$$

3. Bayes’ theorem

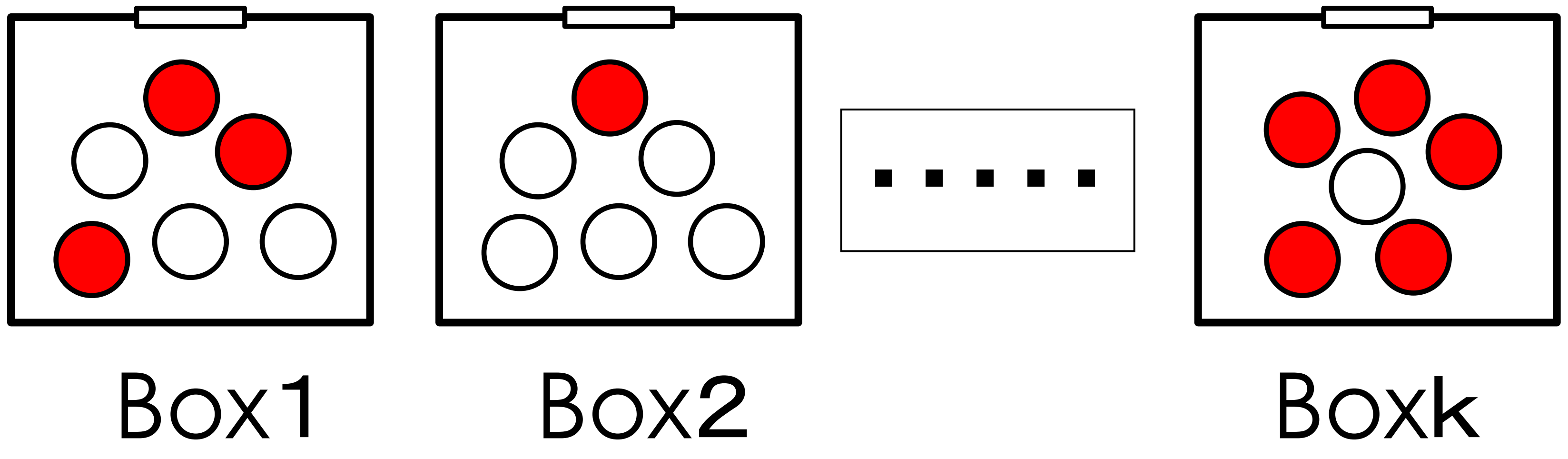
Bayesian approach

We go on to Bayes' theorem here after.
By using the exchangeability A and B in the multiplication theorem, we obtain a formula (Bayes' theorem).

In Bayesian statistics, they regard it as the one to Compute the "likeliness of factors of cause" based on the "probability of results from the cause".

Bayes' theorem [1/2]

- Ex) Suppose that some red and white balls are in some boxes, and you take some balls out of a box randomly chosen.
- Then, estimate the most probable box based on the color of the ball.



Bayes' theorem [2/2]

- Suppose that the events $H_1, H_2, \ldots, H_k$ are exclusive with each other, and satisfy $H_1 \cup H_2 \cup \ldots \cup H_k = \Omega$.
- Then, for a certain event A, the following holds.

$$P(A) = \sum_j P(A \bigcap H_j) = \sum_j P(H_j)P(A|H_j)$$

Summation theorem

Multiplication theorem

- Thus,

$$\boxed{P(H_i|A)} = \frac{P(H_i \bigcap A)}{P(A)}$$

←Multiplication theorem

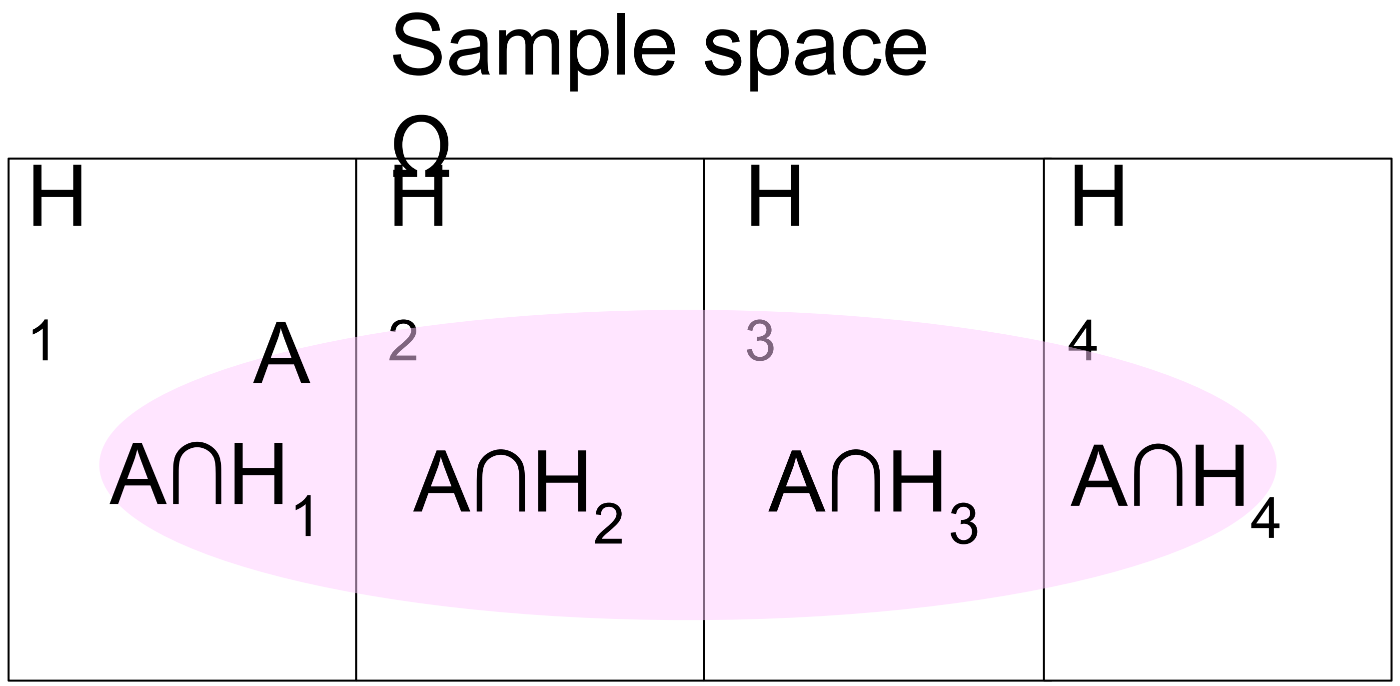
Posterior probability

$$= \frac{P(H_i \bigcap A)}{\sum_j P(H_j)P(A|H_j)}$$

Prior probability

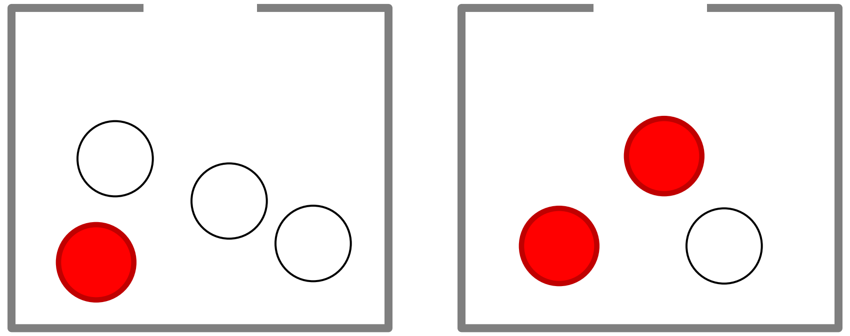
$$= \frac{\boxed{P(H_i)}P(A|H_i)}{\sum_j P(H_j)P(A|H_j)}$$

←Multiplication theorem


Sample space Ω
H1
H2
H3
H4
A
A∩H1
A∩H2
A∩H3
A∩H4

Example of Bayes' theorem

- When you took out a ball from a box chosen from the two boxes depicted below, it was a white ball. Then, estimate from which box you took out the ball.



Box 1 Box2

The event taking out a ball from box 1 : H_1
The event taking out a ball from box 2 : H_2
The event taking out a white ball : A

$$P(H_1)=P(H_2)=1/2$$

$$P(A \mid H_1)=3/4$$

$$P(A \mid H_2)=1/3$$

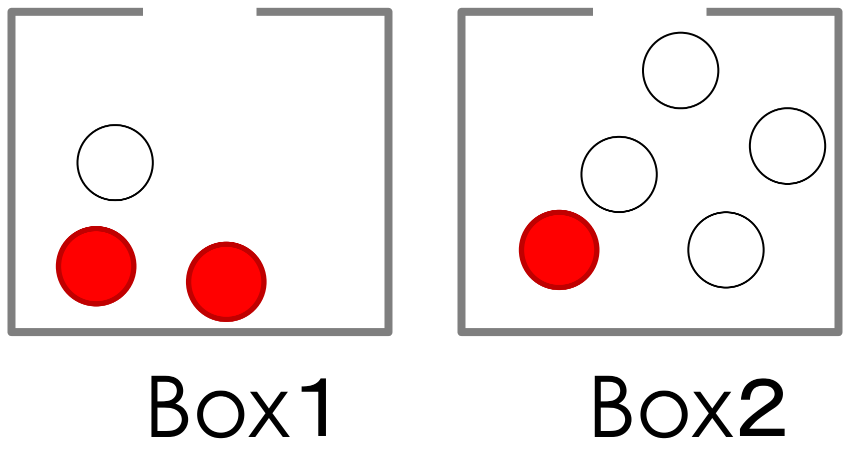
$$P(H_1 \mid A)=\{(1/2)\cdot(3/4)\}/\{(1/2)\cdot(3/4)+(1/2)\cdot(1/3)\}$$
$$= 0.692$$

$$P(H_2 \mid A)=\{(1/2)\cdot(1/3)\}/\{(1/2)\cdot(3/4)+(1/2)\cdot(1/3)\}$$
$$=0.308$$

Example

There are two boxes (say, A and B) filled with red and white balls.
Two red balls and one white ball are in box 1.
One red ball and four white balls are in box 2.
The coordinator randomly choose one box, which you cannot know.

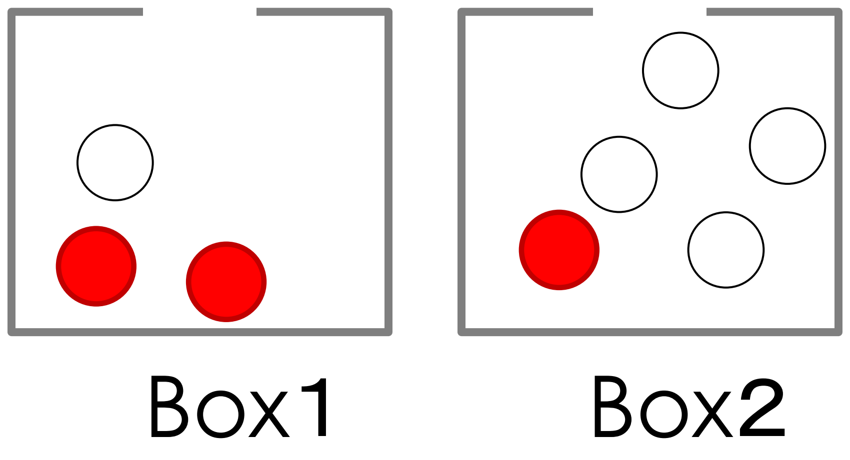
・The participant take out a ball from the chosen box, see the color, and return it into the box.
・Then, estimate the box in case the color of the balls is red.



Example【Hint】

There are two boxes (say, A and B) filled with red and white balls.
Two red balls and one white ball are in box 1.
One red ball and four white balls are in box 2.
The coordinator randomly choose one box, which you cannot know.

・The participant take out a ball from the chosen box, see the color, and return it into the box.
・Then, estimate the box in case the color of the balls is red.
→ You can compare only the numerators!



Example【Ans.】

From box1 : H_1
From box2 : H_2
Red ball : A

$P(H_1) = P(H_2) = 1/2$
$P(A \mid H_1) = 2/3$
$P(A \mid H_2) = 1/5$

$$P(H_1|A) = \frac{P(H_1)P(A|H_1)}{P(H_1)P(A|H_1) + P(H_2)P(A|H_2)} = \frac{\frac{1}{2} \times \frac{2}{3}}{\frac{1}{2} \times \frac{2}{3} + \frac{1}{2} \times \frac{1}{5}}$$

$$P(H_2|A) = \frac{P(H_2)P(A|H_2)}{P(H_1)P(A|H_1) + P(H_2)P(A|H_2)} = \frac{\frac{1}{2} \times \frac{1}{5}}{\frac{1}{2} \times \frac{2}{3} + \frac{1}{2} \times \frac{1}{5}}$$

$\frac{2}{3} > \frac{1}{5}$ Thus, H1 is more likely, i.e., the ball may be taken from box1.

Application of Bayes' theorem

- A certain e-mail includes the word "free". Then, decide whether or not this e-mail is a spam.
  - Consider the items corresponding to "boxes", "red balls" and "white balls" in the previous example.
  - 「red ball」= an e-mail including the word "free"
  - 「white ball」= an e-mail not containing the word "free"
  - 「Box 1」=A set of spams
  - 「Box 2」=A set of e-mails other than spams

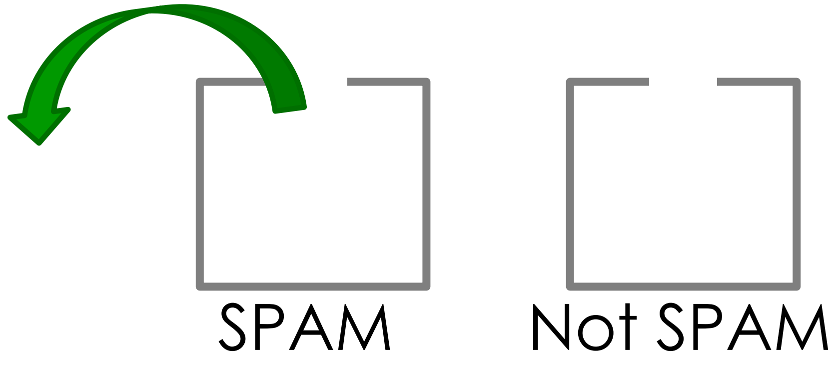
Application of Bayes' theorem

- Suppose that a certain mail to you includes the word "free." Then, determine whether this mail is a spam or not.
  - You should find

P(that mail is a spam | it contains the word "free")

> In the previous example,
> P(Taking a ball from box1 | taking out a red ball)

The event taking out a ball from box 1(=spam) : H_1
The event taking out a ball from box 2(=not spam) : H_2
The event taking out a red ball(containing "free") : A



Application of Bayes' theorem

- We again list the events.
  - The mail is a spam : H_1
  - The mail is not a spam : H_2
  - The received e-mail contains the word "free" : A

- $P(H_1 \mid A)$: the probability that the received mail is a spam given that it contains the word "free".
- $P(H_2 \mid A)$: the probability that the received mail is not a spam given that it contains the word "free".

Application of Bayes' theorem

- By virtue of the Bayes' theorem,

$$P(H_1|A) = \frac{P(H_1)P(A|H_1)}{\sum_j P(H_j)P(A|H_j)}$$

$$P(H_2|A) = \frac{P(H_2)P(A|H_2)}{\sum_j P(H_j)P(A|H_j)}$$

Which is larger?

$P(H_1)$ = "The ratio of spams in the overall e-mails."
$P(H_2)$ = "The ratio of e-mails other than spams in the overall e-mails."

$P(A \mid H_1)$ = The probability that the word "free" is contained in a spam.
$P(A \mid H_2)$ = The probability that the word "free" is contained in a non-spam mail.

Application of Bayes' theorem

$$P(H_1|A) = \frac{P(H_1)P(A|H_1)}{\sum_j P(H_j)P(A|H_j)}$$

$$P(H_2|A) = \frac{P(H_2)P(A|H_2)}{\sum_j P(H_j)P(A|H_j)}$$

Since the denominators are same, we just compare the numerators!

Application of Bayes' theorem

- By virtue of the Bayes' theorem,

$$P(H_1|A) = \frac{P(H_1)P(A|H_1)}{\sum_j P(H_j)P(A|H_j)} \text{ ------ ①}$$

$$P(H_2|A) = \frac{P(H_2)P(A|H_2)}{\sum_j P(H_j)P(A|H_j)} \text{ ------ ②}$$

We assume

$P(H_1)$ = "The ratio of spams in the overall e-mails." = 0.6
$P(H_2)$ = "The ratio of e-mails other than spams in the overall e-mails."=0.4
$P(A | H_1)$ = The probability that the word "free" is contained in a spam.=0.7
$P(A | H_2)$ = The probability that the word "free" is contained in a non-spam mail.=0.1

Application of Bayes' theorem

$$P(H_1|A) = \frac{P(H_1)P(A|H_1)}{\sum_j P(H_j)P(A|H_j)} \quad \text{------①}$$

$$P(H_2|A) = \frac{P(H_2)P(A|H_2)}{\sum_j P(H_j)P(A|H_j)} \quad \text{------②}$$

$P(H_1)$ = "The ratio of spams in the overall e-mails." = 0.6
$P(H_2)$ = "The ratio of e-mails other than spams in the overall e-mails."=0.4

$P(A \mid H_1)$ = The probability that the word "free" is contained in a spam.=0.7
$P(A \mid H_2)$ = The probability that the word "free" is contained in a non-spam mail.=0.1

① = 0.6×0.7=0.42
② = 0.7×0.1=0.07

➡ スパムに分類！

Summary

- Addition theorem

  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

- Conditional probability

  $P(A \mid B) = P(A \cap B) / P(B)$

  $P(A \cap B) = P(A \mid B) \cdot P(B)$ ： Multiplication theorem

- Independence

  $P(A \cap B) = P(A \mid B) \cdot P(B) = P(A) \cdot P(B)$

- Bayes' theorem

$$\frac{P(H_i)P(A|H_i)}{\sum_j P(H_j)P(A|H_j)}$$

Summary (checklist)

- You can state the definition of probability?
- You can state the addition theorem of probability?
- You can state the definition of the conditional probability?
- You can state the Bayes' theorem? Also, can you answer the exercises of “red and white balls in the boxes”?