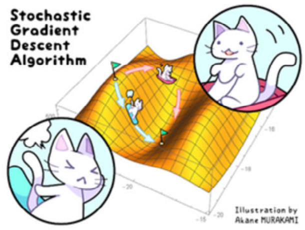
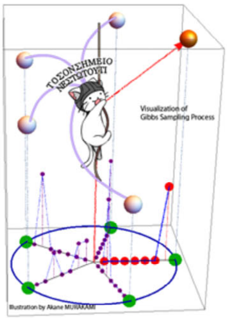
Public Lecture in
Department of Economics
Faculty of Economics and Business Universitas Indonesia



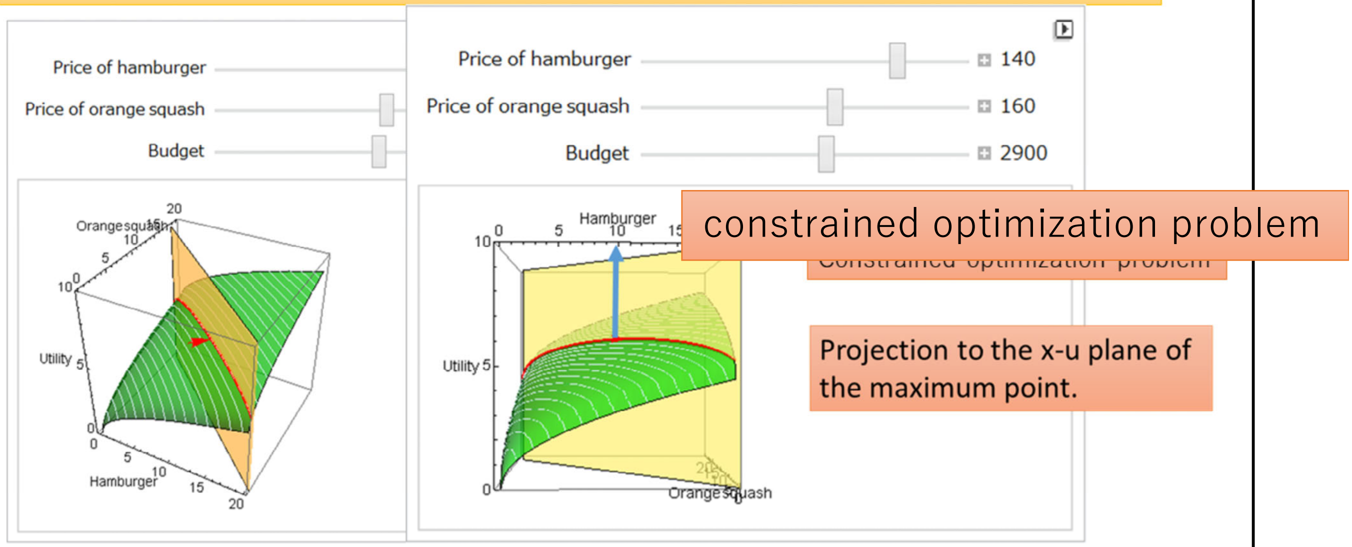
Visually Do Economics Mathematics !



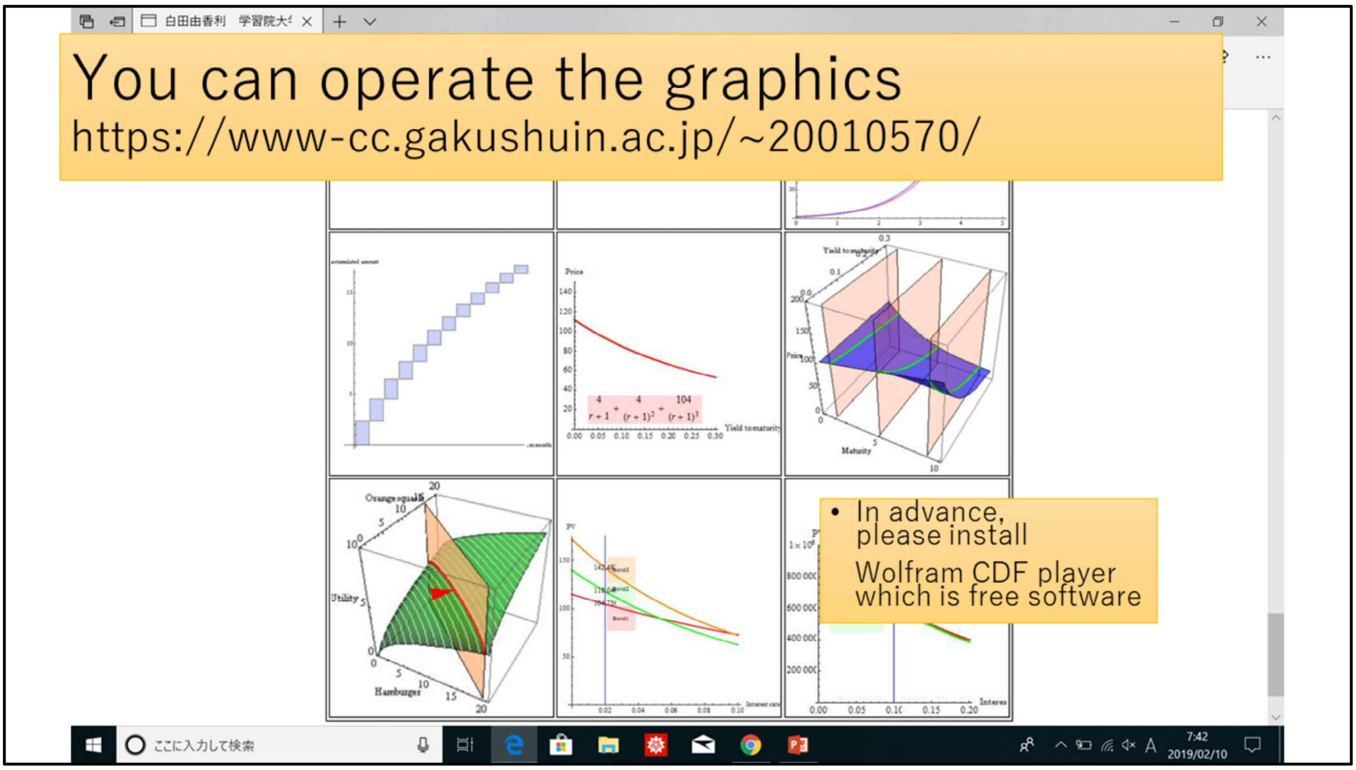
27th February 2019
Gakushuin University
Faculty of Economics
Prof. Yukari Shirota

Contents

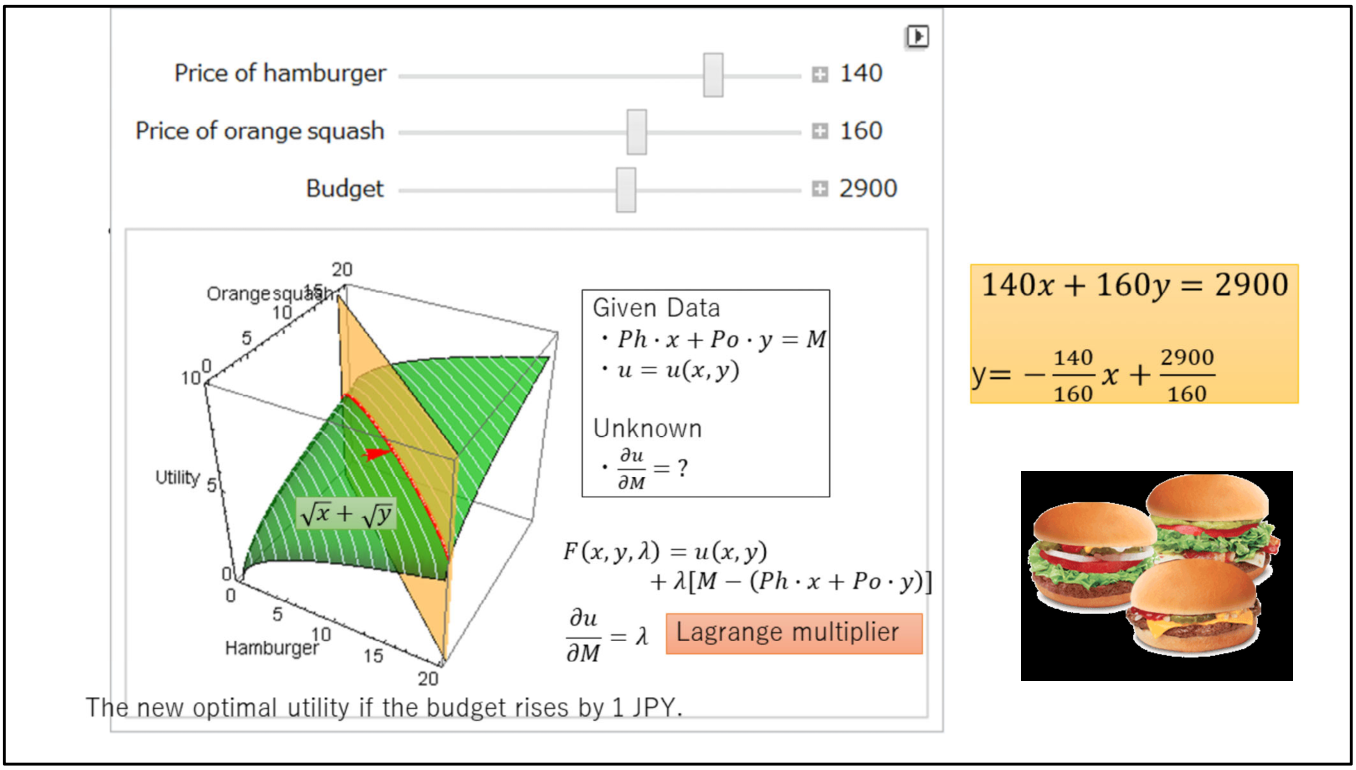
1. Visualization of Lagrange multiplier methods.



First let's start from the Lagrange multiplier problem. That is a constrained optimization problem.
I suppose that you have studied the things already. Today's theme is visualization of the economics mathematics. I would like to show you many visual materials. So let me skip explanations in detail.
I hope that by the visual teaching materials you can understand more deeply the mathematics.

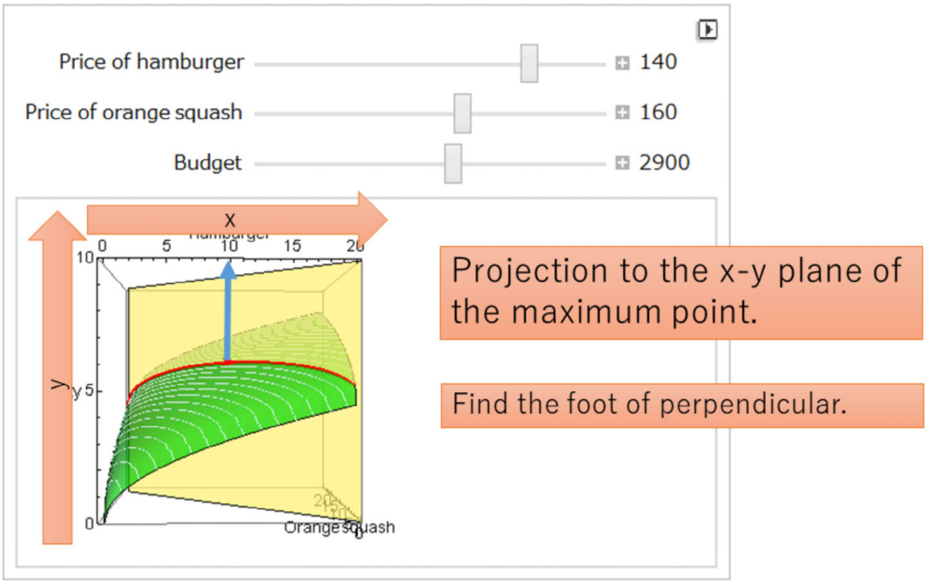


The today’s teaching materials gave been published already so you can access via my web sites.
In advance, please install the Wolfram CDDF player.



Let me explain the word problem.
Suppose that you have the budget 2900 JPY. You will buy the hamburgers and orange squashes. The prices are 140 JPY and 160 JPY.
A consumer's utility function is given by square root of x plus square root of y where x is the number of ***hamburgers*** and y is the number of orange squashes. Find the maximum point of utility under the budget limitation.
Estimate the partial differentiation of curl u by curl M, the budget money. In other words, the new optimal utility if the budget rises by 1 JPY.
Let's see the graphics. The green surface shows the utility function. And the yellow plane shows the budget condition.

Visual Solution of Lagrange Multiplier Problem



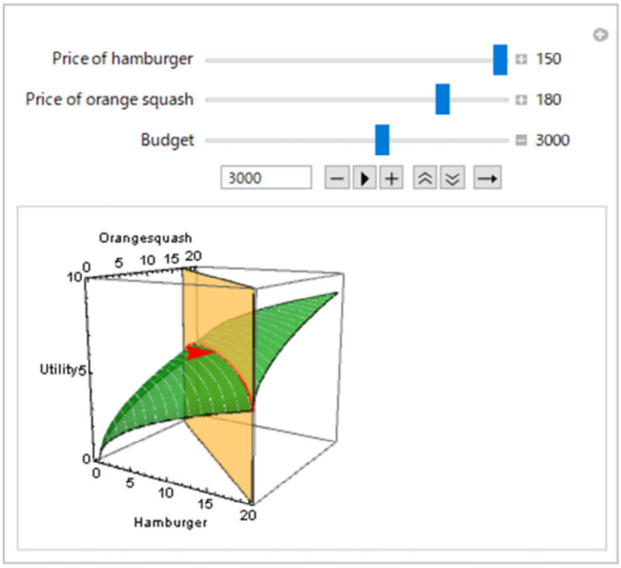
The red curve is the intersection part between the green surface and the yellow plane. That is the available area. Find the maximum point of utility on the curve. This is the maximum point. Then project the graphics to the x-u plane. By the foot of perpendicular of the point, you can see the answer x and y visually. Of course you can calculate the exact figures using the Lagrange method.

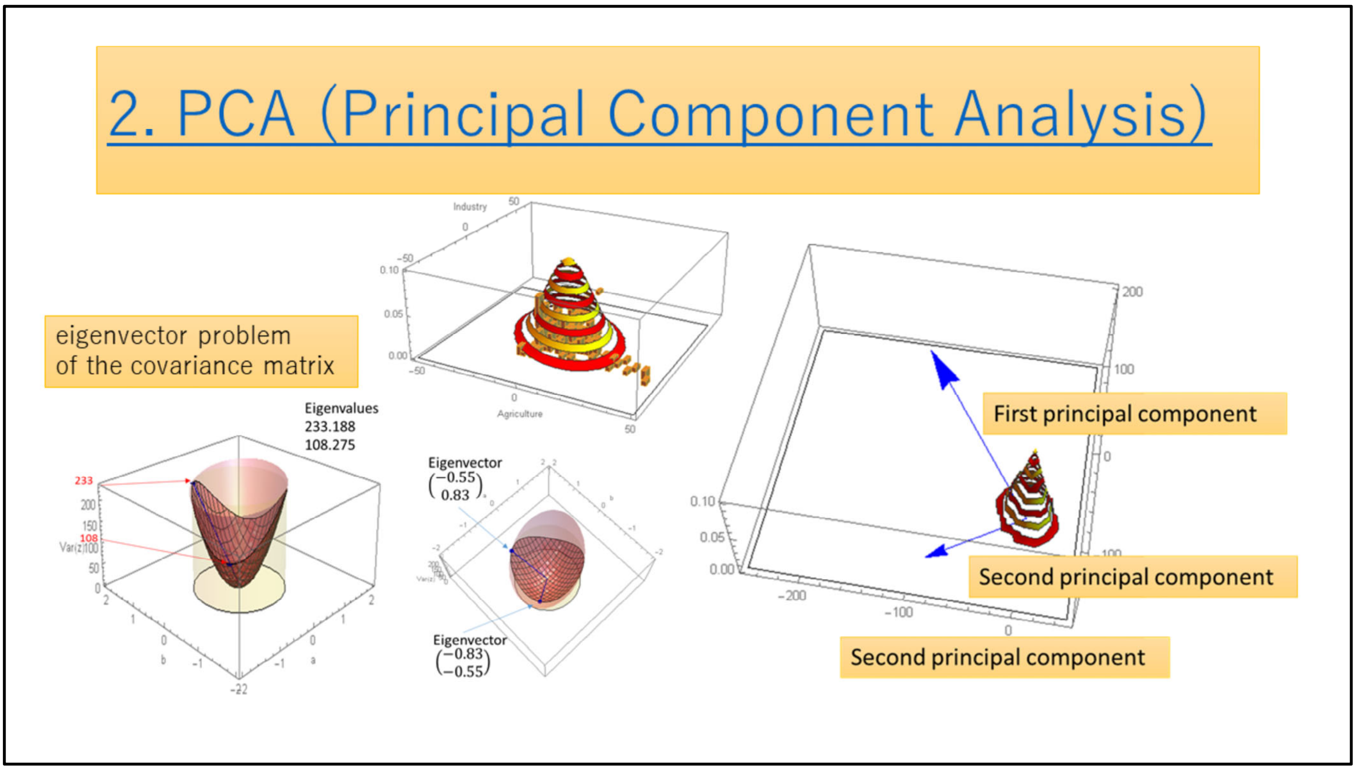
Lagrange multiplier

- In economics, this shows the change in the value of the target function when the value of the limitation rises by 1 unit at the maximum point.

$$\frac{\partial u}{\partial M} = \lambda$$

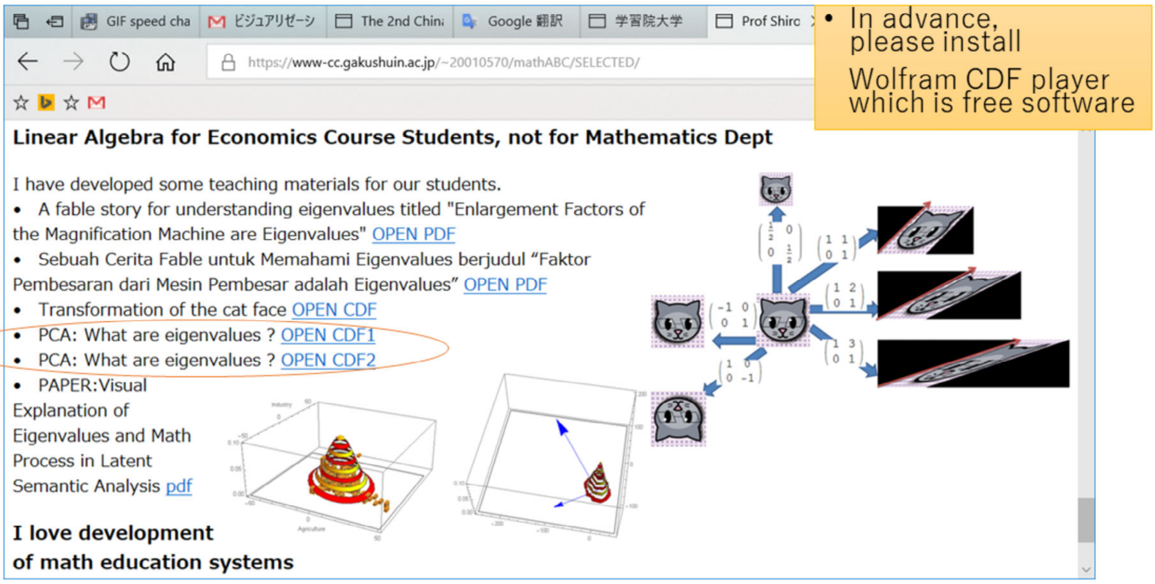
- By 1 more JPY, how much can the utility increase ?



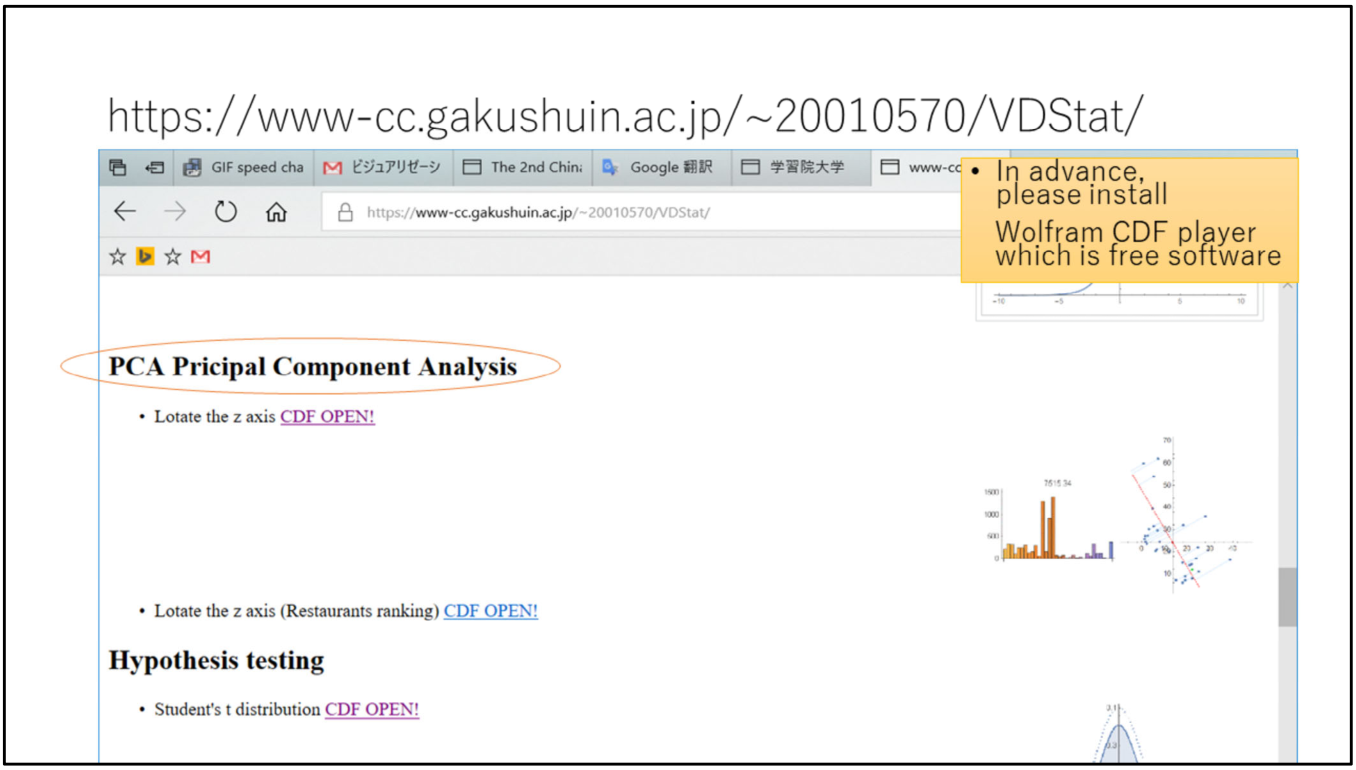


Let me move on to the next topic, principal component analysis.

https://www-cc.gakushuin.ac.jp/~20010570/mathABC/SELECTED/



The visual teaching materials are available here in my web site. Concerning the concept of the eigenvalue, I wrote the a fable story. Then Mr Lubis Muhammad translated that in Indonesian.



First let operate the PCA material.

PCA

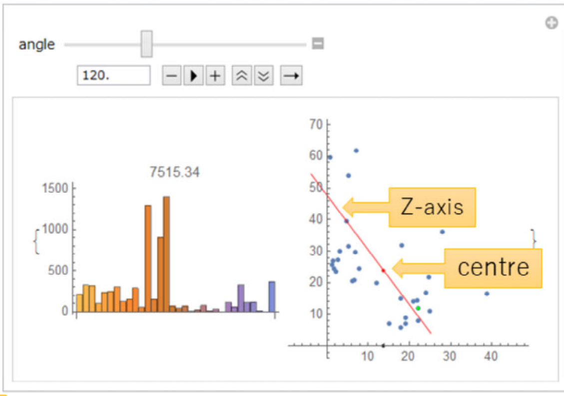
- Dimensional reduction methods
  ex: 2 dimensional data set -> 1 dimensional data set
- Calculation of the eigenvectors/eigenvalues from the given data set
- Eigenvectors are called principal components which are the essence of the given data

What is PCA ?
That is a dimension reduction method and
PCA calculates the eigenvectors of the give data set.
The eigenvectors are called principal components.
I will explain these using visual materials.

PCA: Find the best blending ratio

2 dimensional case

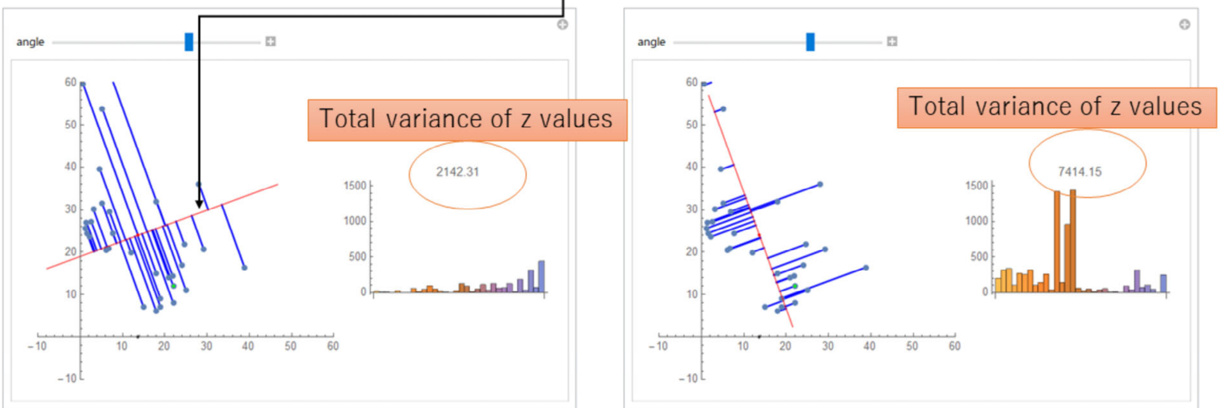
- A point has two attribute values, x and y.
- PCA: projection on the new axis *z*
  z= ax+ by
- Projection from {x, y} data to {z}
- How ?
- <u>Maximization of the variance on the z values</u>,
  so that we can spot the difference among sampling points as much as possible.



https://www-cc.gakushuin.ac.jp/~20010570/VDStat/

Suppose that the given data is 2 dimensional. X and y.
We would like to one line to be called x-axis.
Each data will have its z value which is the foot of perpendicular.
So we would like to maximize the variance of the z values/

Which angle line has a large Var(z) ?
Unknown: a and b



Let’s rotate the candidate line around the centre of the data.
In the left case, the total variance of z values are about 2000.
In the right case, the total variance of z values are about 7400.
The right one is bigger than the left.
Actually the right figure shows the maximum point of the value.
Then what is the ratio a and b at the maximum point ?

PCA: projection on the new axis *z*

z= ax+ by

- Maximize the total variance of the projected data

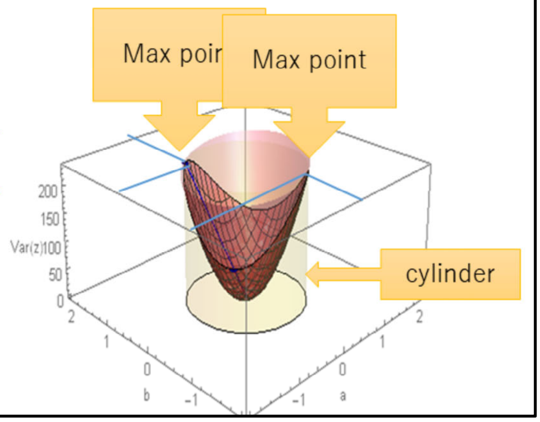
$$Var(z) = Var(a \cdot x + b \cdot y)$$
$$= a^2 \cdot Var(x) + b^2 \cdot Var(y) + 2\,a \cdot b \cdot Cov(x, y)$$

- The constraint is

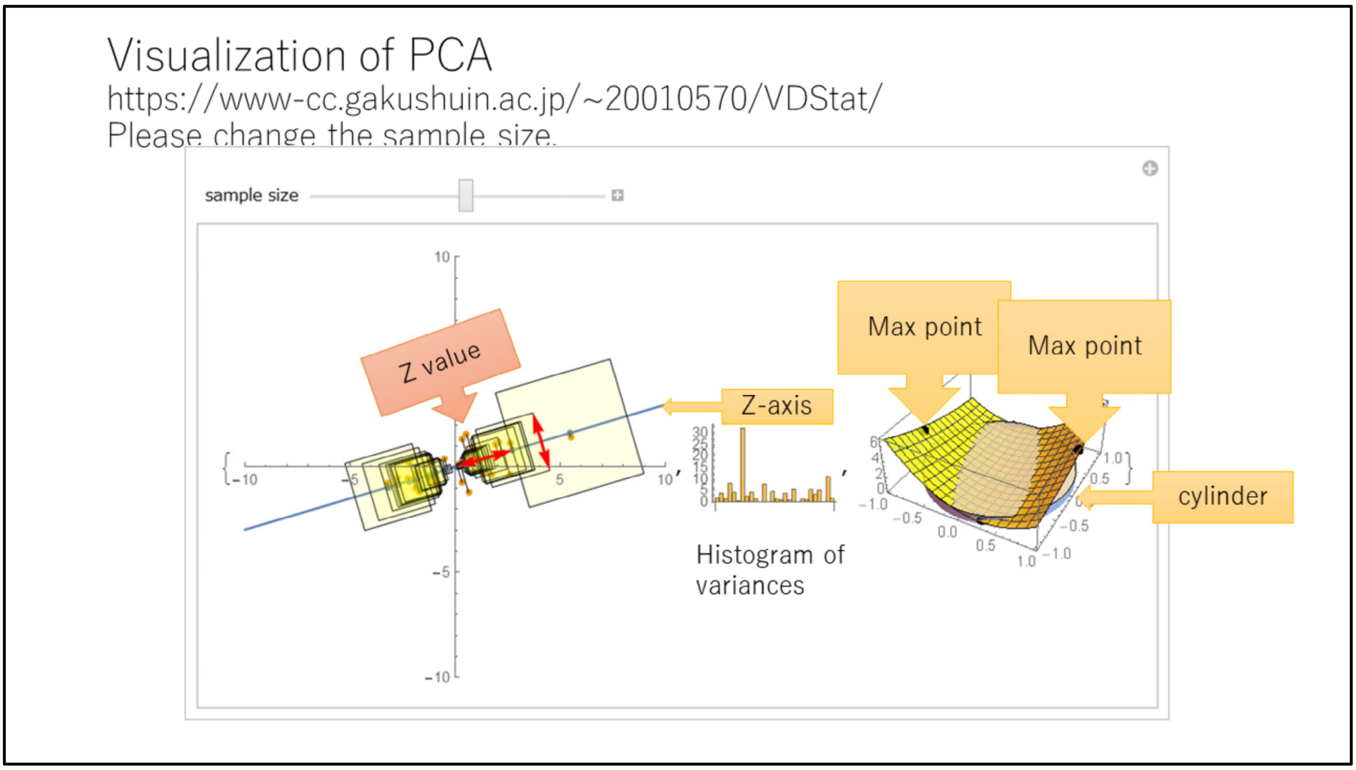
  which is a cylinder.

$$a^2 + b^2 = 1$$

- Lagrange multiplier can be used to find the a and b values at the maximum points.



The expression of total variance of z values is shown here.
The expression includes a variance of x and a variance of y, and the covariance of x and y. We can get the three values' exact figures such as 3212. So the expression's parameters are a and b.
The figure shows the var(z) in the 2 dimensional space.
We need the limitation of a and b so that we can fix one solution.
So the a squared plus b squared equals to 1 is added. Then the problem becomes the constrained optimization problem.
Visually we can solve the problem roughly. The cylinder is the limitation.
The intersection part is a complicated curve which has 2 max points and 2 min points. At the max point, we can estimate the foot of perpendicular.



Another visual material of the same PCA explanation. In the right figure, you can see the cylinder and the 2 max points. On the visual material, you can change the given data, using the slider.

PCA: projection on the new axis z

$z=ax+by$

- By Lagrange multiplier method, find the maximum blend ratio

$$\frac{\partial}{\partial a}F = 2\cdot a\cdot Var(x) + 2\cdot b\cdot Cov(x,y) - 2\lambda a = 0$$
$$\frac{\partial}{\partial b}F = 2\cdot b\cdot Var(y) + 2\cdot a\cdot Cov(x,y) - 2\lambda b = 0$$
$$a^2 + b^2 = 1$$

$$\begin{cases} a\cdot Var(x) + b\cdot Cov(x,y) = \lambda a \\ a\cdot Cov(x,y) + b\cdot Var(y) = \lambda b \end{cases}$$

The Lagrange multiplier method transformed to the determining equation.

$$\begin{bmatrix} Var(x) & Cov(x,y) \\ Cov(x,y) & Var(y) \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \lambda \begin{bmatrix} a \\ b \end{bmatrix}$$

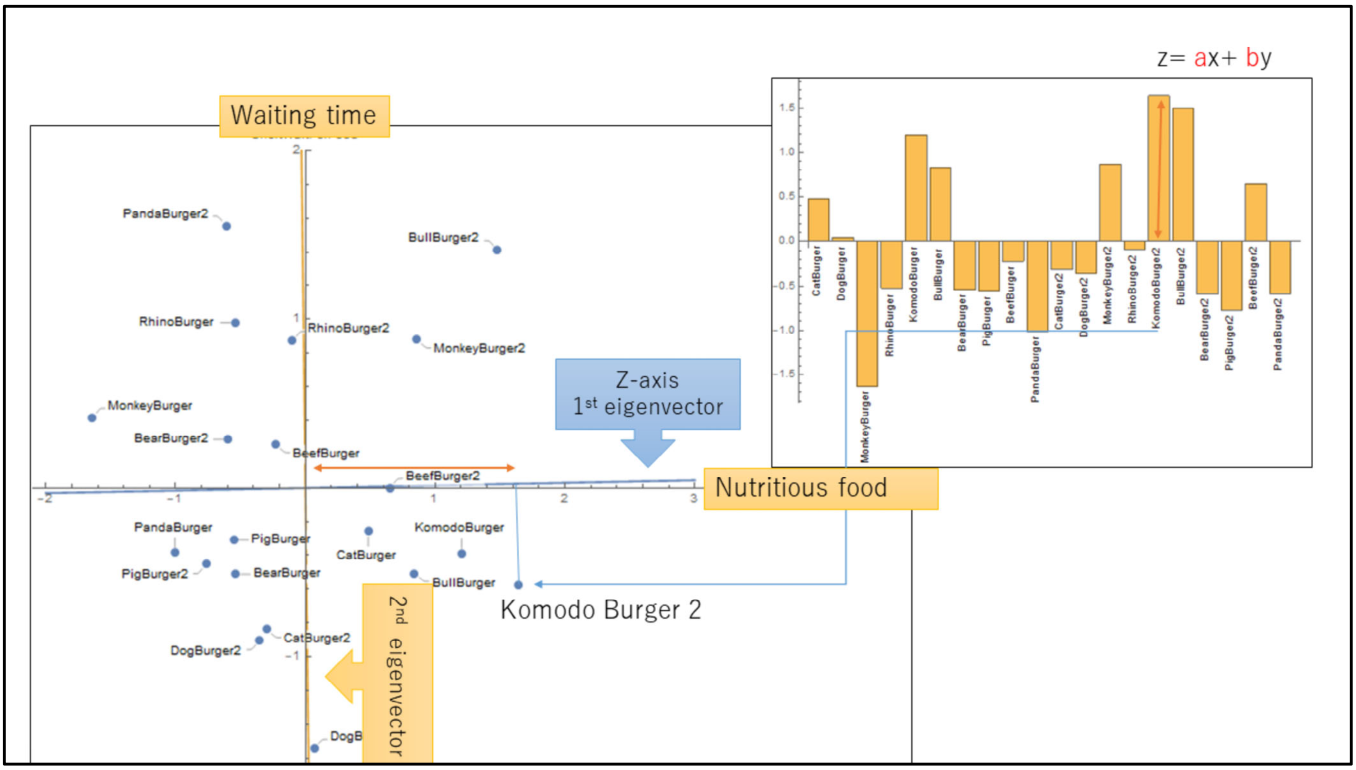
- PCA becomes a problem to find the eigenvalues and eigenvectors of the variance-covariance matrix.

Back to the Lagrange method.
To solve the problem, we work out the three first-order partial derivatives.
And equate these to zero.
Then the first and second equations are transformed to the determining equation concerning the variance-covariance matrix. It makes relationships very clear.

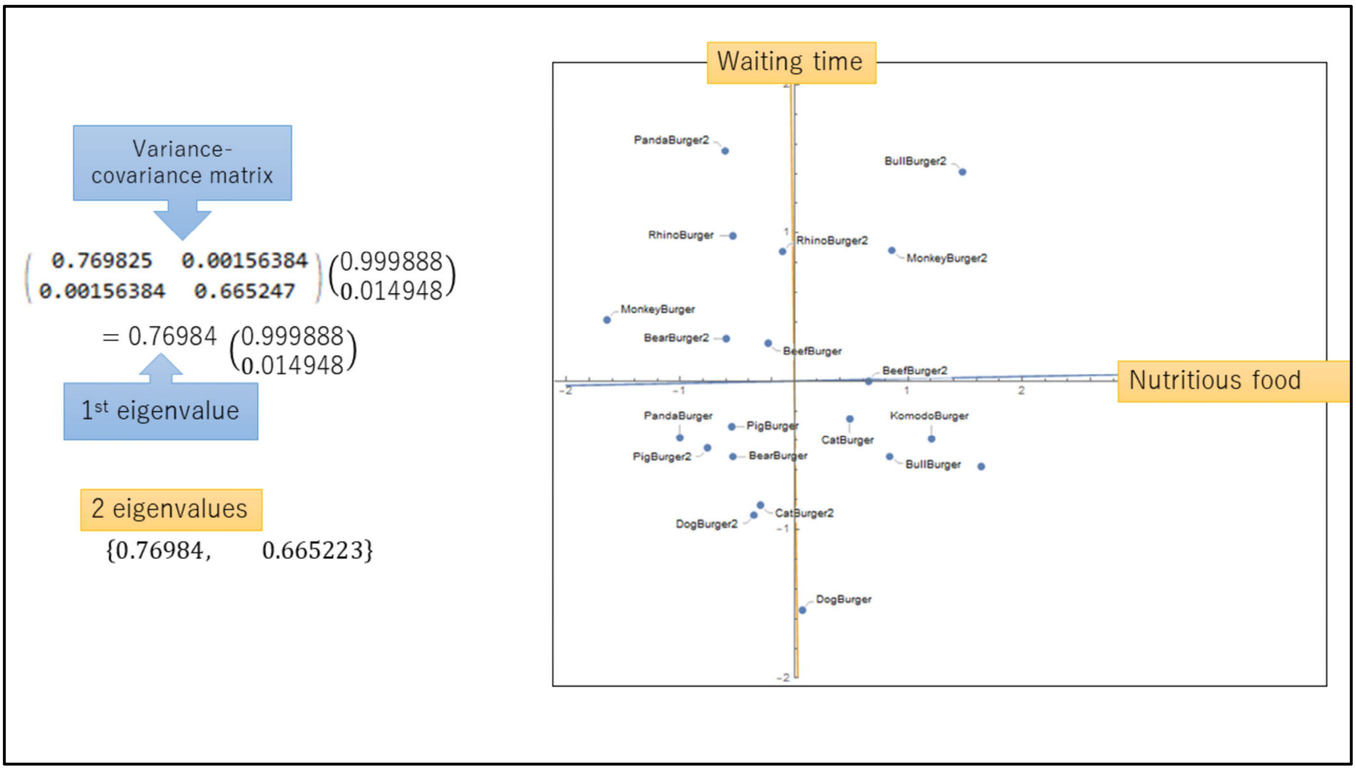
PCA another example

- Fast Food Restaurant
- Attributes
    1. Nutritious food
    2. Short waiting time
    3. Good tasted food
    4. Get what I ordered
    5. Reasonable price
- PCA of 1 and 2

Let me show you here another example of PCA.
That is the estimation of fast food restaurants.
There are many attributes by which measures the quality of restaurants.
Let me use the nutritious food and waiting time.



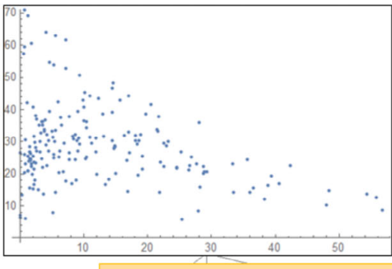
The data are artificially made ones.
After the PCA, we have the first principal and the second principal components. They are so similar to the x and y axis.
Concerning the first principal component, which has the biggest value ?
The Komodo Burger 2. Please have a look at the foot of perpendicular.
The second biggest is BullBurger 2.



Let have a look at the variance-covariance matrix.
The PCA is the determining equation problem.

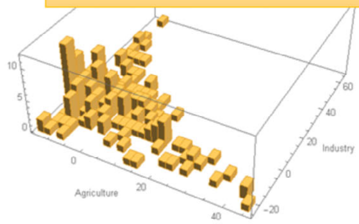
Yet Another Example
Agriculture and Industry values added (percentage of GDP)

A point corresponds to a country.



$$\begin{bmatrix} Var(x) & Cov(x,y) \\ Cov(x,y) & Var(y) \end{bmatrix} = \begin{pmatrix} 146.9 & -57.7 \\ -57.7 & 194.4 \end{pmatrix}$$

3D Histogram of 2 indices



$$\begin{bmatrix} Var(x) & Cov(x,y) \\ Cov(x,y) & Var(y) \end{bmatrix} \begin{bmatrix} -0.55 \\ 0.83 \end{bmatrix} = 233 \begin{bmatrix} -0.55 \\ 0.83 \end{bmatrix}$$

$$\begin{bmatrix} Var(x) & Cov(x,y) \\ Cov(x,y) & Var(y) \end{bmatrix} \begin{bmatrix} -0.55 \\ -0.83 \end{bmatrix} = 108 \begin{bmatrix} -0.55 \\ -0.83 \end{bmatrix}$$

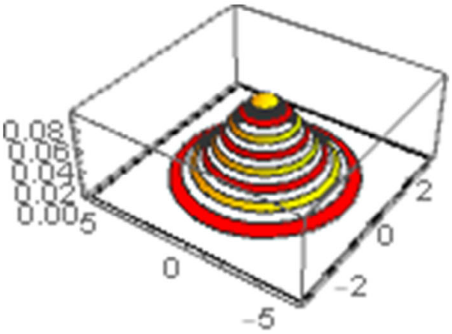
白田「主成分分析の理解に必要な数学知識の視覚的教材」,日本経営数学会誌 31(2), 59-70, 2009-11

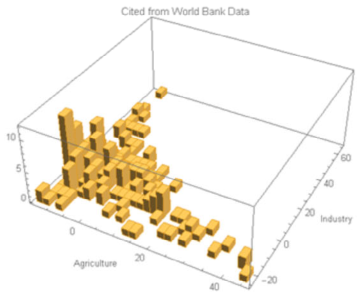
This is yet another example. The country's GDP percentages are plotted here. The x-axis shows the agricultural percentage and the y-axis shows the industry percentage. The bottom figure shows the 3 dimensional histogram of the above plot.

2D Normal (Gaussian) Distribution

- a=b=0, $\rho = 0$

$$\frac{e^{\frac{1}{2}\left(-\frac{(-b+y)(b-y-2a\rho+2x\rho)}{4(-1+\rho^2)}-\frac{(-a+x)(2a-2x-b\rho+y\rho)}{2(-1+\rho^2)}\right)}}{2\pi\sqrt{4-4\rho^2}}$$





Formula of 2D normal distribution

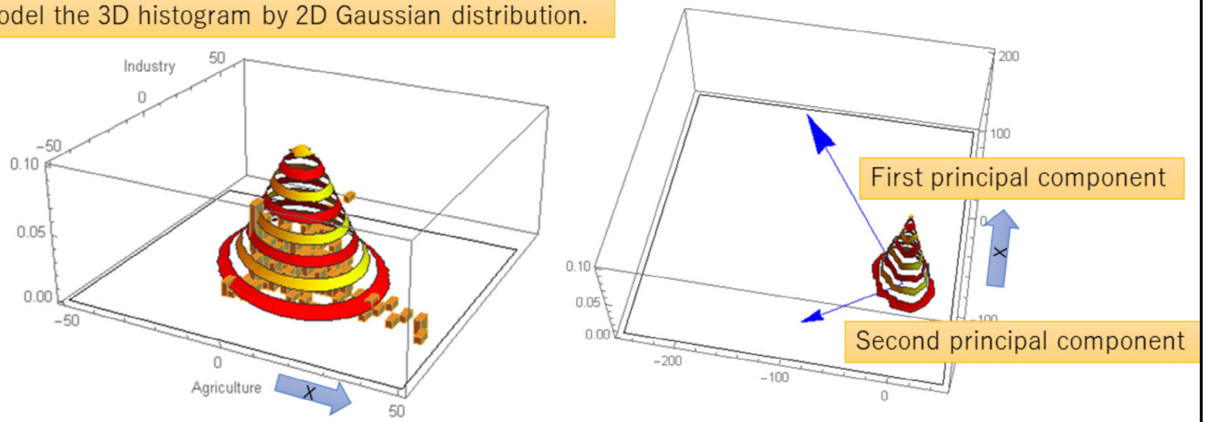
$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|\Sigma|}}\exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu})^\top\Sigma^{-1}(\vec{x}-\vec{\mu})\right\}$$

What is the meaning of the principal components.
Let's remember the 3D histogram shape.
Using the 2dimensional normal distribution, please make the similar shape of the 3D histogram.

Eigenvectors Eigenvalues

To which directions and by how much magnitudes

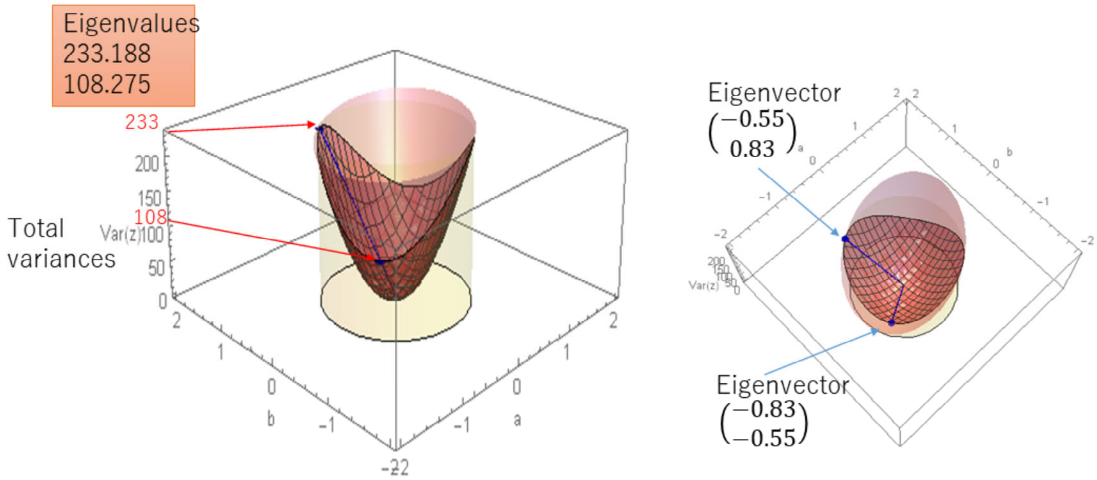
Model the 3D histogram by 2D Gaussian distribution.



The eigenvectors/eigenvalues are the essence of the given data distribution.

We made the similar shape. The shape is like a hat.
The hat shape's direction follows the 2 eigenvector directions.
And what is the meaning of eigenvalues ?
I expanded the hat-shape. The expanding amplitude is the eigenvalue.
Given the data, I will simply express the data distribution. Then, which directions and how much amplitude you expand the shape, which is the eigenvector and eigenvalue.

Surface of Var(z) with a and b

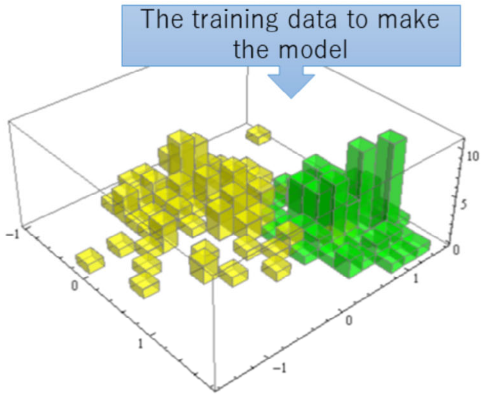


Back to the first visual materilas.
Given the data, the total variance of the z values make the surface like this. Then we find 2 max points which is the first eigenvalue, 233. And the direction from the centre to the max points are eigenvector directions.

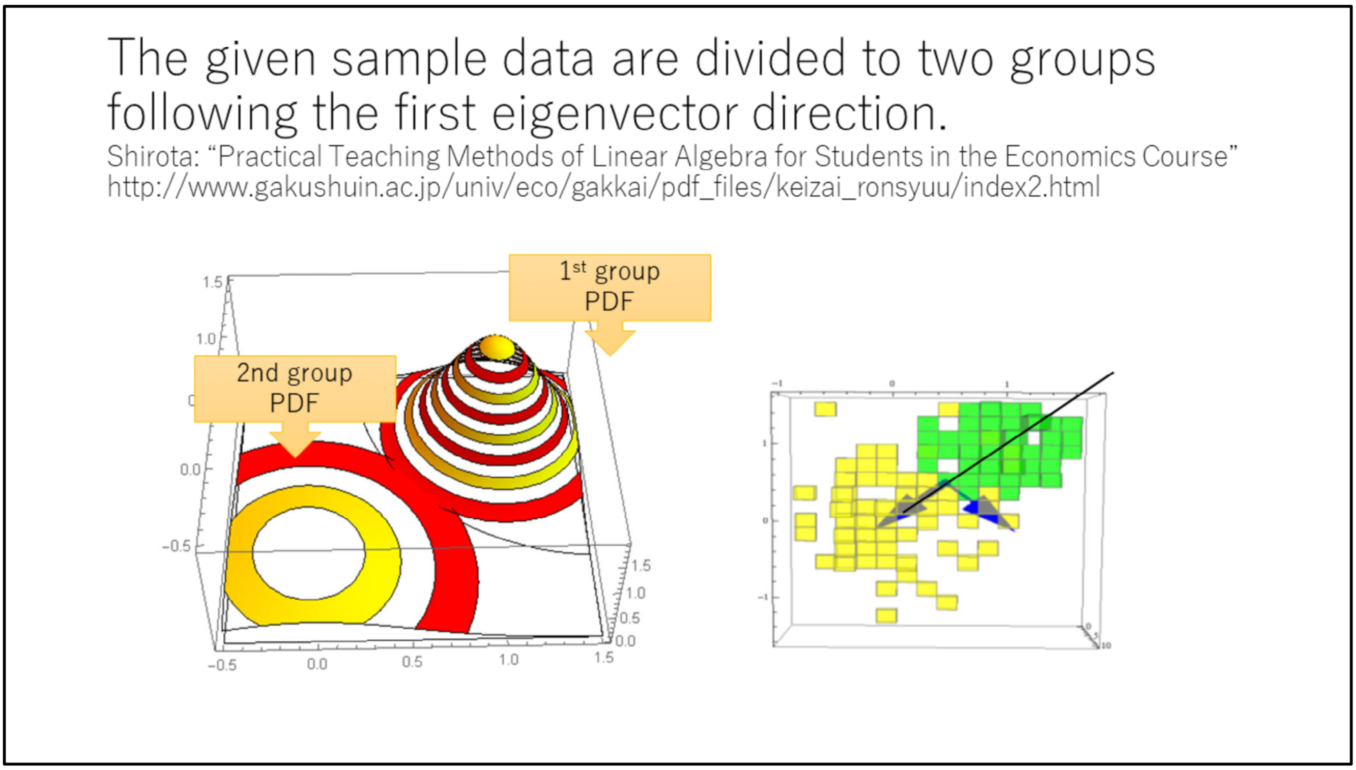
Another Usage of PCA
Clustering

- Want to make 2 groups; they are good ones (green) and no-good ones (yellow).
- First, we have no idea of the two groups.

1. PCA analysis
2. By the new z axis, make the groups.



Another usage of PCA is clustering.
Suppose that there are 2 groups. One is a good one group and another is a no-good group. We conduct the PCA on the data to get the z-axis.



Following the z-axis, we can make the clustering. The boundary is the centre. Over the centre is the one group and this side is another group. Some data will not follow this clustering.
In the left figure, the probability density functions of the 2 group are shown. The boundary is also drawn.

These visualization would help you handle the eigenvector and eigenvalue in economics mathematics.

National Income Determination

Let me move on to the next topic, national income determination problem.

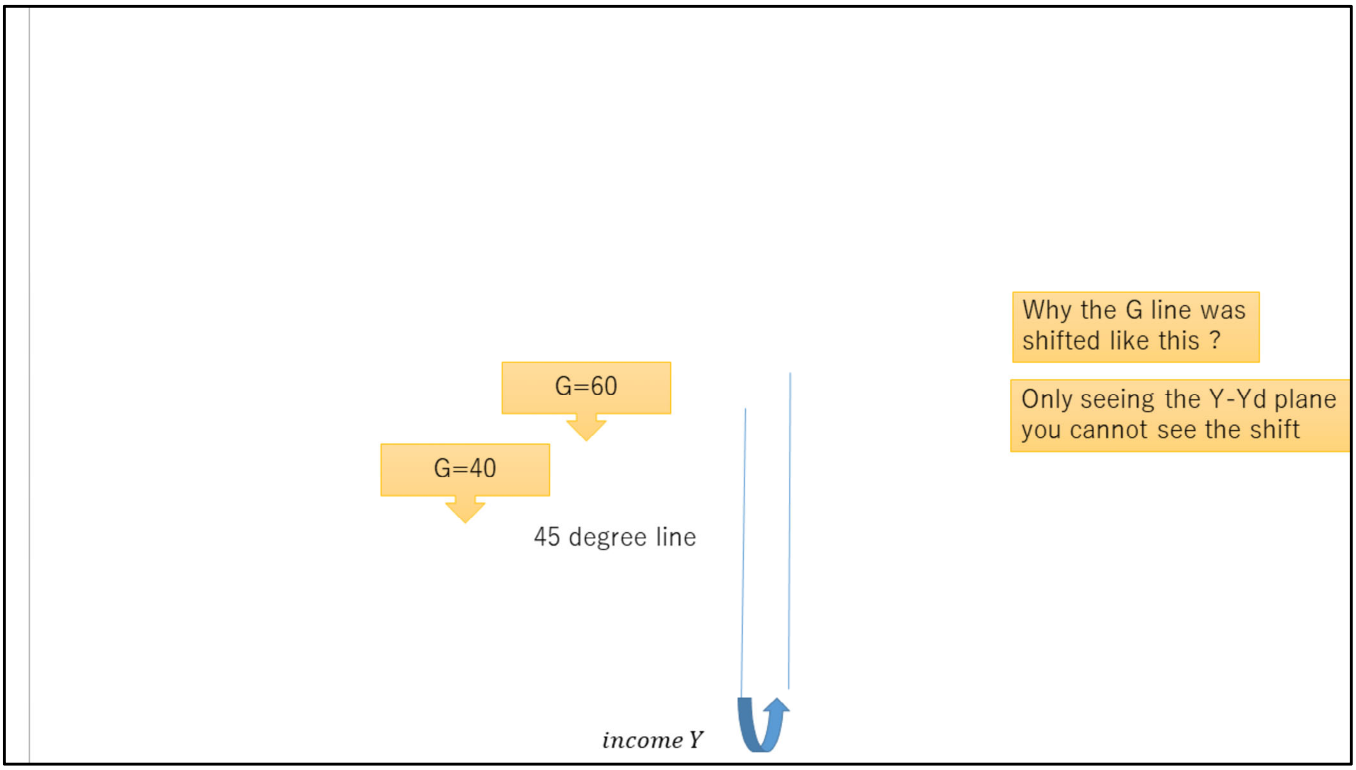
National Income Determination

Given that

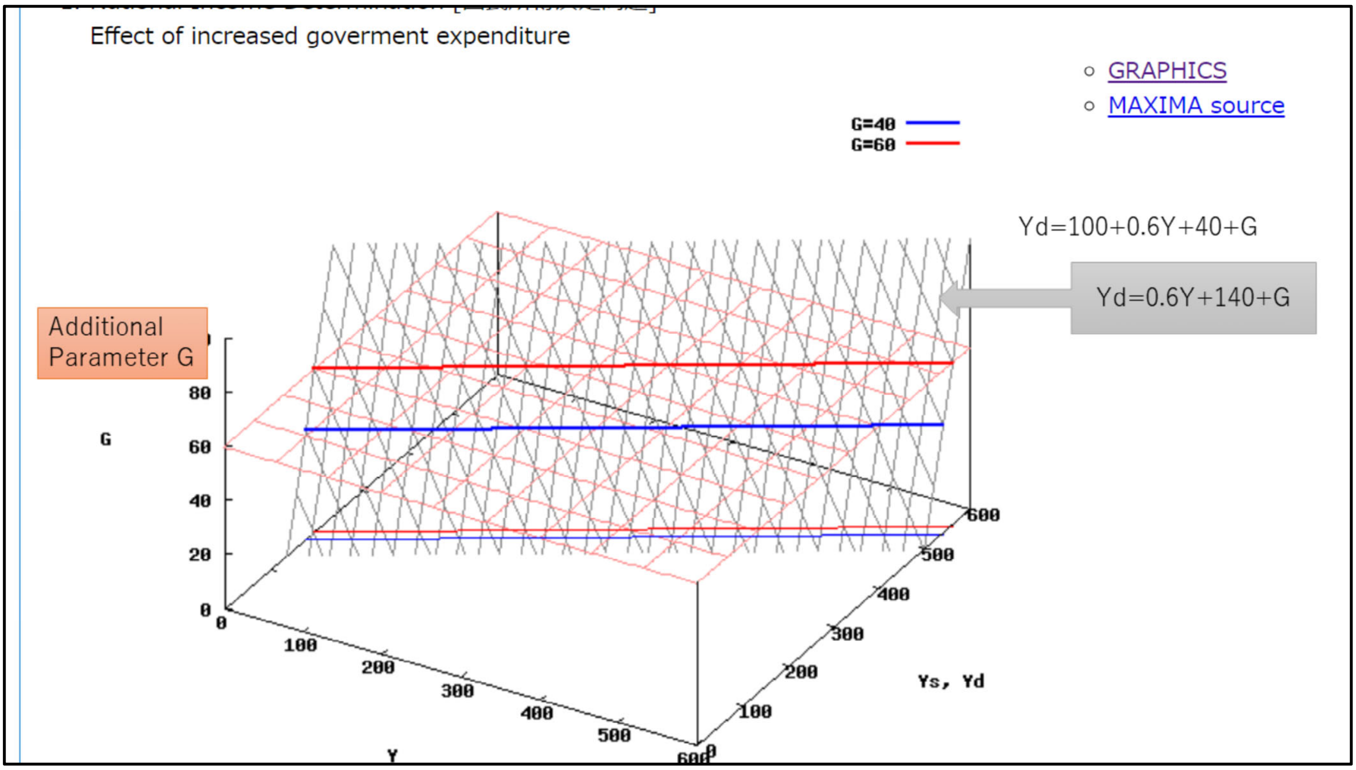
- $Yd=C+I+G$
- $C=100+0.6Y$
- $I=40$
- $G=40$
- $Y=Ys$, $Ys=Yd$

calculate the equilibrium level of national income. Then, when G increases 60, what is the equilibrium level of national income ?

Let's see the given data.
Total demand Yd is supposed to equal to consumption C plus investment I and government expenditure G.
The consumption function is this.
We assume that G and I are autonomous with fixed values 40 and 40.



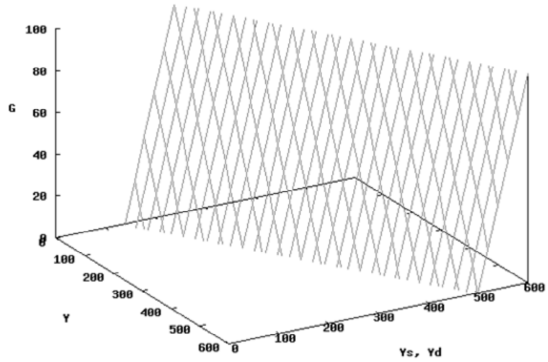
The graphics shows the Y and Yd plane.
The increase of G is expressed by the line shift like this.
So the equilibrium income level is changed.
But what makes this shift. Only seeing the Y-Yd plane, you cannot understand the shift.



So let's see the 3D space with the parameter G.
The gray plane is the Yd expression.
First make the intersection with G=40 and secondly with G=60.
Can you see the projection of the two lines on the Y-Yd plane ?

National Income Determination
https://www-cc.gakushuin.ac.jp/~20010570/private/MAXIMA/part2/

See the GIF animation



This is the GIF animation of the process.

Md=0.2*Y+400−200*r

IS schedule-LM schedule problem

Determine the equilibrium income and interest rate given the following information about the commodity market

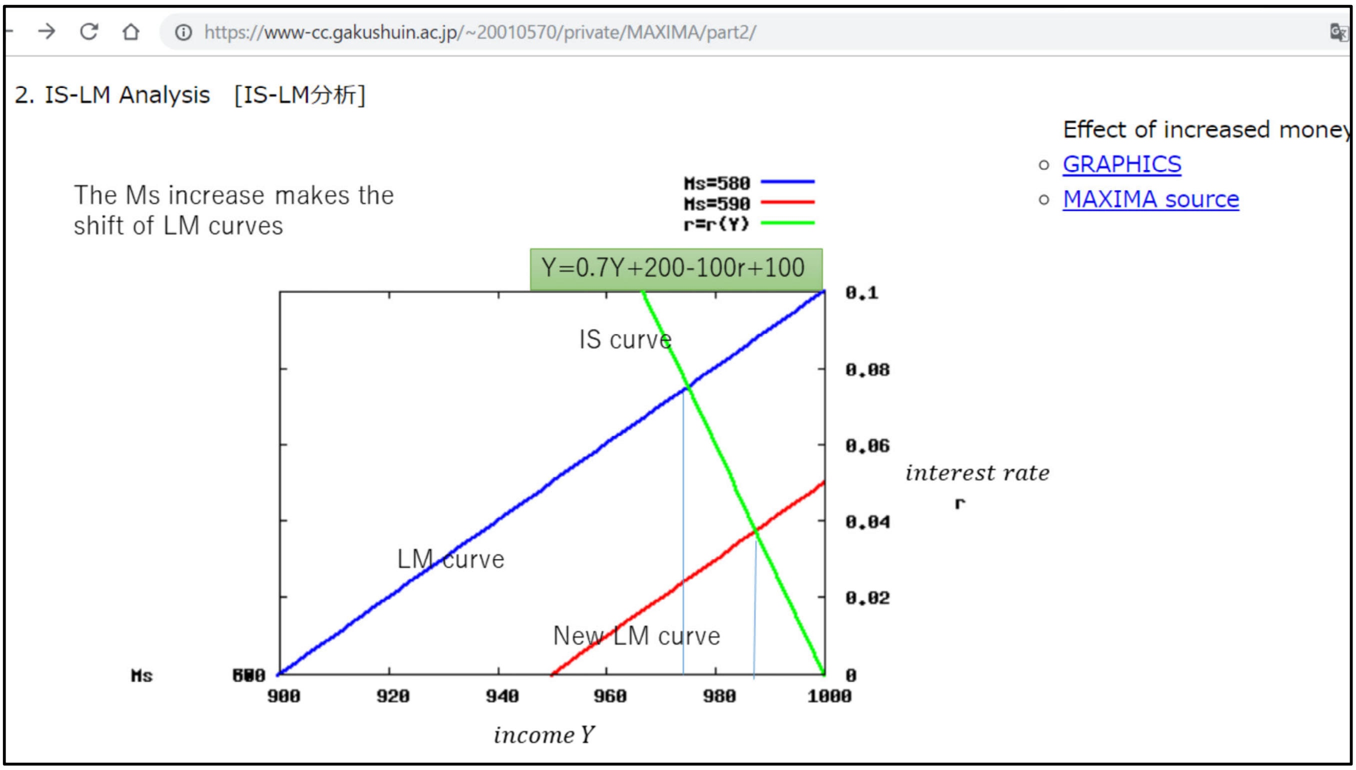
- Y=C+I
- C=0.7Y+200
- I= -100r+100

And the money market

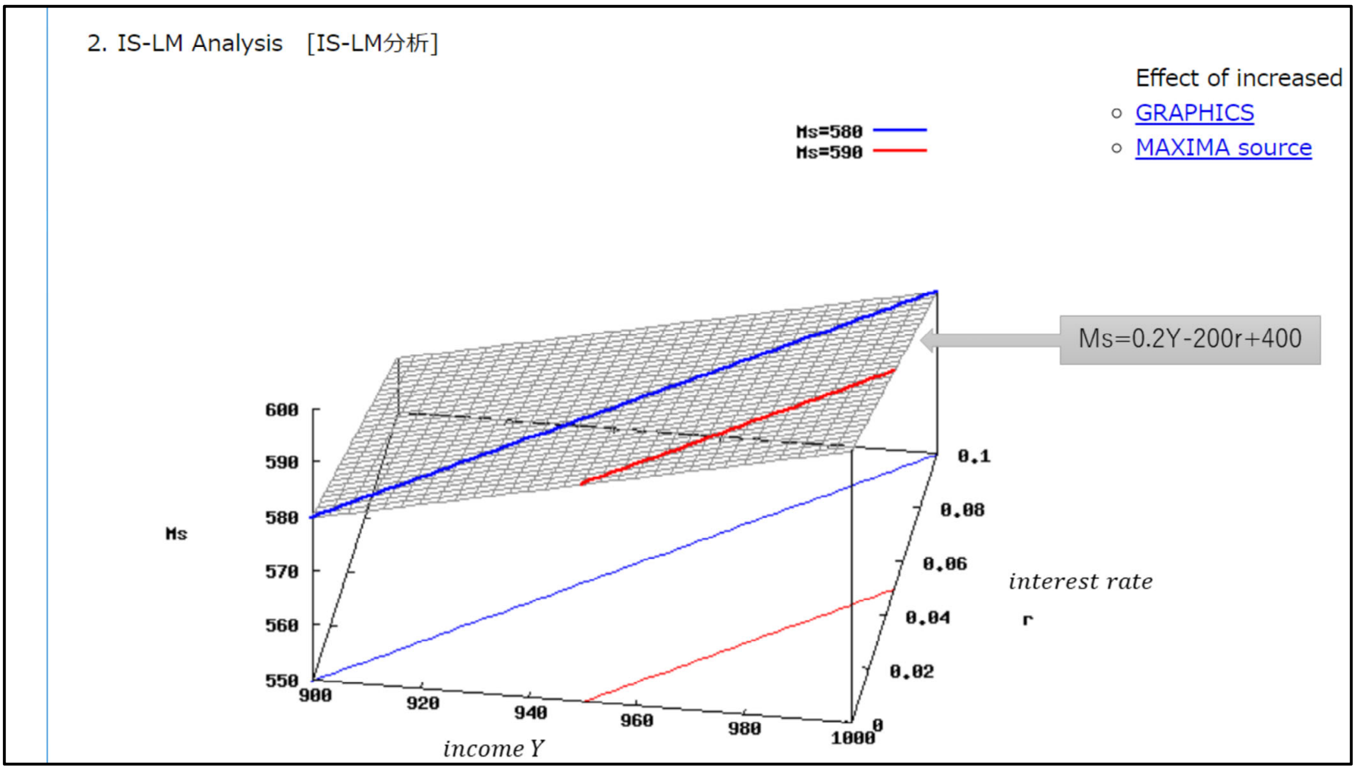
- Ms=580 → 590
- Md=L1+L2
- Precautionary demand L1=0.2Y
- Speculative demand L2= -200r+400

What effect would an increase in the money supply have on the equilibrium levels of Y and r ?

Let me move on to the next topic, IS and LM schedule problem.
From the commodity market, we can get the IS schedule, the relationship between Y and r, interest rate.
From the money market, another relationship between Y and r can be obtained, which is called LM schedule.



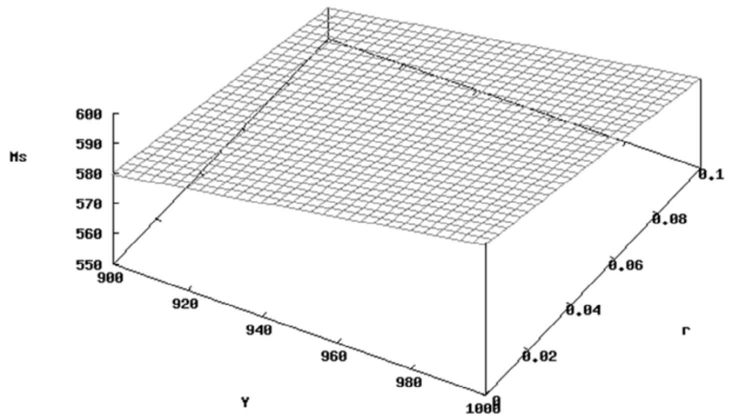
The green line is IS-schedule. The increase of money supply makes the shift of LM schedule. Then the change of the equilibrium income level happens.



On the problem, let's see the 3D space.
The additional parameter is Ms.
The gray plane is the relationship among Ms, Y and r.

IS-LM Analysis
https://www-cc.gakushuin.ac.jp/~20010570/private/MAXIMA/part2/

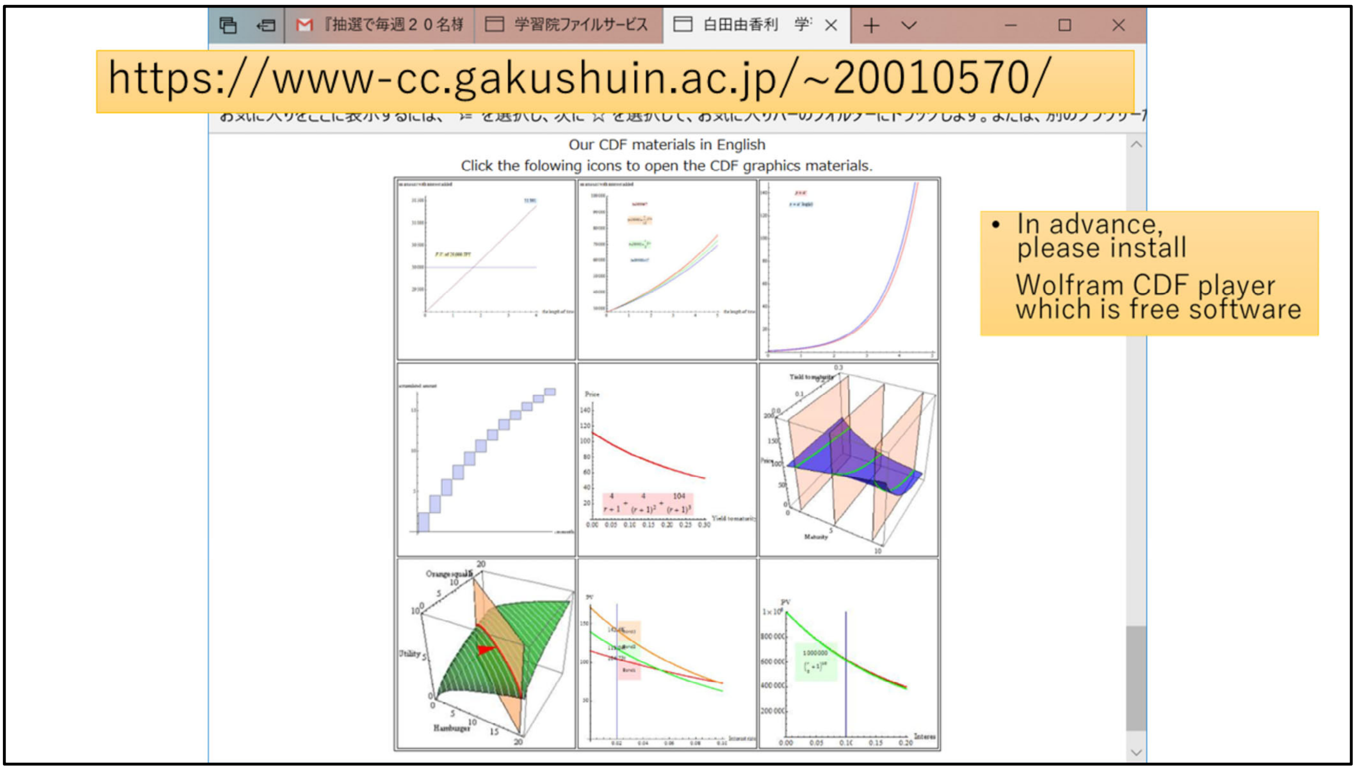
The shift happens when another variable changes.



The shift of line happens by another parameter’s change.
Drawing the 3D space like this, you may understand the shift more deeply.

Compound Interest

Let me move to the another topic, financial calculation. The topic is compound interest calculataion.

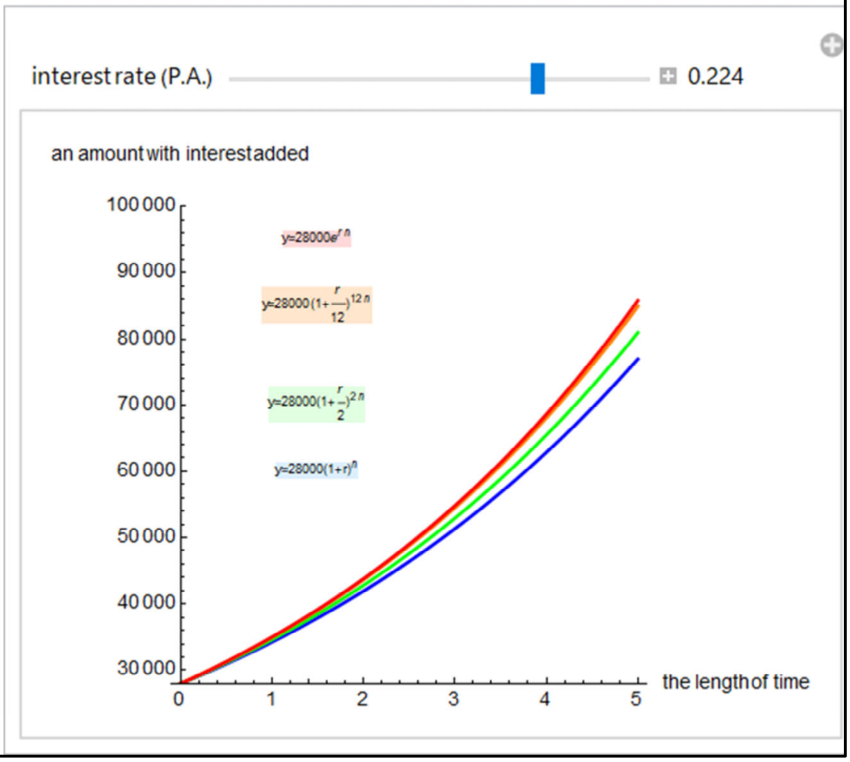


The visual materials are available here.

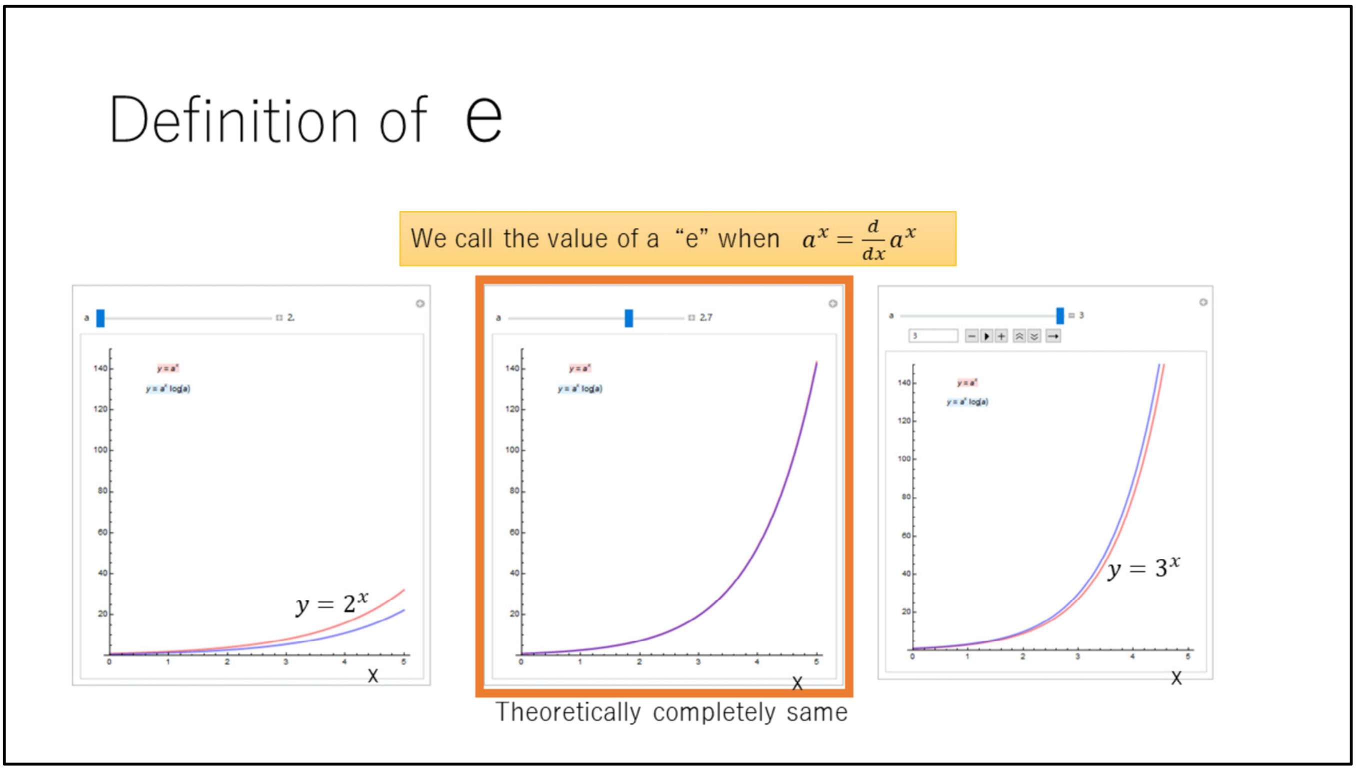
Future value if the interest is compounded

- Annually
- Semi-annually
- Monthly
- Continuous compounding

$$y = Ae^{rn}$$



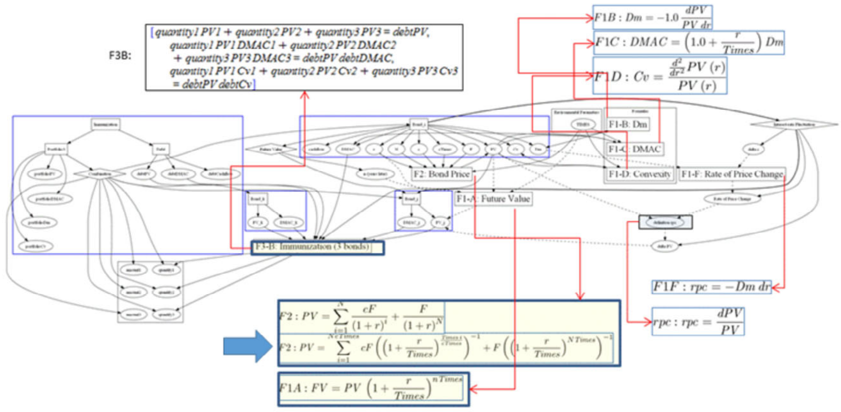
The future value of 28000 JPY under a convention of compounding m times per year is calculated. Then m=1 means one time per annually, m=2 means semiannually, m=4 means quarterly, and m=12 means monthly. The bigger the m value is, the bigger the future value becomes. The difference of the amount value between per annually and semiannually becomes bigger, if the interest rate per annually (P.A.) becomes bigger. However there is the high limit when the m value approaches infinity. Then we call the compounding "continuous compounding".



This is a graph of a^x
and its differential. Let us increase the value "a" from 2.1 to 2.8 gradually by using a slider.
When the value "a" approaches to 2.7, the curve of a^x overlaps completely its differential.
This overlap is not approximate one. That is theoretically happens.

Bond Mathematics: Price-Yield Surface

A Concept Model of Bond Mathematics



Shirota Y., Hashimoto T., Stanworth P. (2013) Knowledge Visualization of the Deductive Reasoning for Word Problems in Mathematical Economics. In: Madaan A., Kikuchi S., Bhalla S. (eds) Databases in Networked Information Systems. DNIS 2013. Lecture Notes in Computer Science, vol 7813. Springer, Berlin, Heidelberg

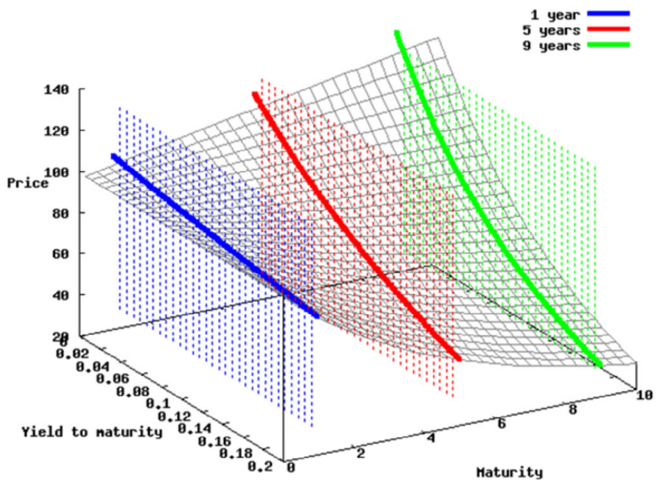
Let me move on to the next topic, bond mathematics.
When I teach bond mathematics, I use this formula diagram.
What is the given data ? What is the unknown ?
Please tick them on the diagram and find the missing link between them.
The formula on the missing link should be used to solve the problem.
That is the deductive reasoning. By deductive reasoning, I will teach the bond mathematics.
Today I will explain this formula of present value.

https://www-cc.gakushuin.ac.jp/~20010570/private/MAXIMA/

$$PV = \sum_{k=1}^{n} \frac{cF}{(1+r)^{i}} + \frac{F}{(1+r)^{n}}$$

Coupon rate c : 4%

- Price-yield curves with yield to maturity and three maturities
- The longer the maturity becomes, the bigger the modified duration is.



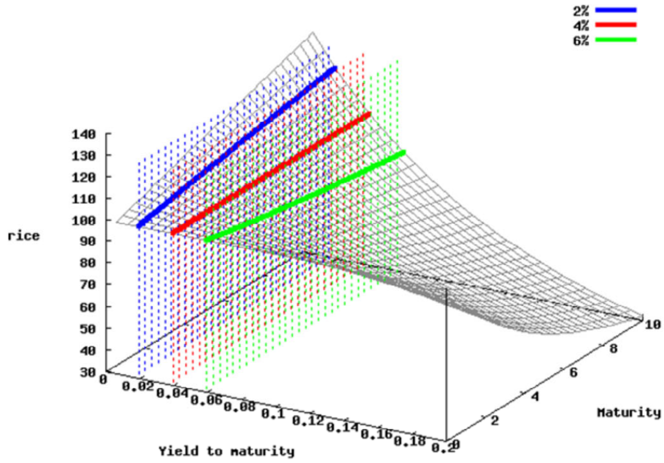
Let me explain the PV formula which is the summation of coupons and the face value. The maturity is n years.
This function has parameter a yield to maturity r and maturity n years.
The formula is expressed by the 3D surface. When we cut the year=1, the intersection is obtained like this.
Compare the intersection curves among different maturity years.
At the par point, the long maturity one has the steepest decline.

Evolution of the bond price over time: A bond price approaches the face value over time.

$$PV = \sum_{k=1}^{n} \frac{cF}{(1+r)^i} + \frac{F}{(1+r)^n}$$

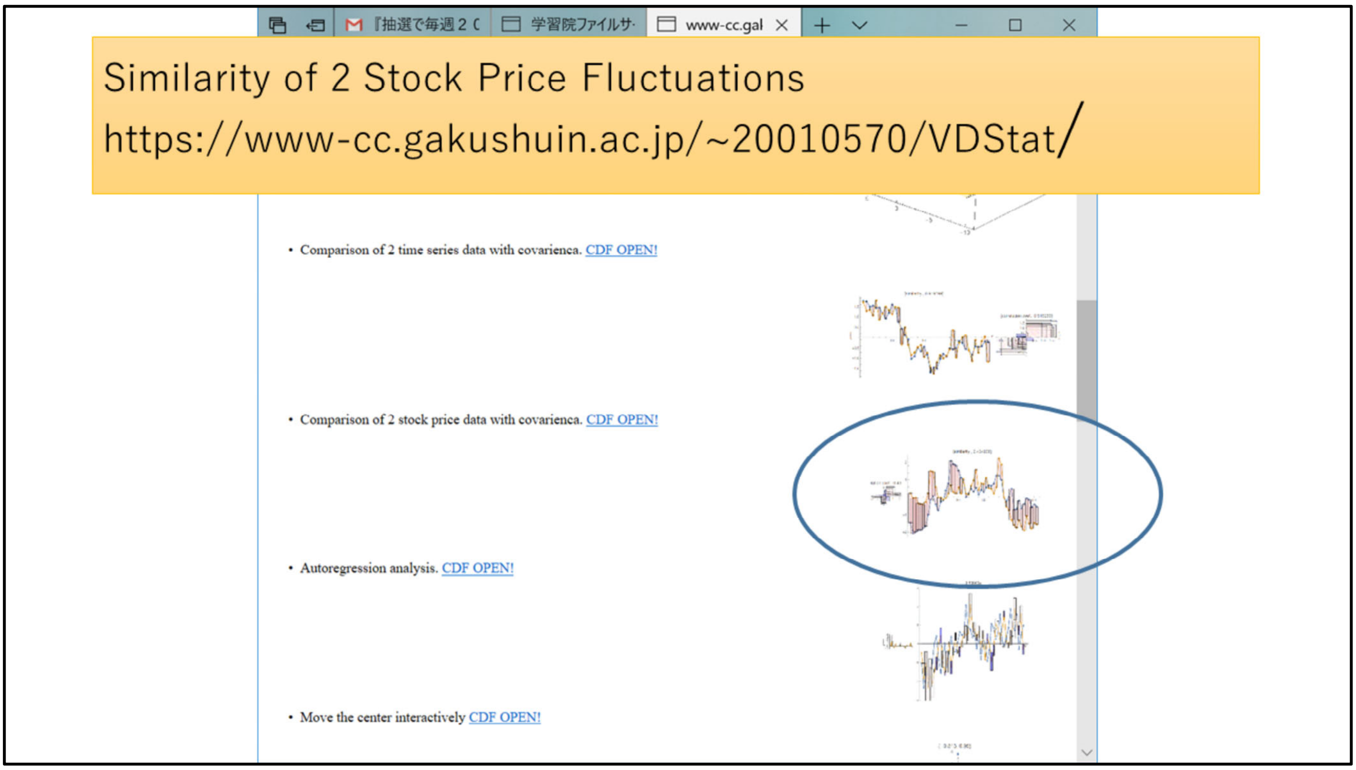
Coupon rate c : 4%

Comparison of three yield to maturities 2%, 4%, and 6%



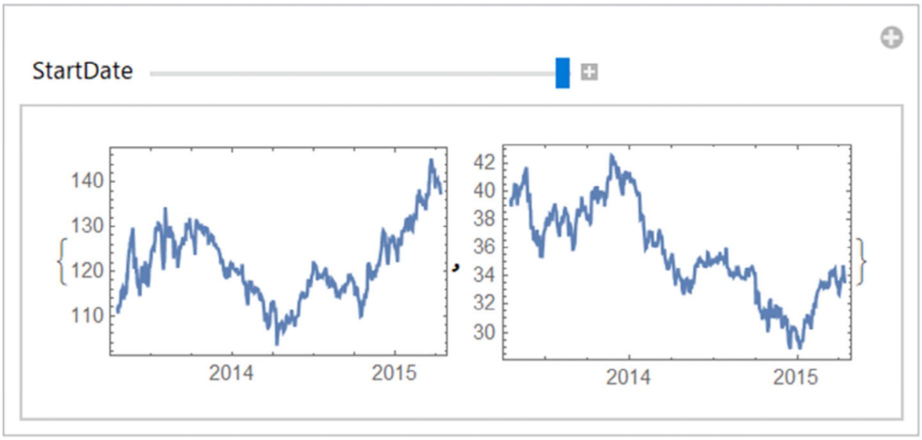
Next let's see the price change over time.
A bond price approaches the face value 100 over time.
The 3D surface will be cut by the different yield to maturity planes.
Let's see the intersection curves.
Because in the case the coupon rate is 4 %, the yield to maturity 4% one keeps the horizontal line.

In the bond mathematics, visualization of the given function will make you understand the current and future situations like this.



Let me move on to the next topic, stock price fluctuations.

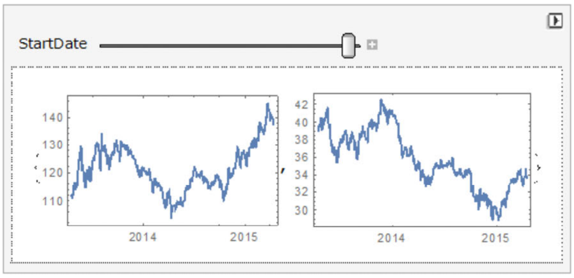
Two Companies Stock Price Fluctuations



Let's consider two companies stock price movement.
I would like to measure the similarity level between them.

Another Interpretation of "Covariance"
Similarity between Time Series Data $\{x_i\}$ and $\{y_i\}$

Stock Price Time Series Data $\{x_i\}$ and $\{y_i\}$

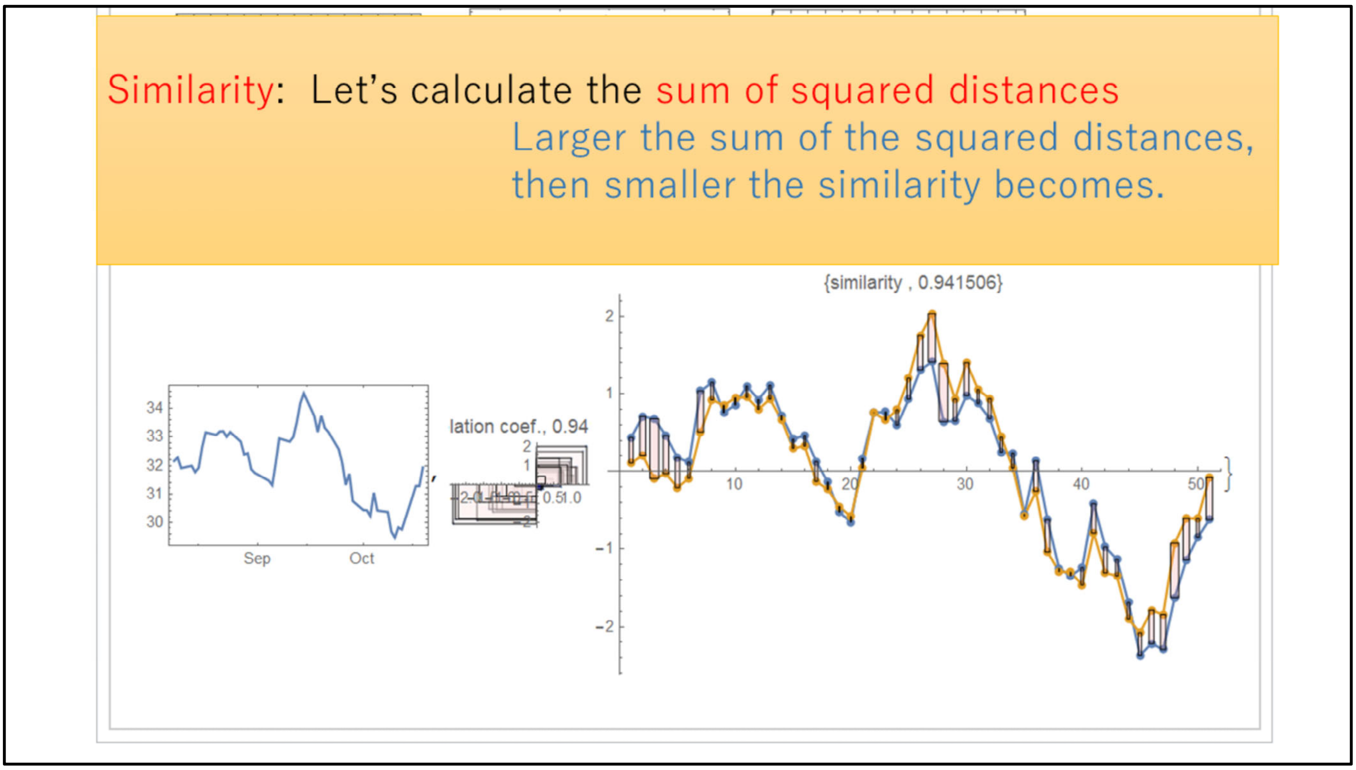


Initially, standardization

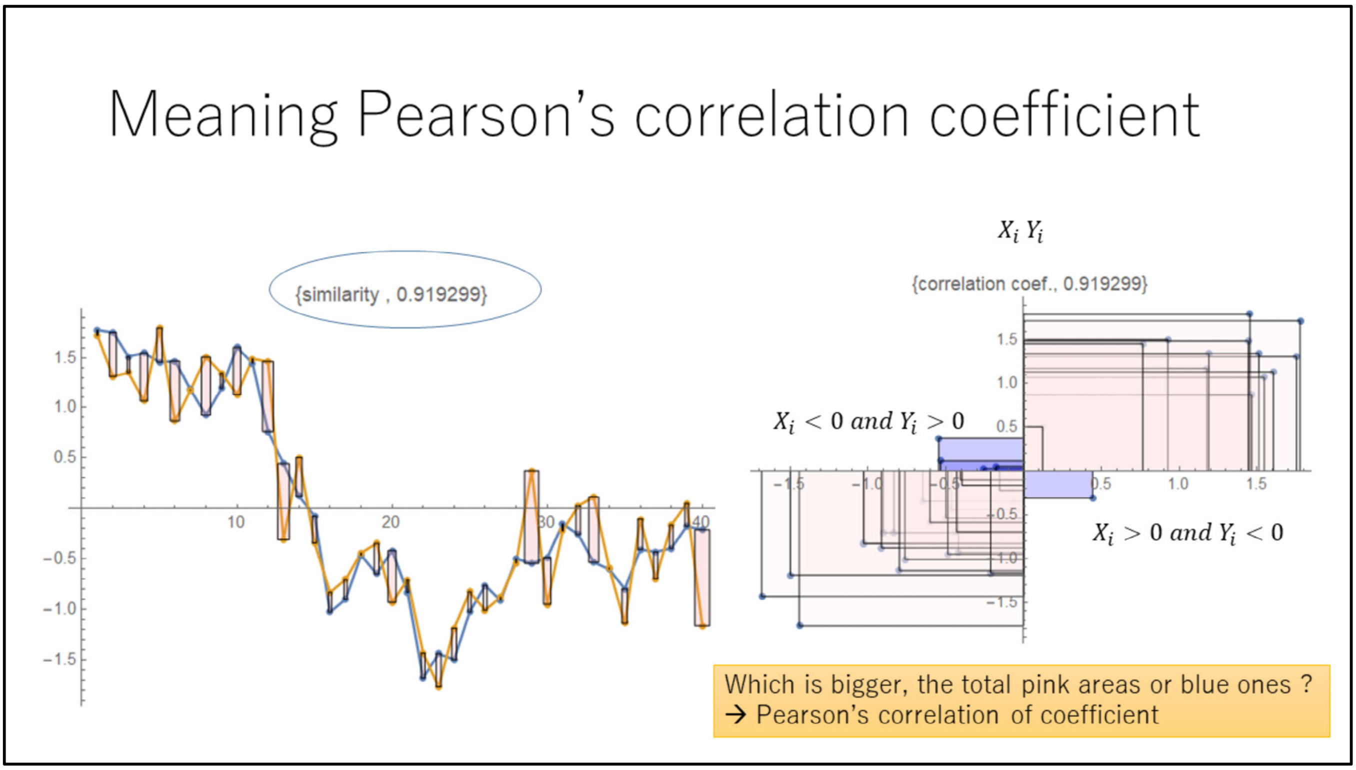
$$\bar{x} = 0, \quad \bar{y} = 0, \quad \sqrt{\frac{\sum (x_i)^2}{n}} = 1, \quad \sqrt{\frac{\sum (y_i)^2}{n}} = 1$$

I will show you one method of similarity level calculations.

Currently, we make the robots make a decision of trading instead of human beings. Then algorithms are so important. The similarity level calculation is also needed as a tool function.
Variable x represents A company and variable y does B company.
As you can see, there is a big gap between them. So standardization is needed in advance. Averages are set to be zero and the standard deviations to be 1.



This is a shot of after the standardization.
The data move around the zero horizontal line.
In this case, the similarity level is high, the value is 0.94.
The gap which is expressed by the pink rectangle is small.



Let me explain the Pearson's correlation coefficient.
Let's make the scatter graph of each point and paint the area x times y.
If both are positive, color the area in pink. If they are positive and negative, then pain it in blue. So calculate the total areas of which is bigger.
In the above case, the pink area is bigger. The correlation also becomes positive. The similarity level becomes 0.919.

Similarity: Let's calculate the sum of squared distances

$$\sum (x_i - y_i)^2$$

Larger the sum of the squared distances, then smaller the similarity becomes.

$$= \sum (x_i{}^2 - 2x_i y_i + y_i{}^2) = \sum x_i{}^2 - 2\sum x_i y_i + \sum y_i{}^2$$

$$= n - 2\sum x_i y_i + n = 2n - 2\sum x_i y_i = 2\left(n - \sum x_i y_i\right)$$

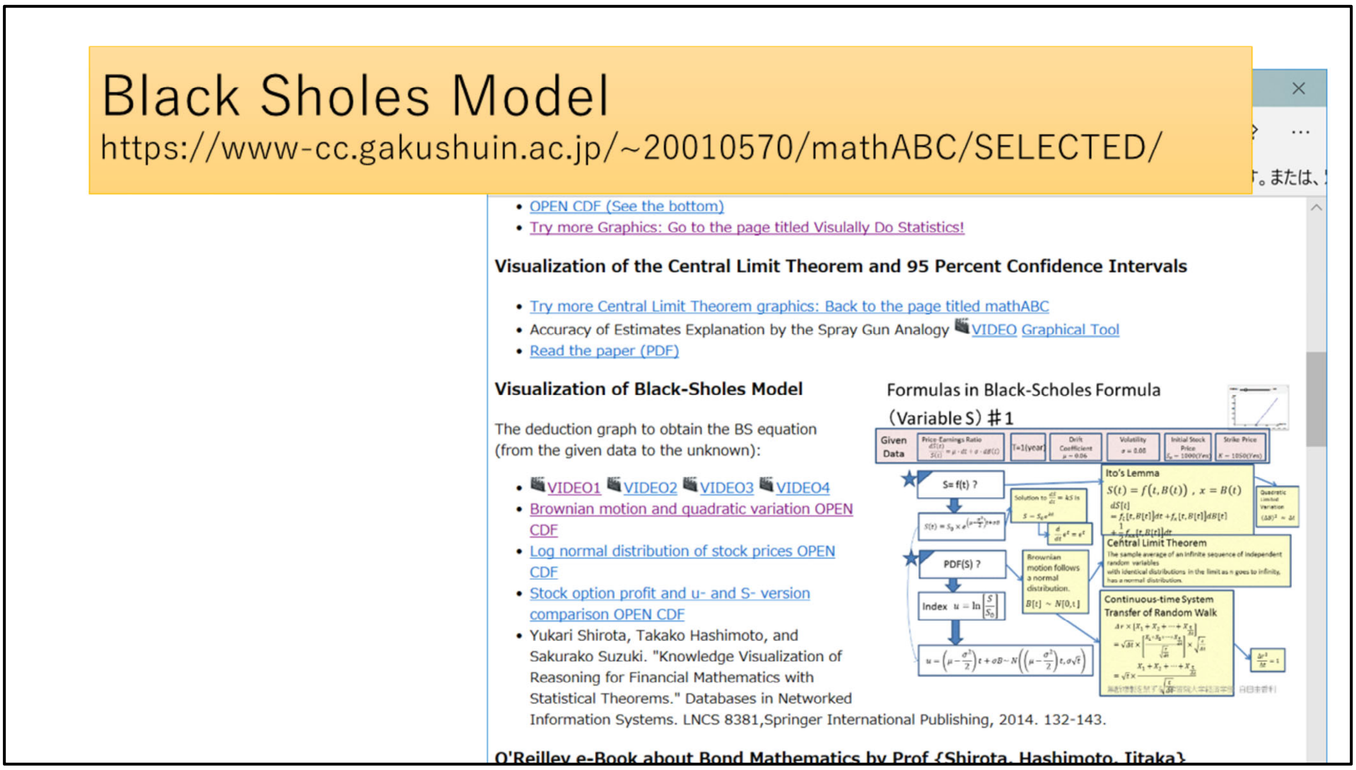
Then

$$\sum x_i y_i = n - \frac{1}{2}\sum (x_i - y_i)^2$$

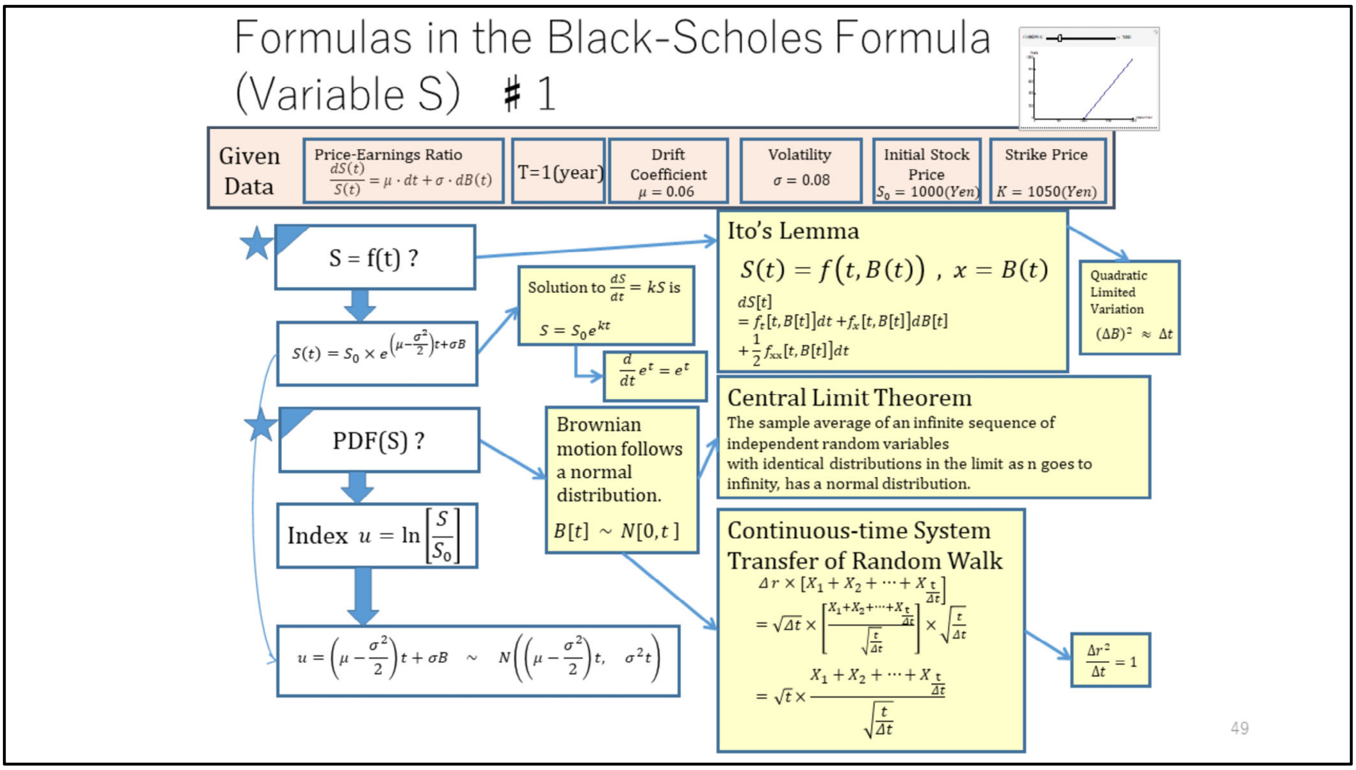
The Correlation Coefficient which shows the similarity

$$\boxed{\frac{1}{n}\sum x_i y_i} = 1 - \frac{1}{2n}\underbrace{\sum (x_i - y_i)^2}_{\text{Sum of squared distance}}$$

if every $x_i = -y_i$ then $1 - \frac{1}{2n}\sum(-2y_i)^2 = 1 - \frac{1}{2n}\cdot 4n = \underline{-1}$



Lastly, I would like to show you the Black Sholes model visually.
In my class, I use the deductive reasoning graph between the given data and the unknown which is the expected value of the option.
In the deductive graph, the formula is colored in yellow. They are Central Limit Theorem and Ito’s Lemma or so.
Today I will show you a part of the materials.



We can see this derivation as a **deductive reasoning process**.

This problem has these **given data**…
<click for given data boxes to appear>

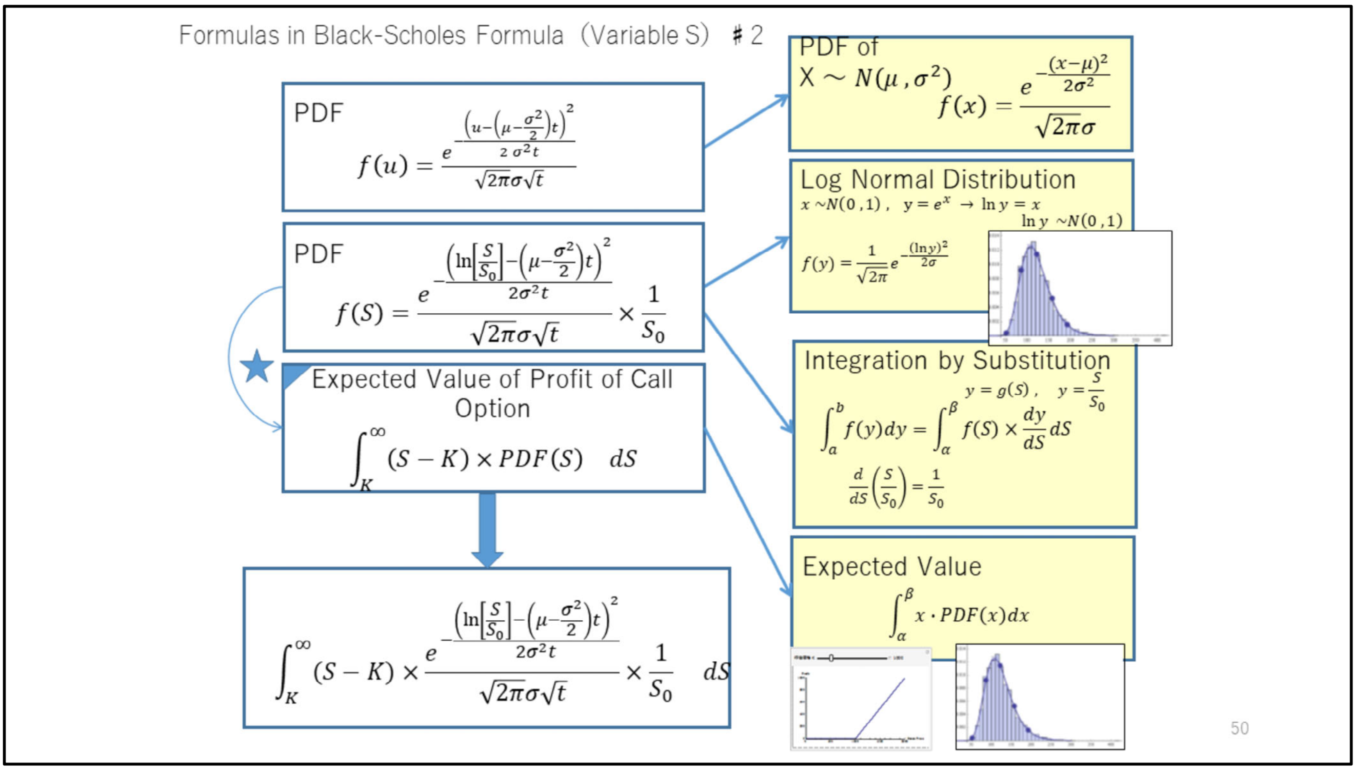
and the **unknown**.

So what is the unknown in this problem ?
It is S, the **expected value** of the **call option p**rofit.
<click for white boxes to appear>

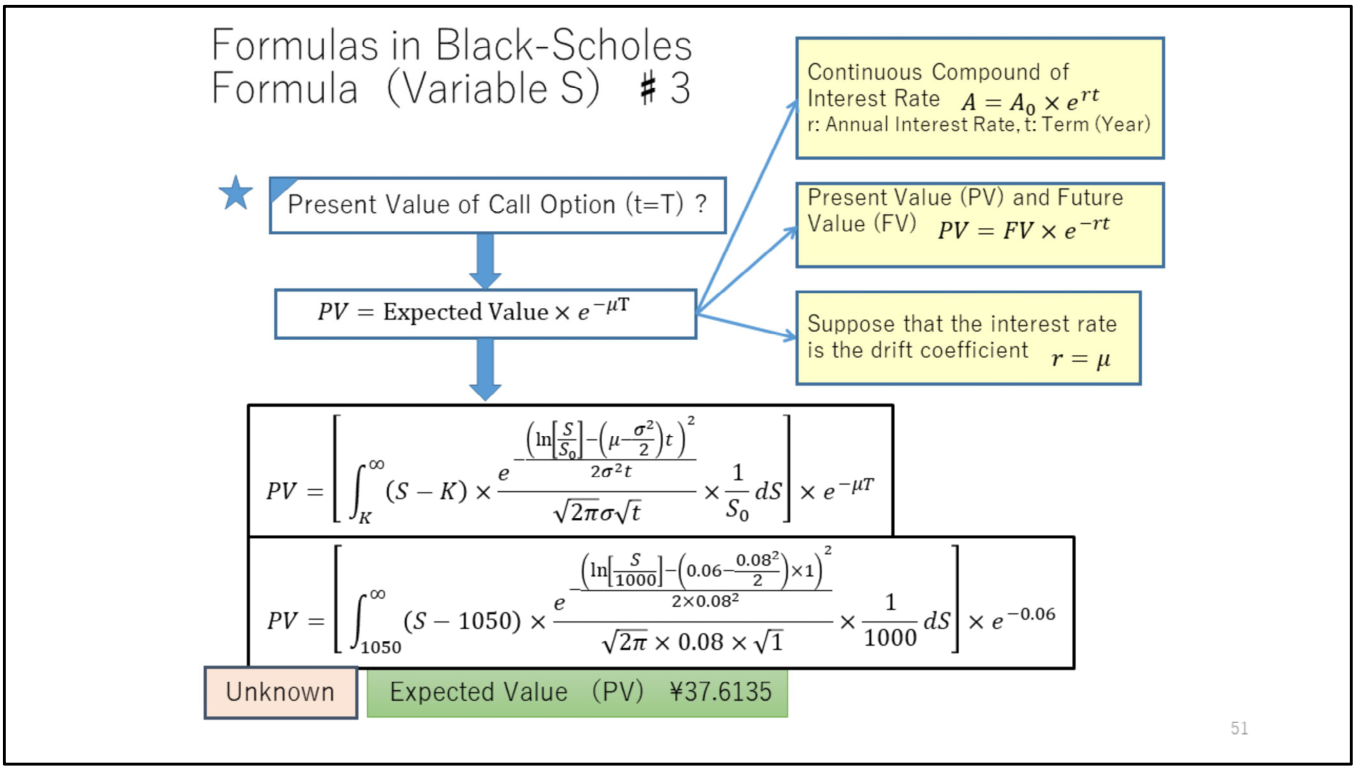
So your task is to connect between the given data <pause> and the unknown <pause> by combining several theories and formulas. <click for yellow boxes to appear>

We will make use of the Central Limit Theorem, Ito's Lemma and other important formulas.

When you're using these teaching materials, you can click or select the name of a theory you are interested in, then you will hear an explanation (the way we do in YouTube).



<Not sure what to say here>

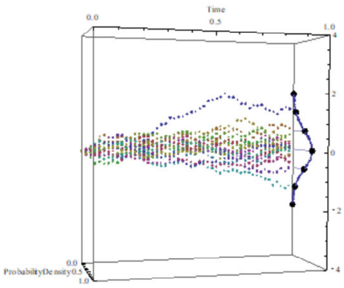


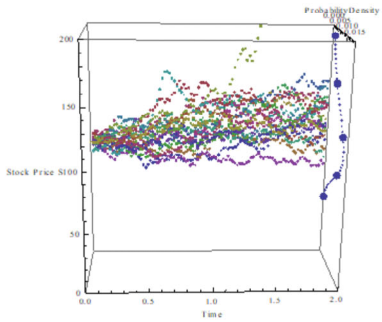
<Not sure what to say here>

Brownian Motion

Where is a particle at time t ? B(t)

How much is a stock value at time t ? S(t)

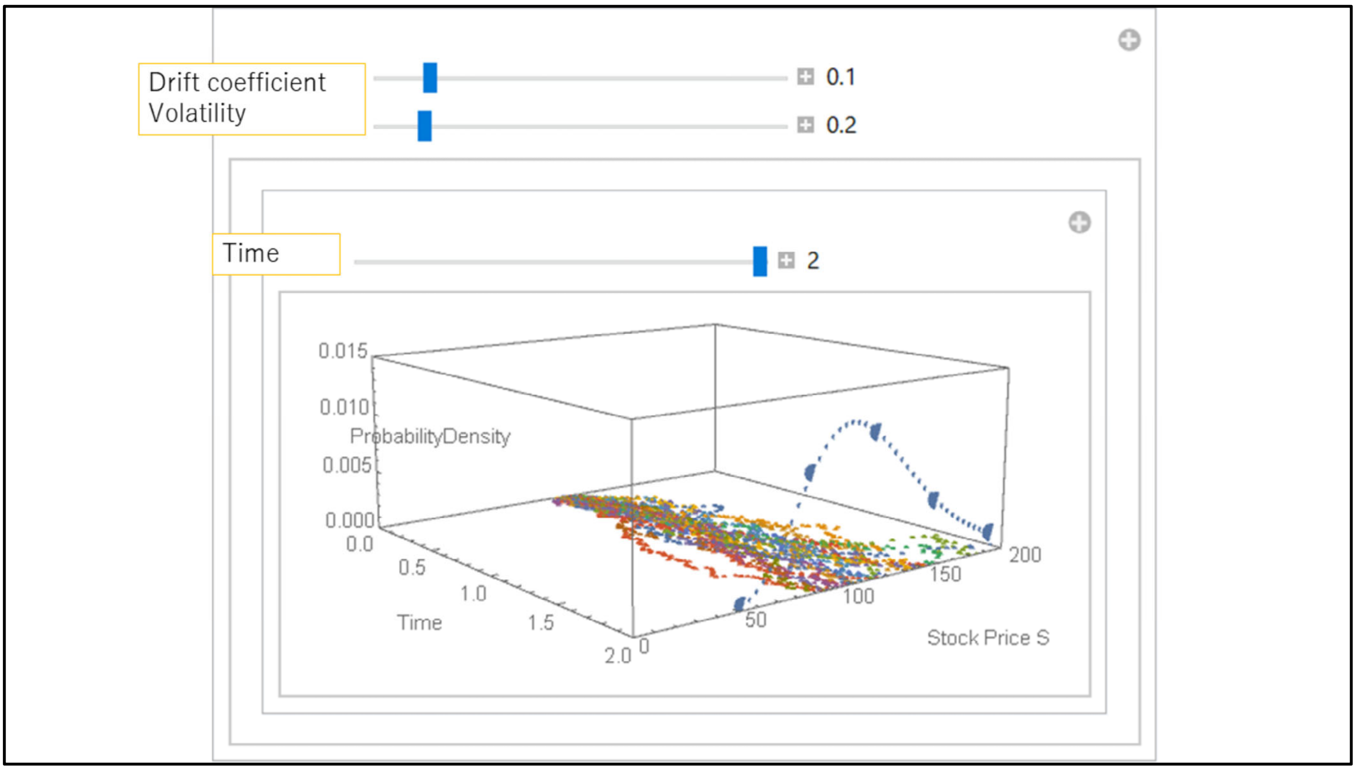




52

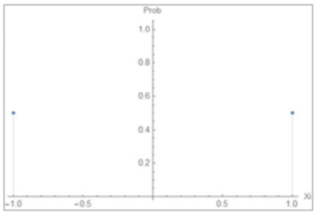
I will use B-of-t for Brownian motion, and S-of-t is the value of the stock, at a given time t.

The movement of the stock price consists of many small steps. The individual steps are not important, but the overall trend is interesting if we want to predict the value of a stock.
The Brown motion follows the normal distribution. On the other hand, stock prices follow a log-normal distribution.



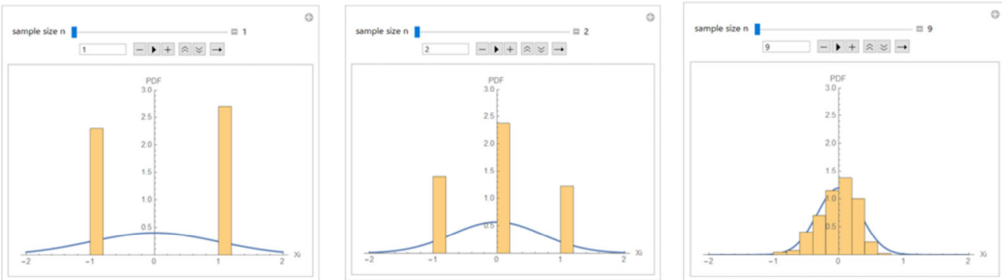
That is a simulation result of stock price movement.
At the maturity T, the distribution follows a log-normal distribution.

The Random Walk of a Drunken Person



Discrete random variable X_i

$P(X_i = \Delta r) = \frac{1}{2}, \quad P(X_i = -\Delta r) = \frac{1}{2}$



$$E(X_i) = 0, \qquad Var(X_i) = \frac{1}{2}((\Delta r)^2 + (-\Delta r)^2) = (\Delta r)^2$$

$$B_n = (X_1 + X_2 + \cdots X_n)$$

B_n *will have the normal distribution* $N(0, \ (\Delta r)^2 n)$

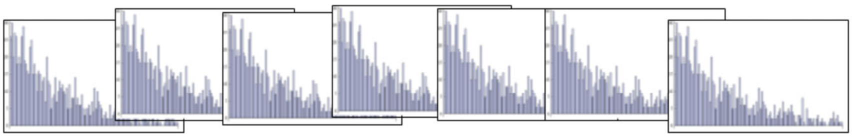
54

Suppose that a discrete random variable has value Xi at a given time.
The total time is divided into n equal time intervals, each Δt long.
During Δt, our drunken person moves with the probability of plus-delta-r 0.5 and the probability of minus-delta-r 0.5.
And again, if we repeat the experiment many times, the PDF probability graph approaches a normal distribution.
The Central Limit Theorem tells us that these values will be normally distributed.

Central Limit Theorem (CLT)

- Let $X_1, X_2, \cdots$ be an infinite sequence of independent random variables with identical distributions (each Xi has mean μ and variance σ^2)

independent and identically distributed (**i.i.d.**) random variables



- Then let $\bar{x} = \frac{\{X_1 + X_2 + \cdots + X_n\}}{n}$.
- Then $E(\bar{x}) = \mu$ and $Var(\bar{x}) = \sigma^2/n$.
- The central limit theorem says that, in the limit as n goes to infinity, $\bar{x}$ has a normal distribution.

55

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma B(t)$$

S(t) is the stock price
B(t) is Brownian motion

μ is the drift coefficient
σ is the volatility

56

We will model a very small change, or **perturbation**, in the stock price, as **Brownian motion**.

It combines with our two parameters, volatility and drift coefficient, in this differential equation for the stock price.

Using Ito's Lemma

$$S(t) = S_0 \times e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B(t)}$$

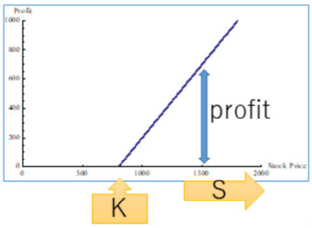
57

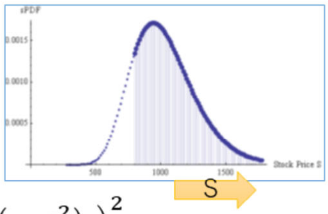
Using Ito's Lemma we can solve the equation. The solution is this stock price equation.

Call option: right to buy the stock at a fixed price, the strike price

$$E[X] = \int_{-\infty}^{\infty} x \times PDF(x)\, dx$$

$$E[S-K] = \int_{-\infty}^{\infty} (S-K) \times PDF(S)\, dS$$





$$= \int_K^{\infty} (S-K) \times \frac{e^{-\frac{\left(\ln\left[\frac{S}{S_0}\right]-\left(\mu-\frac{\sigma^2}{2}\right)t\right)^2}{2\sigma^2 t}}}{\sqrt{2\pi}\sigma\sqrt{t}} \times \frac{1}{S_0} \quad dS$$

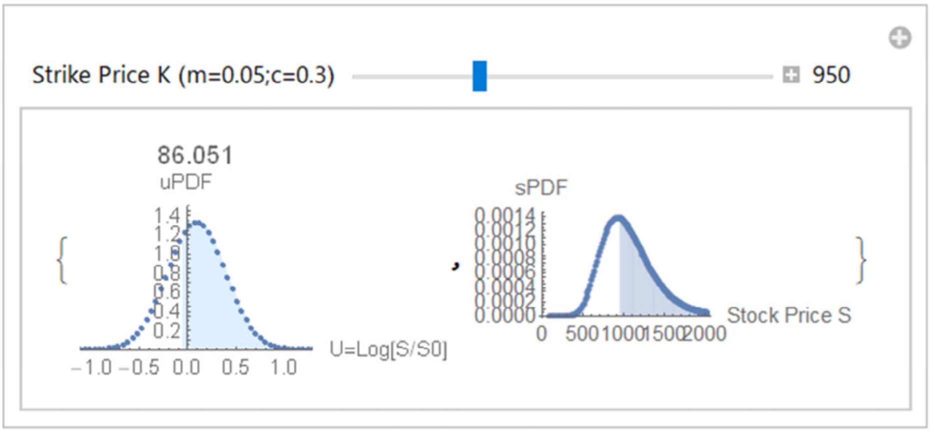
58

If a variable x has a probability distribution density function PDF, then the expected value is given by this expression.
In the case of the profit on a call option, (S-K) times PDF(S) is the expression.

While the stock price is less than the strike price, the profit is zero. At prices above the strike price, the profit goes up.

So we only need to integrate across the range from K up to infinity
The PDF of the profit is the same shape as the PDF of the stock price.

https://www-cc.gakushuin.ac.jp/~20010570/mathABC/SELECTED/BS-CDF/BlackS.cdf



Let's have a look at the right figure. That is the stock price PDF.
When the strike price increases, the expected value decreases because the integration range decreases.
On the other hand, the x axis is variable u, log of (S/S0). The distribution follows a normal distribution. Integration by variable u can get to the same correct answer. You can use variable S or u to get the correct answer. But I think that using S is easy to understand the math.

Conclusions

- Visualization is useful to understand the mathematics !