# Creating a New Hospital Costing System

## - Interacting with Complex Systems -

Edited by Takashi Tahara

Takashi Tahara, Kouji Hirai, Yutaka Tachimori

This text is an English translation of Part III of the above book.

# Part III: Complex Systems in Healthcare

The term "medical uncertainty" has been used in the field of healthcare economics for a long time. This uncertainty has a significant impact on the management of healthcare, including healthcare costs.

If you are in the medical field, uncertainty is an everyday occurrence. For example, it is impossible in medicine to determine one's domain (support area). A company or store's domain is defined outside the medical field. McDonald's only deals in hamburgers. Even if a customer dislikes hamburgers, McDonald's will not serve anything other than hamburgers. However, it is impossible in medical care if a doctor wants to treat only appendicitis. Even if he operates on a patient with appendicitis, it could be an ovarian axis torsion, it could be cancer, or it could be diverticulitis. You may receive an ophthalmology patient by mistake, even if you are a neurosurgeon. If doctors only had to deal with specialized areas, it would be easier to provide medical care, but this is not the case.

It is not only physicians. There are often situations in medical care where the scope of patients cannot be limited. In Japan, emergency medical care is divided into primary care for patients with minor illnesses, secondary care for those requiring hospitalization, and tertiary care for advanced life-saving medical care. However, it is not uncommon for patients with minor ailments to visit tertiary care institutions, and it is not unusual for patients with myocardial infarction to visit primary care institutions. This is not only because patients are seen by mistake. Even if triage is performed in advance by the medical staff, mistakes are bound to occur. Medical care would be much simpler if only minor illnesses were seen at primary care facilities and only life-threatening patients were seen at tertiary care facilities.

The length of hospital stay has become an issue. However, as we experience in our daily clinical practice, the size of treatment cannot be limited: a patient admitted to the hospital expecting to be cured in a week may take three months to recover. Accidents sometimes occur in which a C-section results in massive bleeding and death. It is an unfortunate accident. If the blood transfusion had been completed in time, the patient's life might have been saved. You may think that it is as easy as deciding how much blood to prepare for transfusion during surgery. Many hospitals have been trying to determine the amount of blood to be ready for transfusion for a long time. However, deciding how much blood to be prepared is still impossible. Assuming that sufficient blood amounts are provided for all surgeries, most of the blood prepared would be discarded without being used.

Convenience stores in town also determine appropriate inventory quantity by analyzing databases of products sold. In healthcare, however, proper inventory quantity still cannot be determined. You remove an item from your inventory one day because it will not be used for months; the next day, you need it. This situation often occurs. Is medicine lagging? The limitation of work, control of inventory, and appropriate cost management, which other industries can do, are not so well done in healthcare.

As described above, medical uncertainty is commonplace in the field. Traditionally, there has been a widespread belief that science is unquestionable. This has led some to believe that medical uncertainty arises because medicine is not a science. However, the idea that "science is certain" collapsed with the development of chaos theory in the mid-20th century. In this section, We will show that uncertainty in medicine stems from the nature of the complex systems inherent in medicine.

# Chapter 1: Linear, Nonlinear, and Dynamical Systems, then Chaotic and Complex Systems

**Y. Tachimori**

## 1 Linear, nonlinear

### 1-1 Reductionism

Descartes' reductionism, which laid the philosophical foundation for modern science, is as follows.

1) To accept as accurate only that which is clear and known.

2) The problem should be divided into as many detailed parts or elements as possible.

3) To know the whole, start with the simplest part or element and stack to the most complex.

4) At this time, all the parts or elements should be taken up and composed without dropping even one of them.

The main points are twofold. The first is to break it down into minor elements, and the second is to stack the broken-down pieces to understand the whole.

There is a difference between linear and nonlinear in this second stacking method. When stacking, linear theory considers that there is a linear relationship between the elements and stack them. On the other hand, nonlinear theories, such as chaos theory and complex systems theory, stack elements by assuming a nonlinear relationship between the parts. The two views are the same in that they both decompose into components.[1] The idea differs from holism (holistic approach), which denies decomposition.

### 1-2 Linearity

A linear relationship is intuitively in which elements' properties remain unchanged when stacked into a whale. For example, in the case of mass, the overall mass is simply the mass of each component piled (added up). The linear relationship (of a function) is mathematically expressed as follows.

$$f(x+y)=f(x)+f(y) \qquad f(cx)=cf(x) \qquad \text{(III-1-1)}$$

This is the definition[2] of the function $f(x)$ being linear, which, in the first example, can be interpreted as follows. Let $f(x)$ denote the mass of an element $x$. The first equation says that the group of two elements is the sum of their packs. The second equation states that the abundance of $c$ elements *is* $c$ times the mass of that element. If $x$ and $f(x)$ are real numbers, then it is proved that the only function that satisfies such a relation is the following equation, where a is a constant.

$$f(x)=ax \quad . \qquad \text{(III-1-2)}$$

This is the well-known expression for a straight line. In other words, a linear relationship between the variables x and y means that these two variables are in a proportional relationship (i.e., y=ax).

---

[1] In reality, however, most of the linear stacking was done until recently.

[2] For real-to-real functions, you don't need two expressions. One of them is sufficient. For functions from complex numbers to complex numbers or from vectors to vectors, these two defining expressions are necessary.

The two variables can represent anything[3]. If the variable $x$ *is the* amount of change in the length of the spring and $y$ *is* the pulling force of the spring, then this relationship is Hooke's law of springs. When a quantity changes over time, it can also be the relationship between the quantity at one time and the quantity at the next time. For example, consider savings. If the savings in one month is $x$, the amount of savings in the next month is $y$, and the interest in one month is a, then,

$$y= (1+a)\,x.$$

This is also a linear relationship.

Where each month is denoted by $n$, the deposit amount for month $n$ *is* expressed as $x_n$ instead of $x$. Then, y is represented as $x_{n+1}$, and the following equation describes the above relationship.

$$x_{n+1} = (1+a)x_n \quad . \qquad \text{(III-1-3)}$$

If we determine the deposit amount $x_0$ in the first month, we can find the $x_n$ in any month $n$ as a function of $n$.

$$\boldsymbol{x_n = (1+a)^n x_0} \quad . \qquad \text{(III-1-4)}$$

This equation expresses the deposit amount in month $n$ as a function of $n$. Equation (III-1-3) is the relationship governing the change in savings, which is a linear relationship. The obtained equation (III-1-4) is the relationship between time (with month as the unit) and the amount of savings, a so-called exponential and nonlinear function.

Equation (III-1-3) is an equation to know the change of one variable, but it can also be used to understand changes in multiple variables. For example, suppose you have savings in two banks, A and B (interest rates are *a and b*, respectively). Let $x_n$ and $y_n$ be the respective deposit amounts in month $n$. Suppose further that a specific percentage $c$ of B's deposit amount is transferred to A monthly. In this case, the changes in the deposit amount in banks A and B can be expressed as follows.

$$\boldsymbol{\begin{cases} x_{n+1} = (1+a)x_n + cy_n \\ y_{n+1} = (1+b)(1-c)y_n \end{cases}} \qquad \text{(III-1-5)}$$

This equation allows us to determine the changes in deposits at the two banks each month. This is another linear relationship that represents a change in time. We call the two variables x and y linearly interacting in such a case.

In physics and other fields, equations such as (III-1-3) or (III-1-5)[4] are used to investigate changes in motion or state. In physics, it is called the motion equation or simply the equation, but here, it is called the "change-equation"

---

[3] Although $y=ax+b$ *is* not strictly a linear relationship, it is generally considered to be a linear relationship because $y=ax$ if the origin is shifted. While $y=ax$ is a relationship between one variable $x$ *and* one variable $y$, we also call $(x,y)$ and $(x',y')$ a linear relationship when the two pairs of numbers $(x,y)$ and $(x',y')$ are connected by the relationship $\begin{cases} x' = ax + by \\ y' = cx + dy \end{cases}$. This is a vector-to-vector linear relationship. Furthermore, the same is true for more pairs of three numbers and more pairs of four numbers.

[4] These equations are discrete in time, but when time is continuous, the equations are called differential equations.

in the sense that it is an equation that represents change. If that change-equation is linear (mainly when it represents a change in multiple variables), we call those variables linearly interacting.

Thus, the term linear is used in two primary senses. One is the linearity of the function, and the other is the linearity of the change-equation. As mentioned before, the time variation of a linearly interacting variable may not be represented by a linear function[5].

### 1-3 Nonlinearity

Non-linear means “not linear”. A nonlinear function means a function that is not linear, and a nonlinear interaction means an interaction that is not linear. In fact, linear is an extraordinary relationship. There are very few linear relationships or linear interactions around the world. The world is full of nonlinearities.

On the other hand, when we look at what humans do, linearity abounds. The machines humans make comprise many parts; in many cases, these parts interact linearly (this does not mean that each piece moves linearly). When we make a machine, we first make parts that do one thing. Then, those parts are combined to make a whole machine. Even after the parts are assembled into a whole, they still perform the same function. For example, the contents of a radio may look complicated, but it comprises several parts. They are divided into parts that supply power, detect waves, amplify waveforms, and play sound. These parts are combined to make a single radio, but the finished radio is nothing more than a machine that supplies power, detects waves, amplifies the waveform, and plays sound. In other words, it is a linear machine in that the entire radio shows only the function of the sum of its parts (linearity).

Linear analysis is also used when looking at and analyzing social phenomena. Many use linear approximation since dealing directly with nonlinear functions and interactions is very difficult. This method is used everywhere because it is powerful and easy to handle.

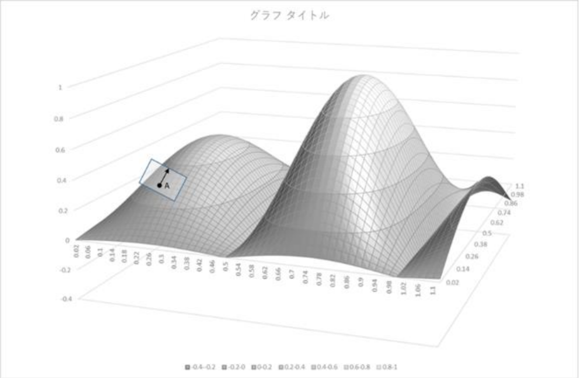


Fig. Ⅲ-1-1　nonlinear function

## 2. linear functions and linear approximation

---

[5] However, the function that represents the change in time of a linearly interacting system is fixed. It can only be a function consisting of a constant, a linear function, a trigonometric function (sin or cos), an exponential function and their sum.

In practice, we use linear approximation to deal with nonlinear phenomena. A linear approximation of a function is the idea that the function is approximated by a tangent line (tangent plane or tangent surface) at that point.

Figure III-1-1 shows a tangent plane having contact with the function (surface) at point A. The surface is winding when viewed broadly, but it is close to a plane near point A. The idea is to approximate the surface by the value of the tangent plane at that point instead of the value of the surface. For example, let us assume that this surface represents a company's total sales operating a chain of stores, where the number of stores is $x$, and the store size is $y$. Suppose that the company is currently at point A. At this point, we consider changing the number of stores and the size of one store to increase the total sales in the future. For this kind of problem, we should consider the tangent plane at the current point A and move in the direction where the slope of the tangent plane becomes the largest. However, this method increases the error between the tangent plane and the surface as one moves forward. Therefore, the tangent plane is found again at a slightly advanced point. Then, the next direction is determined by determining the direction again with the new tangent plane.

We said that there is a surface that determines sales, but in practice, it is rare to know what the shape of the surface is. However, the decisive point of the linear approximation method is that it can be performed even if the surface's shape is unknown. In the above example, a small experiment (or survey) is conducted at a particular time. In other words, the number of stores is slightly increased, or the size of the stores is increased somewhat, and how sales change is investigated. After conducting many of these small studies, the shape of the tangent plane is roughly determined.[6] It is possible to draw a tangent plane without knowing the shape of the entire surface.

The PDCA cycle (Plan-Do-Check-Action cycle) is used in production and quality control in business activities. In a sense, this is also a management method based on a linear approximation. We develop a plan for improvement in a business, do it, and evaluate whether the improvement has been obtained. Then, the degree of improvement is considered. The plan is slightly modified according to the evaluation. This is a process of overall improvement through repetition. To reach the top of the function, we construct a tangent plane at that point and move forward to the point where the slope is the greatest. Then, the tangent plane is again built at a slightly advanced stage, and the way forward is found. This process is repeated. This is the essence of the PDCA cycle, a linear approximation method for dealing with nonlinear phenomena.

Linear approximation is also commonly used in the medical field. Comparative studies are conducted to investigate disease causes, which is also a type of linear approximation. One variable (e.g., smoking rate) is varied to see how the incidence of disease or mortality rate changes. Then, if there is a correlation between the change in that variable and the change in incidence, we determine that the variable affects the disease. This is only a local validity study. For example, the health hazards of cigarettes have been mentioned, and banning cigarettes may decrease lung cancer mortality rate. However, it is not known whether the overall mortality rate will decrease. There is also a theory that medical care costs will increase. Similar methods are often used in the public. For example, the economic effect of Hanshin Tigers (Japanese pro baseball team) winning a championship can be considered a linear approximation.

---

[6] Such a method is what is called in statistics a linear regression method. All general multivariate methods are linear approximation methods.

On the other hand, the linear approximation method has several drawbacks. One is the misalignment (error) between the tangent plane and the actual surface. However, this is a minor problem. As in the PDCA cycle described above, linear approximation does not end once. Going further, we can repeat the linear approximation and correct the misalignment. However, in some cases, such as medical care, where linear approximation is an extensive and costly process, it may be challenging to repeat the linear approximation. In such cases, if the fact is established when the misalignment becomes large, there is a danger that the point with the significant misalignment will be followed as it is.

The more significant problem is that the solution obtained by linear approximation is only locally optimal and may need a better global solution. If you are climbing a mountain without a map and following the slope of an oblique face, you may reach a low mountain peak that is not the summit you aim for. Also, when you are at the top or bottom, the tangent plane will be horizontal, making it impossible to find a new solution. Furthermore, the system under consideration may sometimes undergo a catastrophe, a sudden breakdown. These possibilities are the essential problems of linear approximation.

There is an even more severe and critical problem. It is a problem unique to systems that interact with each other in complex ways (complex systems), such as human activity and biology. The problem is that when one moves toward the top of a curved surface, the shape of the surface itself is changed by that activity [1][2]. This problem typically occurs in complex systems like those we will discuss later. It is like the phenomenon of election predictions changing the election results.[7] This can be considered as the problem does not fit the linear approximation (theory breaking).

Thus, the linear approximation method has significant limitations, but at a stage when the broad functional form is not known (and probably will not be in the future), this method cannot be said to be wrong. Furthermore, a better way has yet to be found. However, it is necessary to be aware of the method's limitations as it is applied.

## 3. Dynamical system

Next, we consider the change-equation. In physics and mathematics, a system of equations for determining time changes is called a dynamical system (Appendix III-1). A dynamical system does not only deal with physical phenomena such as the motion of particles but also with social movements, economic movements, changes in the number of living organisms, and so on. These are all treated mathematically in differential equations or regression equations[8]. The term dynamical system is a tribute to mechanics, the first to use such a method.

The solution of the equations of a dynamical system reveals the time variation of the system. A time series graph

---

[7] Such a phenomenon is called an internal observation.

[8] As mentioned before, there are two types of change-equations: differential equations when time is continuous and those that treat time as discrete (1, 2, 3, ...). When time is discrete, we deal with equations of the form $x_{n+1} = f(x_n)$, where time is represented by an integer $n$. Such an equation is the so-called recurrence formula in high school mathematics. It is also called an autoregressive equation. This means that if we try to find $x_{n+1}$, $x_{n+1} = f(x_n) = f(f(x_{n-1})) = f(f(f(x_{n-2}))) = \ldots = f(f(f(f\ldots(f(x_0)))\ldots)$. This is because it goes back autoregressively. Also, in the world of computer programs, such a relationship is called a recursive call or a recursive function.

is obtained by writing the time on the horizontal axis and the values of the variables on the vertical axis. Although it is possible to think of the properties of a dynamical system in terms of a time-series graph, it is easier to think of it in terms of the concept of phase space. Now, for $n$ equations (whose variables are $x_{1, \dots,} x_n$ ), the values of the variables at each point in time are represented by the points ($x_1$, $x_2$, ..., $x_\mathrm{n}$ ) in the n-dimensional space. This is called the phase space of the dynamical system. The dimension $n$ of this space is called the dimension of the dynamical system. This phase space gives a better picture of the properties of the dynamical system than a direct representation of the time series data. For example, if the solution of a dynamical system has $x = \sin{(t)}, y = \cos{(t)}$, then the solution can be expressed in phase space as a circle (Appendix III-2).

The idea of representing the world in terms of such differential or regression equations began with Newton's mechanics, but the goal is to find the time variation of an object. As I mentioned earlier about linear and nonlinear functions, the change-equation is one tool to find the very functions. This kind of thinking has led physics to the center of natural science.

This idea of a change-equation shaped the so-called deterministic worldview. It is as follows. A differential equation represents the motion of all particles. Solving that differential equation will elucidate all future particle motions. This is equivalent to being able to predict the changes in the world entirely.

In the change-equation, the nature of the time variation obtained is very different between linear and nonlinear. Therefore, the use of linear approximations is limited (theory breaking). The solution for the time variation of a linear equation is completely solvable (precisely only for constant coefficients). It is also known that the stable destination of the time variation (often crucial in practice) can be expressed in phase space in two ways: 1) it leaves away to infinity, or 2) it approaches a certain point and becomes fixed.

“Stable” means that it approaches there spontaneously and returns to the original point as soon as it is displaced. The previous fact shows that for a linear equation, the only stable state is either a fixed point or infinity. It is easy to understand if we take a pendulum as an example. Without friction, the pendulum would continue swinging with a constant period and amplitude (periodic motion). However, this periodic motion is unstable. “Unstable” means that the pendulum will move to another position after a slight change. For example, if a small amount of force is applied to this pendulum to change its amplitude, it will continue to swing with the new amplitude. However, if there is even a tiny amount of air friction, the amplitude will gradually decrease and eventually stop. Since there can be no friction in the real world, the far future state of the pendulum is one in which it comes to a standstill. In other words, this stopped condition is stable. Thus, in the linear equation of change, the only stable state is a stopped state, except for the case where the pendulum leaves an infinite distance.

This stationary state appears in many places. Chemical equilibrium, homeostasis in living organisms, and stabilization of states through feedback all resemble the static state of a pendulum. However, the real world is full of change. Linear equations cannot explain such changes.

In nonlinear equations, the situation changes drastically. Again, we consider the final state, excluding transient states. The first stable state found in nonlinear equations is periodic motion. In the case of a pendulum, even if it is in periodic motion, it changes to a motion of a different amplitude as soon as it is shaken slightly. However, the periodic motion found in the nonlinear equation is stable. Even if the motion is somewhat shifted, it immediately returns to its original orbit. In this sense, the motion is quite different from a pendulum's. Here, a stable and changing final state (steady state) is found for the first time. This is called the “limit cycle”.

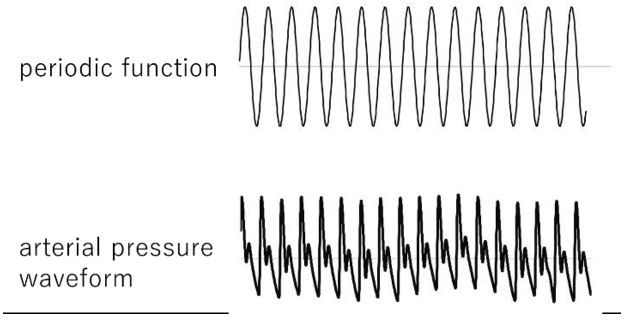


Fig. III-1-2　periodic function and arterial pressure waveform

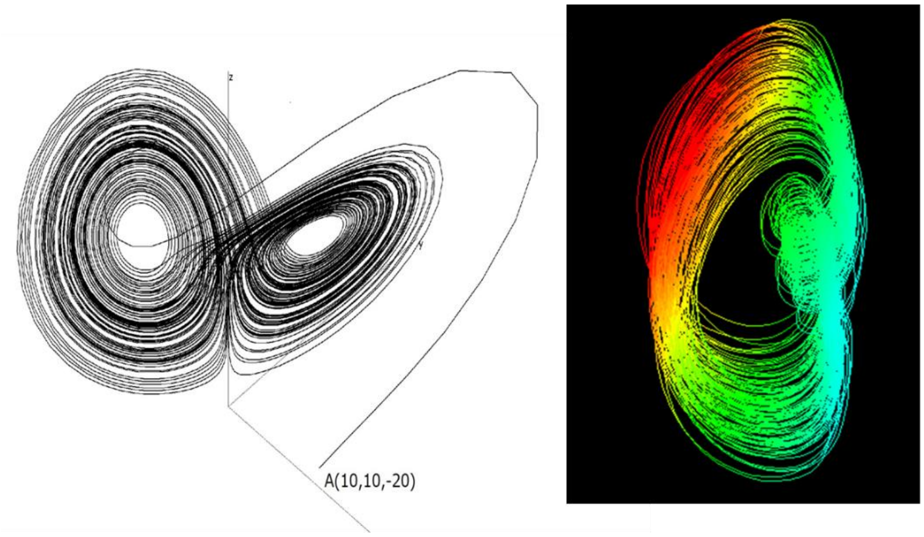
Many phenomena in the world appear to be periodic. The heart's movement, respiratory motion, etc., appear at first glance to be periodic motions. For this reason, it was once thought that these motions were limit cycles. However, a closer look at either the heart or respiratory motion reveals that they are not cyclic (Figure III-1-2). Movements that appear to be cyclic at first glance but, upon closer inspection, are different from cycles. The cause of such motion was not easily understood. In such a situation, a state called chaos was discovered.

Chaos is a stable state found in the solution of a nonlinear equation (differential equation or discrete dynamical system) with highly complex behavior [3]. A steady state does not mean it stays at a specific value. It means it is moving in a very complex way within a specific range, but if you pull it away from that state a little bit, it will return to its original state.[9]

Such a state that keeps moving is not periodic and returns, when you pull it apart, is called a "strange attractor" (Figure III-1-3). This means a strange state as if attracted to that state (Appendix III-2). At first glance, such a chaotic state appears to be a periodic motion. However, a closer look reveals that the motion differs from periodic motion. These motions are like those of the heart, respiration, brain waves, etc., and it is said that they may be chaotic motions.

---

[9] “A little” means that if you move it a lot, it may not come back. In nonlinear equations, there is often more than one stable state, and if you move it too far, it will be pulled into another stable state.



Lorenz attractor Attractor of finger plethysmogram

Fig. Ⅲ-1-3 Strange Attractor.

## 4. from chaos to complex systems

The characteristics of chaotic motion appearing in nonlinear equations (differential equations or discrete dynamical systems) are as follows.

(1) The time series (trajectory) appears random at first glance.

From simple equations, seemingly random and complex motions occur.

(2) Initial value sensitivities: any small initial error magnifies greatly.

Therefore, the future cannot be predicted. In chaos, the magnitude of the error increases exponentially with time. That is, the error is proportional to $e^{\lambda t}$. The average of the coefficients $\lambda$ of this exponential function is called the Lyapunov exponent. This value is a significant number that characterizes chaos. This fact that small changes are greatly magnified is called the butterfly effect. This term is a metaphor for a butterfly flapping its wings in Japan, which may cause a storm in the United States.

Because of the above properties of chaos, it isn't very sensible to carefully follow the changing trajectories. To observe and treat the state, some statistical treatment is necessary. Moreover, it is possible to keep a statistical structure that differs from the normal distribution. It is also seen that there is a relationship between chaos and fractals since so-called fractal structures can be found in the geometry of strange attractors[4][5][10].

The following differential equations of three variables describe Lorenz chaos, a typical chaos.

---

[10] For more information on fractals, please refer to references [4][5].

$$\begin{cases} \frac{dx}{dt} = -px + py \\ \frac{dy}{dt} = -xz + rx - y \\ \frac{dz}{dt} = xy - bz \end{cases} \quad . \qquad \text{(III-1-6)}$$

The nonlinear differential equations that produce chaos must contain at least three variables. It has been proven that attractors created from differential equations with two variables are limited to a single point or limit cycle (Appendix III-2). In discrete dynamical systems, chaos can also emerge from just one equation. The most straightforward such equation is called the logistic map and is described by the equation $x_{n+1} = ax_n(1 - x_n)$.

In general, chaos is described by three or four equations (the number of variables in the equations is called the degree of freedom), but how does a system with a large degree of freedom behave? A gas comes to mind as a system with significant degrees of freedom. The number of particles in a gas is in the order of $10^{23}$. This means that the degree of freedom is in the order of $10^{23}$. The behavior of a system with such a large degree of freedom is quite different from that of a system with a small degree of freedom. In the case of a gas, individual particles' motion characteristics are lost. The global behavior of the gas as a whole can be expressed in terms of only three variables: pressure (P), volume (V), and temperature (T), which are defined as the famous gas equation $PV = nRT$. There, the individuality of particles disappears, and the motion of each particle is described only statistically. That is, the kinetic energy of each particle follows a normal distribution, and the individual particles' motion is neglected. This is because the interactions between individual particles are small or almost negligible.

Then, even for systems with significant degrees of freedom governed by nonlinear equations, is it possible to lose the individuality of each variable and become describable in some macroscopic small number of variables? Tsuda, Kaneko [2], and others have investigated this question by taking an example of a simple system with a large degree of freedom. They first studied the motion of a discrete dynamical system with a large degree of freedom since a system of differential equations would be computationally expensive. Specifically, based on $x_{n+1} = 1 - ax_n^2$, which generates chaos with one degree of freedom, this is extended to N degree of freedom. Depending on the extension, the system is divided into CML (Coupled Map Lattice) and GCM (Globally Coupled Map). The equations are as follows.

Let $f(x) = 1 - ax^2$ and the time is *n,* and the variable number *is i*. Then, each system is represented by the following *N* equations.

CML: (i=1, …, N) $x_{n+1}(i) = (1-\varepsilon)f\big(x_n(i)\big) + \frac{\varepsilon}{2}\big(f\big(x_n(i+1)\big) + f\big(x_n(i-1)\big)\big)$ *i=1, …, N* (III-1-7)

GML: (i=1, …, N) $x_{n+1}(i) = (1-\varepsilon)f\big(x_n(i)\big) + \frac{\varepsilon}{N}\sum_{j=1}^{N} f(x_n(j))$ *i=1, …, N* (III-1-8)

CML represents a local interaction in which the value of each variable is influenced only by both neighbors (i+1, i-1). In contrast, GML represents a global interaction in which the average of the values of all variables affects each variable. The constant ε is a constant that represents the magnitude of the influence. The larger it is, the more significant the impact on the other variables. Global influences such as those in GML are often found in biology, society, etc. For example, in a very simplified model, blood flow in each vessel is determined by the blood pressure applied to that vessel and the vascular resistance. However, the blood pressure in each vessel is affected by the overall blood pressure. The total blood volume and the total vascular resistance determine the overall blood pressure. In other words, the blood flow in an individual vessel meets the requirements of GML because it is affected by the global total blood flow and total vascular resistance, not just the local situation at that site.

Kaneko et al. have simulated a GML consisting of several hundred elements and revealed several important properties [2]. The system is divided into parts (attractors) in which all or part of the elements oscillate synchronously depending on the parameters *a* and ε values. Specifically, it takes one of the following states.

(1) A coherent phase in which almost all initial conditions fall into complete retraction and oscillation.

(2) An ordered phase with a few (k<<N) attractors occupying most of the time. All elements within each attractor are synchronized and have the same time series.

(3) A partially ordered phase in which various attractors with different numbers of clusters coexist depending on the initial conditions.

(4) Asynchronous phase in which each element oscillates entirely discretely.

In the above cases 1)~3), the degrees of freedom decrease, but in case 4), the degrees of freedom remain *N* because all elements move separately. In the first case, 1) and 2), the system is divided into several parts, each moving synchronously. Therefore, once this state is reached, it will remain in it forever. What initially had *N degrees of* freedom is reduced to *k* (usually several). In terms of a phase space, information is being compressed from a graph with several hundred dimensions to a graph at the level of two dimensions (plane) or three dimensions (space). The following case 3) is complex. The part of the elements that move synchronously is the same as in 1) and 2). However, the characteristic of this situation is that each component has a part that moves out of synchronization. The parts that move out of synchronization are not entirely separate from each other, but as time passes, some of the elements move together, and at some time, they are combined with other elements. Tsuda, Kaneko, and others call this phenomenon of coherent parts changing with time "chaotic itinerancy.” In the case of low-dimensional chaos, once an attractor is formed, it remains an attractor forever without changing. However, in the case of multidimensional chaos, a coherent part may appear to be an attractor. Still, it is only temporary, and over time, the attractor may collapse and become completely disjointed, or it may become another attractor. In other words, the attractor state is not stable but metastable.

Such a phenomenon is chaotic itinerancy. During coherent periods over time, the degree of freedom is reduced. However, during asynchronous periods, the degree of freedom is more significant. And it is impossible to predict when the metastable state will break down. The simulation results show that the change from a metastable to an asynchronous state occurs abruptly. This probably corresponds to the fact that the movement of each microscopic element is chaotic, and its trajectory cannot be wholly predicted. Also, in this 3) state, a hierarchical structure of the system is often observed; see reference [2] for details. In the last asynchronous phase 4), even though chaos is present, the individual elements are moving separately, so no reduction in degrees of freedom occurs.

As discussed below, much chaos has been found in living organisms. In many cases, these dimensions are only a few dimensions, and no cases of hundreds of dimensions have been found. Most living organisms are dynamical systems with more than a few hundred million dimensions. The fact that the dimensions of chaos found in such systems are only a few dimensions suggests that a reduction of the dimensions found in GCM research is occurring. In a sense, such a dimension reduction can be regarded as the appearance of an order in the system due to the interaction between the elements.

# Chapter 2: Complex Systems and Power-law Distribution

**Y. Tachimori**

## 1 What is power-law distribution?

Chaos theory was found in the theory of dynamical systems. At that time, a dynamical system's degree of freedom (which can be thought of as the number of equations) was about 3 or 4. Few degrees of freedom means that the system is simple. Then, the question arises as to what happens in a system with many degrees of freedom and a more significant number of elements. The answer to this question is the theory of dynamical systems consisting of many elements, as explained in Chapter 1-4. A system consisting of many elements is naturally a system with many degrees of freedom, and CML and GCM are examples of systems consisting of many elements. What we have learned from the studies of CML and GCM is that a situation like that of chaos theory with small degrees of freedom can be created in such a system. The entire system can be in periodic motion or a chaotic state with few degrees of freedom. Moreover, such a system is characterized by many such states in one system. Therefore, the system settles into different states depending on the initial conditions. More importantly, although the system that has settled down appears stable at first glance, it sometimes suddenly collapses and shifts to another state. The system has many metastable states, and the system transitions between these metastable states over time. This is the phenomenon seen in dynamical systems with many degrees of freedom.

Another important aspect that emerged from chaos theory is how systems are described. In the original Newtonian mechanics, a system is defined by the positions of the particles and their changes in time. Newtonian mechanics is a theory that predicts, and even wholly predicts, changes in place, so-called determinism. However, chaos theory proves that it is impossible to predict the position entirely. This means that it is impossible to describe a system by determining the position of its particles at a particular time. In that case, how should the system be defined? According to chaos theory, the position cannot be determined entirely, but the probability of the position of an element can be determined. This means that the way to describe the system is to change from the position at a particular time to the probability of the position. This is also true for CML and GCM. There is no point in describing the exact values of the elements of the system, which means that we can only predict the probability distribution of the element values and, in some cases, the probability distribution of the index representing the entire system.

Another critical point is the hierarchical structure: studies of CML and GCM have shown that hierarchical structures can be observed in metastable states. It is a matter of course that hierarchical structures are kept outside of dynamical systems, but they must also be observed in dynamical systems. In the world of physics, there is a world of statistical mechanics, which consists of many particles, in contrast to the dynamics of a few particles. The world of statistical mechanics is a world of probability compared to that of mass dynamics. The way of describing this world is different from that of the microscopic world, and it is characterized by macroscopic indices such as temperature, pressure, entropy, etc., which represent statistical averages. Former attempts to explain the laws of the visible world using microscopic dynamics have ultimately failed. There is a hierarchy in the microscopic world, and the macroscopic world, and the logic of the lower hierarchy cannot explain the theory of the upper hierarchy. This means it is impossible to describe the visible world by Newtonian or quantum mechanics. The macroscopic world requires macroscopic logic[2].

The dynamical system cannot represent some worlds and objects in the higher hierarchy of the world that the dynamical system can represent. In the world of living organisms, and especially in the world of populations of living organisms, there are few objects that the dynamical system can represent. The change in the number of individuals, the spread of disease infection, etc., can be expressed to some extent by the dynamical system. However, a dynamical system can't describe the behavior of individuals, the social behavior of a group of individuals, much less a human society or a society created by humans. Then, are there no general laws in such a world? This question gave rise to the concept of complex systems.

Complex systems must be better defined, and different people have different definitions. However, the following definitions are generally accepted.

1) The system[11] is composed of many elements (agents[12]).
2) Each agent interacts locally with each other.
3) This local interaction results in the emergence of overall properties, states, and behaviors that cannot be predicted from the properties of each agent.
4) These overall properties, state, and behavior affect and change individual agents' behavior.

Here, 3) and 4) are automatically determined by 1) and 2), and such an overall state of the system does not exist from the beginning.

The requirements for a system are 1) and 2), but we do not mean that a complex system satisfies only these requirements. There are cases in which an overall property shown in 3) spontaneously comes into existence from a system composed of 1) and 2). Such a phenomenon is called emergence, which is called a complex system. As for 4), even if an overall property like 3) emerges, it does not mean that the behavior of individual agents will change. The agents only interact locally, which is determined from the beginning. However, when viewed from the outside, their behavior appears as if the properties of the whole influence it.

It has been reported that when a bartering model is simulated, one of the items being bartered acts as if it were money [6]. Although each agent is bartering, they appear to sell over a particular item. The behavior of individual agents can then be explained as money-mediated behavior. This phenomenon can be considered the emergence of money through personal and local interactions.

A little caution must be exercised regarding the local interaction in 2) above. Local in the strict sense means local in both time and space. That is, it means that the properties of one agent at a given time affect only other agents in the immediate vicinity. However, it is still being determined whether overall properties emerge in local interactions in this sense. The GCM includes the global property of the overall average in the interaction. The overall average means that each agent's property affects the whole. For overall properties to emerge in a complex system, the range of influence of each agent may need to be global in space or time. In the case of biological groups, including human societies, the function of memory may make the interaction global.

---

[11] The system is the totality of the object we are now considering. It is determined by the arbitrariness of the observer rather than objectively existing.

[12] An agent does not necessarily have to be thought of as an individual with a human-like will. It can be a tissue, a biological organ, or a concept. It can be anything that can be considered as a single entity by the observer.

The following are some of the characteristics of such complex systems.

1) Breaking down the whole into its elements does not give us an understanding of the whole.
2) Overall properties other than the simple sum of the parts emerge.
3) Cause and effect are not uniquely linked.
4) All past phenomena influence present situations and phenomena.
5) Changes in the system and phenomena cannot be predicted in the long term.

Complex systems are a concept derived from nonlinear mathematics, especially chaotic systems. Therefore, they inherit the properties of nonlinear systems. Of the above, 3), 4), and 5) are derived from the initial value sensitivities of chaotic systems. Furthermore, 2) is the concept of emergence and is essential for complex systems [7][8][9].

The case of gases, discussed in Chapter 1, is a system of many degrees of freedom but with small interactions with each other. The particle-by-particle interaction results in an energy exchange between each particle, and finally, the kinetic energy of the particles becomes a normal distribution around a mean energy. This means energy becomes the temperature, an important parameter representing the state of the gas. The reason that temperature is an important parameter representing the properties of a gas is that the energy distribution is normal. In this case, the property of temperature is not a property that is not expected from the properties of each agent. It is simply the sum of the energies of the individual particles. This is consistent with the idea that to understand the whole, the whole can be broken down into more minor elements, which can then be added together. Thus, a gas system is composed of many factors but not a complex one.

What happens when there is interaction between agents? Consider a game of guessing the two sides of a coin. A coin is tossed, and many people are asked to think the flip or front sides. First, consider that these many people are now standing in a line. Then, we asked each person whether they believed the coin was the flip or front sides. There are two cases in this case: one is to ask without telling how others answered, and the other is to ask after describing how the previous person answered. Presumably, the answers would be a random sequence in the first case. However, the neighbor's answer influences each person's response in the following case, so it may not be a random sequence. Thus, the interaction causes a deviation from random.

It is yet to be determined whether there is a property that holds for complex systems in general. However, a power-law distribution is often found in complex systems as a counterpart to the normal distribution for gases mentioned earlier [9][10]. A power-law distribution is a distribution in which a particular variable's frequency distribution (probability distribution) is proportional to $1/x^a$. The properties of the power-law distribution are described in detail in the following chapters. Still, it is also found in many other complex systems besides medical care, as shown in Table III-2-1 and Figure III-2-1. Recently, in particular, many power-law distributions have been known to appear on the Internet. Power-law distributions on networks are known as "complex networks" and are an essential recent topic [11].

- **Magnitude and frequency of earthquakes Gutenberg-Richter rule**
- **Distribution of Exchange Rate Fluctuation**
- **Size and frequency of avalanches on Sandpile**
- **Population of cities and their ranking**
- **Income Distribution(Pareto's law)**
- **Distribution of the number of extinct species per unit time**
- **Distribution of the number of links on the Web**
- **Distribution of Amazon sales**

**Table Ⅲ-2-1 Power-law distribution in non-medical fields**

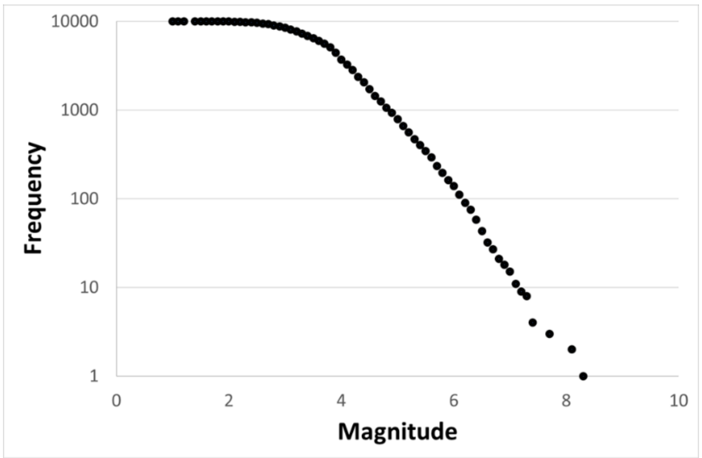


**Fig. Ⅲ-2-1 Earthquake magnitude and frequency. (Complementary cumulative distributions: CCD)**
**Prepared by the author from data in the JMA earthquake database. It can be seen that where the magnitude is 4 or higher, the logarithm of the magnitude and its frequency is a linear relationship. The linear relationship breaks down where the magnitude is less than 4.**

Traditionally, many statistical phenomena have been believed to be normally distributed. In addition, theories were often constructed assuming a normal distribution to explain the phenomena. When a system with a power-law distribution is approximated by a normal distribution, deviations, and leaks occur between the actual value and the distribution. This leakage is more significant than expected, and the prediction is based on the theory assuming normal distribution may break down or the theory construction itself may collapse. Some believe that the recent financial crisis was caused by the failure of financial engineering theories to predict large market fluctuations due to the assumption of a normal distribution of exchange rate fluctuations. In reality, exchange rate fluctuations are said to deviate from the normal distribution and have a distribution similar to the power-law distribution, which can be interpreted as a complex system[10].

The power-law distribution is also called "scale-free distribution" or "fractal distribution" and is closely related to fractals and chaotic systems. Many models have been proposed to explain why the power-law distribution occurs. Some of them are discussed in the following chapters.

## 2 Causes of power-law distribution

### 2-1. Power-law distribution, Long Tail distribution

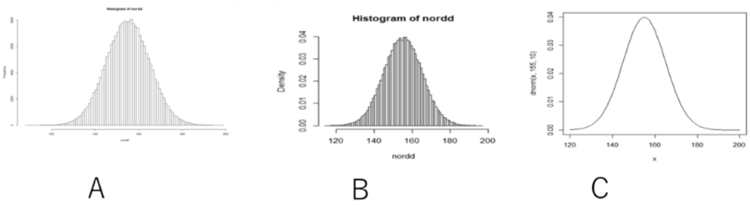


Fig. III-2-2 Histogram and Probability density.

In complex systems, the interaction of elements results in the emergence of overall properties that cannot be predicted from the behavior of the individual components. The most common property is a long tail distribution, such as the power-law distribution. The "Power-law distribution" is found in a wide range of fields, including word frequency distributions in sentences, physical phenomena such as phase transitions, cellular-level phenomena in living organisms, phenomena in biological populations, and phenomena in human society [12] [13] [14] [15] [16] [17].

Consider A in Figure III-2-2 as a histogram of heights from a hypothetical group of 100,000 people (let $N$). Let us assume that the increments are 1 cm. If we randomly select one person from this population, the probability that their height is between 140 cm and 141 cm is $a/N$, where $a$ is the frequency read from 140 cm on the graph. So, to make the chart directly represent probability, all the vertical axes of graph A should be multiplied by $1/N$. This is the graph B. From this graph, we can now read the probability of being a certain height. However, the increments in this chart are every 1 cm. Therefore, reading off the probability of being between 140 cm and 140.5 cm is impossible. In this case, it is necessary to decrease the interval between ticks of the histogram. However, the smaller the gap, the larger the error in the graph. To reduce the error, it is necessary to increase the size of the population being measured. Eventually, the chart becomes a smooth curve as the population's size increases and the interval between ticks decreases. This curve is shown in Graph C.

Such a smooth curve representing probability is called a probability density function (PDF). The probability density function $p(x)$ means that the probability that the value $x$ of the random variable $X$[13] is greater than or equal to $a$ and less than $b$, that is, $a \leq X < b$, is the area from *a to b of* the curve $p(x)$, i.e., Or it means that when $dx$ is a small value, the probability that X is $x \leq X < x+dx$ is $p(x)dx$. Instead of the probability density function, a distribution function (or cumulative distribution function, CDF) may be used. Let $F(x)$ denote the distribution function, which means the area under $x$ in the p(x) graph. In other words

---

[13] A random variable is a mapping from an event to a real number. In the height example, the event represents each individual and the real number represents his or her height. For example, if an individual is represented by $\omega$, $X(\omega)=152$ means that the height of individual $\omega$ *is* 152cm.

$$F(x) = \int_{-\infty}^{x} p(x)dx = \Pr(X \leq x) \ . \qquad \text{(III-2-1)}$$ [14]

In other words, it means the probability that $X$ is less than or equal to $x$. Instead of the cumulative distribution function, the complementary cumulative distribution function (CCDF or CCD) $CF(x)$ may be used. This is defined as follows.

$$CF(x) = \int_{x}^{\infty} p(x)ds = \Pr(X > x) = 1 - F(x) \ . \qquad \text{(III-2-2)}$$

In other words, it means the probability that X will be greater than $x$. When considering probability, you can use PDF, CDF, or CCDF (CCD), but you must know what you use.

A random variable $X$ or $p(x)$ is said to have a power-law distribution when $p(x)$ is proportional to $x^{-\alpha}$ when $x$ *is* above a certain value $x_{\min}$. In other words, $p(x)$ is said to follow a power-law distribution when the following equation represents it.

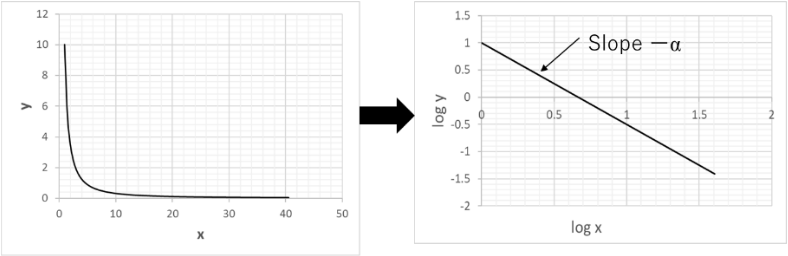
$$p(x) \propto \frac{1}{x^{\alpha}} \qquad (x > x_{min}) \ . \qquad \text{(III-2-3)}$$

The proportionality relation should only hold when $x>x_{\min}$. This indicates that the power-law distribution is a large x (tail) phenomenon.

Here, we call α the "power exponent". Taking the logarithm of both sides of (III-2-3), we obtain

$$\log(p(x)) \propto \log\frac{1}{x^{\alpha}} = -\alpha \log x \ . \qquad \text{(III-2-4)}$$

This shows that a graph of the logarithm of the value $x$ and the logarithm of its probability is a straight line (Figure III-2-3). The same is true for the frequency instead of the probability. From the slope of this line, we can determine α [18][15].



A B

**Fig. III-2-3 Power-law distribution.**
A: power-law distribution diagram B: double-logarithmic graph

Based on actual data, the power-law distribution can be graphically represented by the Zipf plot and CCDF methods in addition to the probability density method. In the Zipf plots and CCDF, both logarithmic graphs of the power-law distribution are straight lines. If the slope of both logarithmic graphs of the Zipf plot for the same data is -ζ and the slope of both logarithmic graphs of the CCDF is -β, the following relationship exists (Appendix III-4).

---

[14] Pr($X$formula) means the probability that $X$satisfies this formula; Pr($X \leq x$) means precisely $Pr\{\omega / X(\omega) \leq x\}$, which means the probability of a set of individuals $\omega$ whose height is less than or equal to $x$. Probability is the ratio of the set of $\omega$ to the whole.
Thereafter, though some integrals appear, $\int_a^b f(x)dx$ is the area bounded by $f(x)$, the vertical lines of $x=a$, $x=b$, and the x-axis in the graph of $f(x)$.

[15] To be more precise[19].

$$\alpha = \beta + 1 = \frac{\zeta+1}{\zeta} \ . \qquad \text{(III-2-5)}$$

From now on, we will refer to the exponentα in the probability density simply as the power-law exponent, the exponent obtained from the slope of the Zipf plot as the Zipfs exponent (or simply ζ), and the exponent obtained from the slope of the CCDF graph as β.

A power-law distribution is essentially a probability distribution. However, there are cases where a power-law connects two variables, *y* and *x*. In other words, it is the case where

$$y \propto \frac{1}{x^{\alpha}} \ . \qquad \text{(III-2-6)}$$

In this case, $y$ and $x$ are said to have a " power-law" relationship. Again, $\alpha$ is called the power-law exponent. The "power-law distribution" can also be described as a distribution in which probability and *x* are related by the "power-law."

The power-law distribution is attractive in complex systems because it is often found in complex systems. Another is that the power-law distribution has properties that differ significantly from normal distributions. The characteristics of the power-law distribution depend on the value of the power-law exponent $\alpha$. Mathematically, there are three points.

(i) 0≤α≤1

If $\alpha$ = 1, then

$$\int_{x_{min}}^{\infty} p(x)\,dx = \int_{x_{min}}^{\infty} \frac{A}{x} dx = [A \log x\,]_{x_{min}}^{\infty} = \infty \ . \qquad \text{(III-2-7)}$$

If $\alpha \leq 1$, then $\frac{1}{x^{\alpha}} \geq \frac{1}{x}$ so the above integral still diverges. After all, *p*(*x*) is not a probability distribution.

② 1<α≤2

In this case, $\int_{x_{min}}^{\infty} p(x)\,dx$ will be a finite value, so *p*(*x*) is a probability distribution.

$$\int_{x_{min}}^{\infty} xp(x)\,dx = \int_{x_{min}}^{\infty} xA/x^{\alpha}\,dx = \int_{x_{min}}^{\infty} A/x^{\alpha-1}\,dx \ . \qquad \text{(III-2-8)}$$

However, the above integral diverges since α-1≤1 in the above equation. In other words, the average no longer exists.

③ 2<α≤3

The mean exists by the same argument as above, but the variance is infinite. This indicates that the variance of the data is enormous.

As described above, the behavior of *p*(*x*) for a power-law distribution varies greatly depending on the magnitude of its power-low exponent. How this affects medical care will be discussed again in Chapter 3.

In social phenomena, including medical care, there are many so-called "long tail distributions" other than the power-law distribution. The lognormal distribution or the truncated power-law distribution are representative of such distributions. There is no significant difference between these distributions in terms of tail. Distributions that are said to be power-law distributions often deviate from the power-law distribution when *x* is large. Therefore, the mean and variance exist even though the distribution is called a power-law distribution. However, since the variance is enormous for both distributions, there will be little difference in application.

### 2-2.Causes of power-law distribution

Many models have been proposed for creating such a power-law distribution [19]. According to Newman, the following mechanisms exist for the power-law distribution.

1) Combination of exponential functions
2) Inverse
3) Random walk model
4) Yule process (rich-get-richer)
5) Phase transition
6) Self-organized criticality
7) Broken stick model

Of the above, 1) and 2) are tricky mechanisms in some respects. The "combination of exponential functions" is a power-law distribution that occurs as follows. Suppose a quantity $y$ has an exponential distribution. That is, there is a relationship $p(y)$~exp($ay$) (where p($y$) is the probability density of $y$). Suppose further that the quantity $x$ has a relationship $x$~exp($by$) with y. Then, the following equation expresses the distribution p(x) of x.

$$p(x) \propto \frac{x^{-1+\frac{a}{b}}}{b} \quad . \qquad \text{(III-2-9)}$$

In other words, $p(x)$ is the power-law distribution with power-law exponent $\alpha = 1 - a/b$.

The "inverse" mechanism of the second is that when the distribution $p(y)$ of $y$ has p(0)>0, given x such that $x=1/y$,

$$p(x) = p(y)\frac{dy}{dx} = -\frac{p(y)}{x^2} \quad , \qquad \text{(III-2-10)}$$

where $y$ *is* close to zero at large $x$, so we end up with $p(y) \cong p(0)$ and so we obtain the following equation.

$$p(x) \cong \frac{p(0)}{x^2} \propto x^{-2} \quad . \qquad \text{(III-2-11)}$$

This can be considered a superficial power-law distribution. Four mechanisms other than these two are essential and will be discussed in more detail.

### 3) Random walk model

The Random Walk Model is a model for the life span of various objects. The random walk model is the following model.

Suppose that the state $x(n)$ at time $n$ can be represented by an integer (..., -2,-1,0,1,2, ...).

In this case, at each time, it changes to +1 with probability $p$ and -1 with (1-$p$).

In other words, it changes as follows.

$$\begin{cases} x(n+1) = x(n) + 1 & \text{probability} \quad p \\ x(n+1) = x(n) - 1 & \text{probability} \quad (1-p) \end{cases} . \tag{III-2-12}$$

Let $x(0) = A$ and consider the first $n$ for which $x(n) = 0$ to be the lifetime $T$ of this system. In this model, when $p$=1/2, the distribution of $T$ is a power-law distribution [20]. Specifically, it is expressed by the following equation.

$$p(T) \propto T^{-1.5} .$$

In other words, it is a power-law distribution with a power-law exponent α=1.5. A slight variation of this model can be used to create a model for the distribution of hospitalization duration (see Chapter 3 for details).

**4) Yule process**

This model is also known as the rich-get-richer model. The phenomena explained by this model include the distribution of populations in cities, the distribution of citations of articles, the distribution of species in genera in biological classification, and the distribution of links on the web. As discussed later, this mechanism may also be related to the distribution of patients with a disease in medical care.

Here is a model of urban population distribution. It is well known that the urban population distribution forms a power-law distribution [8]. A model to explain this is the Yule process. The model is as follows.

For a system consisting of several cities

1. periodically, a specific population of groups is added to the system.
2. the group creates a new city with probability $\pi$ or joins a group already forming a city with probability $1-\pi$.
3. if a group joins an existing city, the probability of joining any city is proportional to the population of each city.

This is all. Figure III-2-4 shows the results of simulating this model.

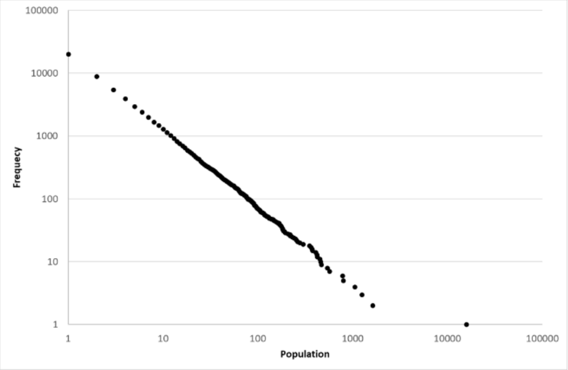


Fig. III-2-4 Simulation results of Population model.

As you can see from the graph, the city's population has a power-law distribution. As mentioned, this simple model is a mechanism for many "power-law distribution" phenomena.

**5) Phase transition**

A phase transition is a physical phenomenon. It represents a situation in which the state of a system changes abruptly. It is a phenomenon in which water becomes ice or gas at a specific temperature. When the phase transition does not occur, the system is said to have an inherent scale (length, size, time) and so on. However, this intrinsic scale is lost at the critical point when a phase transition occurs. At the critical point, all possible lengths and sizes arise, and the length distribution forms a "power-law distribution.

**6) Self-organized criticality**

Phase transitions are unique situations that occur at certain temperatures and environments. Therefore, it is considered a phenomenon that occurs only rarely. On the other hand, self-organized criticality is a phenomenon in which the system automatically creates such a critical state as a property of the system. This phenomenon is a concept discovered by Per Bak in his research on phase transitions [7][21]. The model he created is known as the "sand-pile model." When sand is gradually piled up on the ground, at a certain point, it can no longer be piled up, and an avalanche occurs. This model assumes that the avalanche size distribution at this point forms a power-law distribution. This model explains the relationship between the magnitude and frequency of earthquakes (Gutenberg-Richter law, Figure III-2-1) and the relationship between the size and frequency of wildfires.

In the world of medical care, there is not any phenomenon that this model can explain. However, since it will become important later in explaining the concept of "truncated power-law distribution," I will explain this model here. This model is as follows (Figure III-2-5).

Consider a lattice with one side *N*.

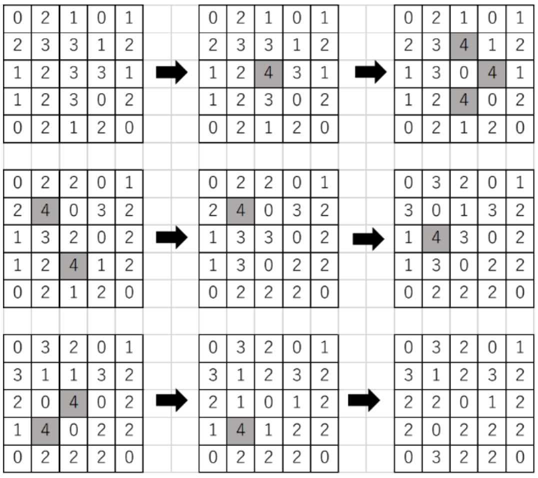
(1) At each point, one block is dropped randomly.

(2) If the number of blocks in the dropped lattice is four due to the drop, the blocks are distributed (avalanche) one by one to the four lattices around that lattice.

(3) If the number of blocks in the distributed grid is 4, perform (2) for that grid.

This operation is repeated until the avalanche stops.

Once the avalanche stops, return to (1).



| | | | | |
|---|---|---|---|---|
| 0 | 2 | 1 | 0 | 1 |
| 2 | 3 | 3 | 1 | 2 |
| 1 | 2 | 3 | 3 | 1 |
| 1 | 2 | 3 | 0 | 2 |
| 0 | 2 | 1 | 2 | 0 |

| | | | | |
|---|---|---|---|---|
| 0 | 2 | 1 | 0 | 1 |
| 2 | 3 | 3 | 1 | 2 |
| 1 | 2 | 4 | 3 | 1 |
| 1 | 2 | 3 | 0 | 2 |
| 0 | 2 | 1 | 2 | 0 |

| | | | | |
|---|---|---|---|---|
| 0 | 2 | 1 | 0 | 1 |
| 2 | 3 | 4 | 1 | 2 |
| 1 | 3 | 0 | 4 | 1 |
| 1 | 2 | 4 | 0 | 2 |
| 0 | 2 | 1 | 2 | 0 |

| | | | | |
|---|---|---|---|---|
| 0 | 2 | 2 | 0 | 1 |
| 2 | 4 | 0 | 3 | 2 |
| 1 | 3 | 2 | 0 | 2 |
| 1 | 2 | 4 | 1 | 2 |
| 0 | 2 | 1 | 2 | 0 |

| | | | | |
|---|---|---|---|---|
| 0 | 2 | 2 | 0 | 1 |
| 2 | 4 | 0 | 3 | 2 |
| 1 | 3 | 3 | 0 | 2 |
| 1 | 3 | 0 | 2 | 2 |
| 0 | 2 | 2 | 2 | 0 |

| | | | | |
|---|---|---|---|---|
| 0 | 3 | 2 | 0 | 1 |
| 3 | 0 | 1 | 3 | 2 |
| 1 | 4 | 3 | 0 | 2 |
| 1 | 3 | 0 | 2 | 2 |
| 0 | 2 | 2 | 2 | 0 |

| | | | | |
|---|---|---|---|---|
| 0 | 3 | 2 | 0 | 1 |
| 3 | 1 | 1 | 3 | 2 |
| 2 | 0 | 4 | 0 | 2 |
| 1 | 4 | 0 | 2 | 2 |
| 0 | 2 | 2 | 2 | 0 |

| | | | | |
|---|---|---|---|---|
| 0 | 3 | 2 | 0 | 1 |
| 3 | 1 | 2 | 3 | 2 |
| 2 | 1 | 0 | 1 | 2 |
| 1 | 4 | 1 | 2 | 2 |
| 0 | 2 | 2 | 2 | 0 |

| | | | | |
|---|---|---|---|---|
| 0 | 3 | 2 | 0 | 1 |
| 3 | 1 | 2 | 3 | 2 |
| 2 | 2 | 0 | 1 | 2 |
| 2 | 0 | 2 | 2 | 2 |
| 0 | 3 | 2 | 2 | 0 |

Fig. III-2-5 Schematic diagram of the sand pile collapse model. Starting from the top-left lattice. A block falls on the middle grid of the upper left grid. Then the avalanche starts in turn and stops at the bottom right. The size of the avalanche is 7.

In this model, the lattice size is considered infinite. The full power-law distribution is obtained only when the lattice size is unlimited. When the lattice size is finite, a "truncated power-law distribution" is accepted, as will be described later.

### 7) Broken stick model

Break a stick of length one at a random place. Break each of the remaining sticks again at a random position. This is repeated. When this process is repeated many times, the lengths of the remaining fragments follow a "power-law distribution." This mechanism is thought to be a mechanism for the distribution of meteorite sizes and the distribution of populations of species living in a particular area[22][23].

## 2-3 Long tail distribution other than power-law distribution

### ◇ Long tail distribution

The power-law distribution is a long-tail distribution. To understand the meaning of "long tail," please refer to Figure III-2-6. Figer A shows the normal distribution and the power-law distribution side by side. At first glance, the power-law distribution appears to be rapidly decreasing, while the normal distribution appears to be slowly falling. However, figure B is the one that is enlarged after x=1000. This figure shows that the power-law distribution decreases more slowly than the normal distribution. This figure shows that the power-law distribution decreases more slowly than the normal distribution when *x is* large (the tail). Such a distribution that decreases gradually at the tail of the distribution is called a "long tail distribution.

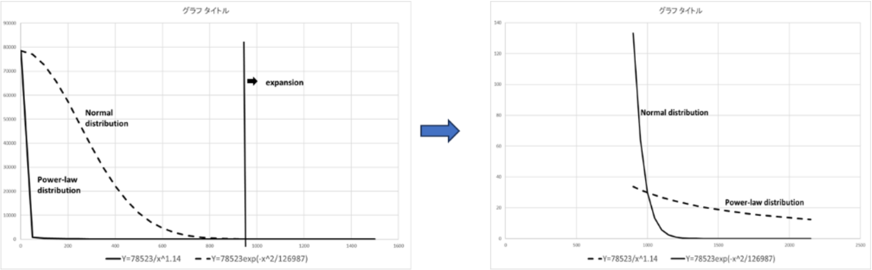


**Fig. III-2-6 Power-law distribution and Normal distribution.**
A：Original figure.
B：Enlarged view of tail in Figure A.

The power-law distribution is not the only distribution with a long tail. Other than the power-law distribution, the log-normal distribution, log-logistic distribution, inverse Gaussian distribution, etc., are known as "long tail distributions." The probability density functions for each are shown below.

**Log-normal distribution**

It means a random variable such that its logarithm has a normal distribution.

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma x}\exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right) \quad . \qquad \text{(III-2-13)}$$

**Log-logistic distribution**

$$p(x) = \frac{\exp\left(-\frac{\ln x-\mu}{\sigma}\right)}{\sigma x\left(1+\exp\left(-\frac{\ln x-\mu}{\sigma}\right)\right)^2} \quad . \qquad \text{(III-2-14)}$$

**Inverse Gaussian distribution**

$$p(x) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right) \quad . \qquad \text{(III-2-15)}$$

These are distributions commonly used in survival time analysis [24][25]. They all have a long tail, but the mean and variance are always present, unlike the power-law distribution. Compared to the power-law distribution, the probability density function seen in the double-logarithmic graph is characterized by a downward shift from a straight line at large *x* (Figure III-2-7).

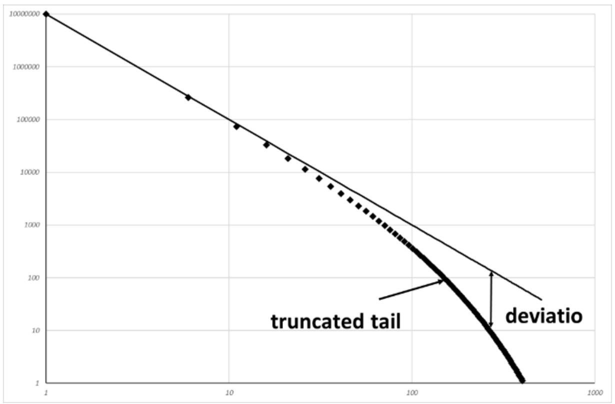


Fig. III-2-7 Power-law distribution and other long tail distribution.

◇ **log-normal distribution**

Among these, the log-normal distribution is typical in complex systems, along with the power-law distribution. Kaneko discusses cell division and evolution problems using a mathematical model of chemical reactions of many substances as an application of the dynamical system consisting of many elements described in Chapter 1 [26]. In it, he states that power-law distributions and log-normal distributions often appear in the results of this model. In general, the change in quantity due to a chemical reaction increases roughly in proportion to the product of the amounts of each component. He states that this results in a log-normal distribution. This is discussed in a little more detail [27].

Suppose that a quantity at time *i is* represented by $X_i$. Let $X_i$ be time-varying according to the following equation.

$$X_i = \alpha_{i-1} X_{i-1} \quad . \qquad \text{(III-2-16)}$$

Where, $\alpha_{i-1}$ is a value (random variable) that varies randomly with time[16] . If the first value of this quantity is $X_0$, then the value at time *n* is

$$X_1 = \alpha_0 X_0 \qquad X_2 = \alpha_1 X_1 = \alpha_1 \alpha_0 X_0 \quad .$$

In the same way

---

[16] Where $\alpha_i$ is a constant, $X_i$ is exponentially increasing or decreasing. However, in intracellular chemical reactions, $\alpha_i$ fluctuates with the surrounding environment and is not considered to be constant.

$$X_n = \alpha_{n-1}\alpha_{n-2} \dots \alpha_0 X_0 \ .$$

Taking the logarithm of both sides of this equation,

$$\ln X_n = \ln X_0 + \ln \alpha_{n-1} + \ln \alpha_{n-2} + \cdots + \ln \alpha_0 \ . \qquad \text{(III-2-17)}$$

$$\ln \frac{X_n}{X_0} = \ln \alpha_{n-1} + \ln \alpha_{n-2} + \cdots + \ln \alpha_0 \ . \qquad \text{(III-2-18)}$$

Here, when *n* is very large, the right-hand side of the above equation has a normal distribution according to the central limit theorem. In other words, let the mean of $\ln \alpha_i$ be $\mu_i$ and the variance be $\sigma_i^2$, and put as follows.

$$m_n = \textstyle\sum_0^{n-1} \mu_k \quad s_n^2 = \sum_0^{n-1} \sigma_k^2 \quad .$$

Furthermore, let $y_n = \ln X_n$ , then this $y_n$ has a normal distribution.

$$p(y_n)dy_n = \frac{1}{s_n\sqrt{2\pi}} \exp\left(-\frac{(y_n-m_n)^2}{2s_n^2}\right) dy_n = \frac{1}{s_n\sqrt{2\pi}} \exp\left(-\frac{(y_n-m_n)^2}{2s_n^2}\right) \frac{dy_n}{dX_n} dX_n$$

$$= \frac{1}{s_n\sqrt{2\pi}} \exp\left(-\frac{(\ln X_n-m_n)^2}{2s_n^2}\right) \frac{1}{X_n} dX_n \ . \qquad \text{(III-2-19)}$$

Eventually, the probability density of $p(X_n)$ of $X_n$ is as follows.

$$p(X_n) = \frac{1}{s_n X_n\sqrt{2\pi}} \exp\left(-\frac{(\ln X_n-m_n)^2}{2s_n^2}\right) \ . \qquad \text{(III-2-20)}$$

Many power-law distributions or long tail distributions are common in medical care. The mechanisms by which they occur are discussed again in Chapter 3.

### ◇ Truncated power-law distribution

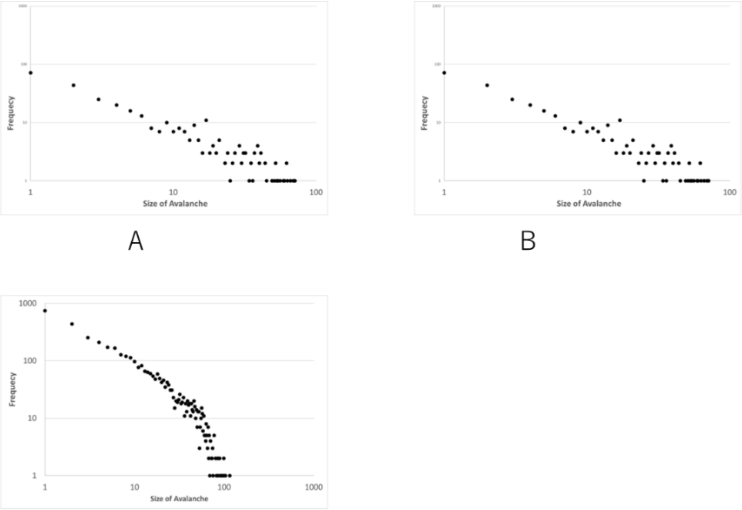
Another type of long tail distribution is generated from the power-law distribution. This type of distribution approaches zero more rapidly than the power-law distribution where *x is* significant (Figure III-2-7). The shape of the distribution is as follows.

$$p(x) \propto x^{-\alpha} F({}^{x}/_{x_0}) \qquad \text{(III-2-21)}$$

where, $F(x) = \begin{cases} 1 & \text{for } x < 1 \\ \exp(-ax) & \text{for } x \geq 1 \end{cases} \ .$

α, $x_0$, and *a* are constants. And *α* corresponds to the exponential of the power-law distribution.

This truncated power-law distribution is created when we place some restrictions on the mechanism that makes the power-law distribution. For example, consider the means of sand-pile collapse. In the sand-pile model, we thought of a lattice with *N* sides. To create a perfect power-law distribution in the sand-pile model, the lattice must have infinite sides. In this sense, this model is restricted from the beginning. Figure III-2-8 illustrates this. The figure shows the results of simulating a sand pile model with a lattice size of 10 x 10. Figures A, B, and C show the results of simulations with 1000, 2000, and 10,000 simulation runs, respectively. The linearity of the frequency distribution is well maintained (i.e., the power-law distribution is carried out) until the number of simulations reaches 2000. However, in (C), where the number of simulations is 10000, the graph is skewed downward at the larger avalanche sizes, and the frequency distribution's linearity is lost. This is because the lattice size of the model is limited to 10, and the results show a truncated power-law distribution.

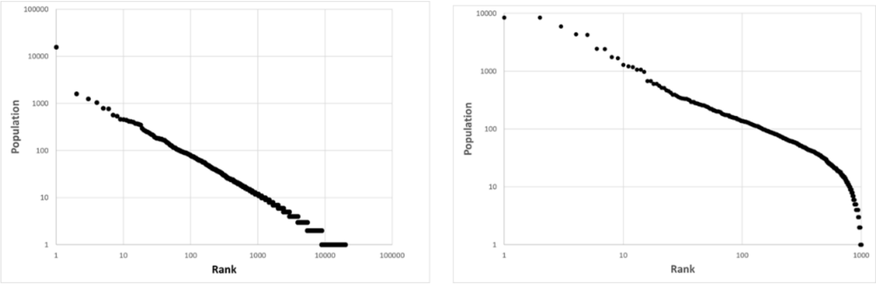


A B

C

**図III-2-8 Simulation results of sand-pile model (lattice 10x10)**
Number of calculations A: 1000 B: 2000 C: 10000

In the urban population model (Figure III-2-4), a truncated power-law distribution is also obtained when the restriction is set up. Figure III-2-9 shows the simulation results of the urban population model. Figure (A) shows the case without restriction, while (B) shows the case where the number of cities is restricted to 1000. In model (B), once the number of cities reaches 1000, new population groups are distributed to existing cities with a probability proportional to their population. As can be seen in the figures, the power-law distribution holds in (A), but in (B), the graph deviates from the power-law distribution downward at the lower ranks. This also indicates that this is a truncated power-law distribution.



A B

**図III-2-9 Simulation results of Urban model(Zipf plot).**
Number of cities A: unlimited B: Up to 1000

Other than this, truncated power-law distributions can be obtained if some restrictions are placed on the mechanism of the power-law distribution. The example in medical care is discussed again in Chapter 3. The models for generating power-law distributions described above are all based on ideal conditions. In the real world, there are almost always limitations due to land area, population, and social issues. Therefore, a perfectly formed power-law distribution rarely exists. Especially in medical care and social phenomena, there will always be limitations. Therefore, such rules should be taken into account when considering actual applications.

# Chapter 3: Complexity in Medical Care

**Y. Tachimori**

## 1. Chaos in living organisms and its analysis

Chaos also occurs in dynamical systems consisting of many degrees of freedom. Moreover, the chaos is determined by a few degrees of freedom. However, the equations that govern them cannot be found. A living body is a complex system consisting of many elements. However, when they interact, some order emerges, which can be described by a few degrees of freedom. If so, chaos may exist in the various signal systems of living organisms. That is how chaos research on living organisms has been conducted [89][17].

Some work needs to be done to determine if time series data is chaotic or to analyze it. Time series data[18] are generally one-dimensional. As mentioned above, some order is created in a dynamical system with many degrees of freedom and can be described with a few degrees of freedom, but even a few are multi-dimensional. To determine whether one-dimensional time series data is chaotic, it is first necessary to reproduce the phase space from the time series data. We do the following to represent one-dimensional data in an m-dimensional phase space. Let $x(k)$ be the original one-dimensional time series data and the delay time $\tau$ is introduced to define the m-dimensional points as follows.

$$u(k) = \big(x(k),\ x(k-\tau),\ x(k-2\tau),\ \dots,\ x(k-(m-1)\tau)\big)\ . \qquad \text{(III-3-1)}$$

This is the point corresponding to the state of the original system concerning the observed data. By reproducing this point in m-dimensional space, the structure of the original dynamical system can be investigated. Takens' "Embedding Theorem" guarantees the correctness of such a method. Please refer to the book [3] for more details on these. By embedding the original data in such an m-dimensional phase space, it is possible not only to reproduce the shape of the attractor but also to calculate the Lyapunov exponent, the dimension of the attractor, etc.[28][29].

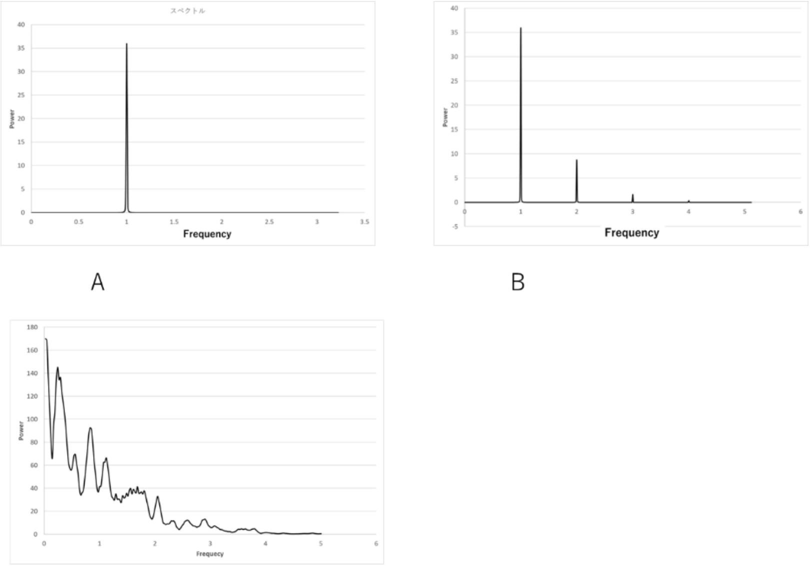
Time series data that exhibit chaos have several characteristics. The first of these is that the power spectrum of the data is continuous. The power spectrum[19] is simply the relative magnitude of the frequency components. In physical terms, it is the energy density of each frequency in electromagnetic and sound waves. For example, a sine wave $\sin(2\pi f t)$ consists of just one wave of frequency $f$. Since the power spectrum of such a wave has only one frequency, it has a value only at the point of frequency $f$ (Figure III-3-1a). However, general sound and electromagnetic waves consist of the sum of waves of many frequencies. For example, the sound of a musical instrument consists of many harmonics in addition to the base frequency. When such waves are expressed as a power spectrum, they consist of the base frequency and its overtone components, so values appear in skipping places (Figure III-3-1b). In contrast, time-series data such as chaos is represented by continuous parts of its power spectrum rather than bouncing around (Figure III-3-1c).

---

[17] Refer to " Chaotic scenarios of complex systems " [2] .

[18] Data changing over time.

[19] The power spectrum is calculated by Fourier transforming time-series data. The Fourier transform is performed by a program, and the most frequently used program is called FFT (Fast Fourier transform), which is used to calculate Figure III-3-1, etc.



A

B

C

Fig. Ⅲ-3-1 power spectra of several data.

Analysis by power spectrum in medical care has been done for EEG and heart rate (HR) for some time [30]. The power spectrum is calculated from time series heart rate data and analyzed by calculating each frequency component. Among the frequency components, the high-frequency band (HF) between 0.15 and 0.40 Hz is said to be strongly influenced by parasympathetic activity. In comparison, the sympathetic nervous system strongly influences the low-frequency band (LF) between 0.01 and 0.15 Hz. This comparison of HF and LF can be used to estimate the state of activity of the autonomic nervous system.

In contrast to this physiological view, Goldberger et al. [31] present an alternative idea. Under normal conditions, HR is not temporally constant but constantly fluctuating. In such a case, the power spectrum of HR shows a continuous spread. However, when the heart condition deteriorates, such as heart failure, HR fluctuation disappears and shows periodic changes or constant values. This was demonstrated by examining the HR of patients with heart failure and sudden death. In the language of chaos theory, this means that HR exhibits chaotic fluctuations when normal and becomes periodic or fixed in a state of crisis. The conventional view is that the body has homeostasis and tries to maintain a constant environment in the body. In contrast, Goldberger stated that fluctuation is normal and fixation or periodic motion is abnormal. This was the first time that he gave a positive meaning to fluctuation. At the same time, it can be considered to have given physiological significance to the phenomenon of chaos.

Another characteristic of time series data that exhibits chaos is that the Lyapunov exponent is positive. One of the characteristics of chaos is initial value sensitivity. Initial value sensitivities are the fact that even though they start from two similar states, they become very different as time passes. In dynamics terms, no matter how close the two initial values are, they move away from each other exponentially as time goes by. In other words, the two orbits move away from each other in the form $\exp(\lambda t)$ $(\lambda > 0)$. The average value of $\lambda$, which determines the rate at which they move away, is called the Lyapunov exponent. If $\lambda$ is negative, the two orbits will move closer together and become virtually identical in the future. However, if $\lambda$ is positive, the two trajectories will move exponentially apart. Therefore, if we can find $\lambda$ from the time series data [28][29][32], we can determine whether the system is chaotic or not.

Tsuda, Tahara, and Iwanaga analyzed the finger blood flow waveforms by finger plethysmogram. They found that the Lyapunov exponent of this time series was positive [33]. A finger plethysmogram is recorded by shining a light on the fingertip to determine changes in the blood flow of the fingertip. It is also used in pulse oximeters that measure blood oxygenation. A waveform corresponding to the changes in blood flow at the fingertip can be obtained by attaching a fingertip pulse meter to the examinee's finger. Tsuda et al. measured the Lyapunov exponent of the finger plethysmogram with an instrument developed by Iwanaga et al. As a result, they proved that the finger plethysmogram is chaotic. Furthermore, the attractor was reproduced by embedding this waveform in a 4-dimensional phase space. Furthermore, Tahara et al.[34][35] found that the Lyapunov exponent and the shape of the attractor varied depending on the subject's mental state and showed that it could be applied clinically.

Since the publication of Tsuda and others, chaos analysis has been performed on many biological signals [111]. Sumita et al. found that changes in pupil diameter were chaotic [36], and chaos analysis was also performed for electroencephalography (EEG) and electrocardiography [37]. Since then, chaos in many living organisms has been discovered, and chaos is a normal phenomenon in living organisms. Furthermore, based on many findings of biological chaos, Tahara et al. view health as "the autonomous formation of relationships through the formation of feedback loops between body and mind, and between individuals and society.” From the viewpoint that "the state of health is somehow expressed on the surface of the body, which is the boundary between the individual and society," they have introduced the idea of "somato-semantic parameters" as an indicator to characterize the state that appears on the surface of the body and are trying to utilize it[38 ][39][40]. He states that most of the chaos in the living body is derived from the measurement and analysis of somato-semantic parameters such as pulse waves, heart sound, electrocardiogram, electroencephalogram, pupil diameter, etc.

Chaos is also seen in engineering and other fields. Still, its manifestation is often viewed as an uncontrollable phenomenon concerning the control of an object, and the problem is how to prevent chaos from occurring in the control. However, in the case of living organisms, chaos itself is thought to have significant advantages for the organism. Traditionally, the concept of homeostasis has been a major factor in living organisms. The concept was regarded as a phenomenon in which various parameters of a living organism are kept constant. However, the importance of stability in a constantly changing environment was recognized with the discovery of chaos in living organisms. Kaneko, Tsuda, Tahara, and others have proposed a new concept of “Homeodynamics” for such phenomena.

Clinical applications of finger plethysmograms are also being made[41][42][43][112][115]. They are often for mental and behavioral problems[113][114][115]. Finger plethysmograms were originally an analysis of human blood flow. The innervation of blood flow is primarily sympathetic and parasympathetic. It is common to assume that the human psyche influences blood flow through these innervations. If this is the case, there should be more studies on the relationship between innervation and finger plethysmograms, but we do not see many of them. Here, we present the results of a preliminary survey conducted when a finger plethysmogram was developed. Since this was a preliminary study and not strictly controlled, the reliability of the data obtained is questionable, but the results are presented for reference.

The stellate ganglion block (SGB) is a technique in the pain clinic area. This is a procedure in which a local anesthetic is injected near the cervical vertebrae in a person's neck. This temporarily paralyzes the stellate ganglion, one of the sympathetic ganglia. The temporary inactivation of the sympathetic ganglion improves blood flow to the face, upper extremities, and other areas innervated by that ganglion. This block is widely used to treat upper body pain, facial Bell's palsy, and idiopathic hearing loss.

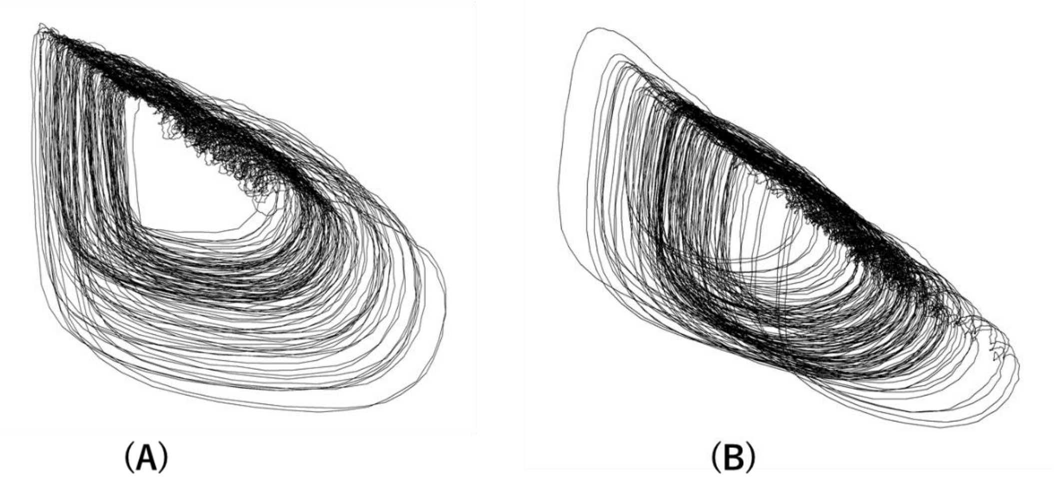


Fig. Ⅲ-3-2　Attractors of finger plethysmogram before and after SGB
(A) before SGB　Lyapunov exponent=0.129
(B) after SGB　　Lyapunov exponent=0.445

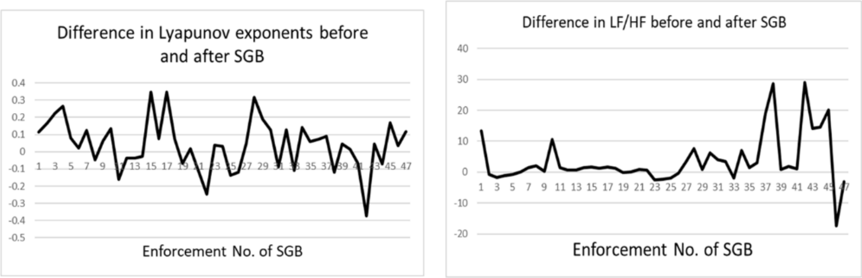
SGB is a block to the sympathetic ganglion, so it naturally affects blood flow to the upper extremity. This would also affect finger plethysmograms. Therefore, we examined the changes in finger plethysmograms before and after SGB. At the same time, Fourier analysis of heart rate was also performed to see the effects on the systemic autonomic nervous system. However, since an electrocardiogram was not recorded, the a-a interval, which correlates strongly with the R-R interval of the electrocardiogram, was measured from the second derivative of the finger plethysmogram waveform and used as a proxy for the R-R interval [44].

Finger plethysmograms were measured before and after SGB, and the Lyapunov exponent was calculated by reconstructing the attractor in 4 dimensions. At the same time, the pre-and post-heart rate (the reciprocal of the a-a interval) was Fourier analyzed to obtain the low-frequency component LH (0.02~0.15 Hz) and the high-frequency component HF (0.15~0.4 Hz) of the spectrum, and their ratio LF/HF was calculated.

Figure III-3-2 illustrates the finger plethysmogram attractors before and after SGB. In this example, it can be seen that the Lyapunov number increased after SGB. Figure III-3-3 (A) shows the result of subtracting the pre-block Lyapunov number from the post-SGB Lyapunov number. As can be seen in the figure, the SGB often increases the Lyapunov exponents of the attractor. The significance test of the difference between the pre-and post-Lyapunov exponents showed p=0.06 (two-tailed), which is not significant but does indicate that the stellate ganglion block tends to increase the number of Lyapunov.

A graph of the difference in LF/HF before and after the block, measured simultaneously, is shown in Figure III-3-3 (B). It can be seen that LF/HF increased after the block ($p$=0.003 two-tailed). It is said that the low-frequency component of heart rate LF mainly represents sympathetic activity, while the high-frequency component HF mainly represents parasympathetic activity. The results of the present study indicate that the block increases the sympathetic activity. The stellate ganglion suppresses local sympathetic ganglion activity. The current results suggest that blocking local sympathetic ganglia increases sympathetic ganglion activity throughout the body. However, interpretation of these results requires consideration of the effects of SGB, the effects of lying in bed, and the effects of needle insertion. In addition, as already mentioned, the data are not very reliable, so it is not yet possible to make a definitive statement. As for the Lyapunov exponent, the conclusion has been reached that the block almost always increases it. The interpretation of this, unfortunately, cannot be fully explained at this point.

The physiological significance of changes in finger plethysmograms is so far unclear. Although dynamical systems theory may be able to predict the occurrence of chaos in living organisms, the physiological or biological significance of this phenomenon has only just begun to be investigated. To investigate the physiological relevance, it is not enough to start with dynamical theory; more physiological and medical science research is needed.



(A) (B)

Fig. Ⅲ-3-3 Differences of Lyapunov exponents and LF/HF before and after SGB

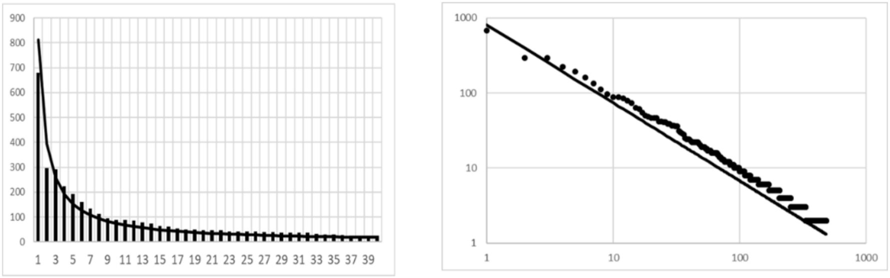
(A) Differences of Lyapunov exponents

(B) Differences of LF/HF

Recently, simultaneous measurements of fetal and maternal heart sounds under various conditions, such as mother's voice stimuli, others' voice stimuli, no stimuli, etc., have also shown that both are deterministic chaos [45][46][47]. This indicates that the heartbeat is both chaotic and altered by external stimuli. There have been few reports on changes associated with the interaction of chaos with the environment, so further development is expected.

### 2 "Power-law distribution" in medical care

Figure III-3-4(A) is a Zipf plot of the frequency of disease names (number of patients with specific disease names) in the ophthalmology department of a hospital. The horizontal axis is the rank of disease names ordered by frequency, and the vertical axis is the frequency of disease names of that rank. The solid curve is the graph of $Y=812/X^{1.04}$. The solid curve graph agrees well with the bar graph. This suggests that $Y=812/X^{1.04}$ represents the frequency of disease names in ophthalmology. As already mentioned in Chapter 2, when the frequency $Y$ of rank $X$ is expressed as $Y = A/X^{\zeta}$ then this data follows a power-law distribution. The relationship between the power-law exponent α and Zipf exponent ζ is $\alpha = \frac{\zeta+1}{\zeta}$. As already mentioned in Chapter 2, the Zipf plot of data following a power-law distribution is a straight line when expressed as a double-logarithmic graph, and this ζ can be obtained from the slope of the straight line. Figure III-3-4(B) is a double-logarithmic graph of the frequency of disease names for ophthalmology. The chart fits the straight line well, indicating that the frequency of ophthalmology diagnoses follows a power-law distribution. The slope of the line shows that the ζ of the ophthalmic diagnoses' distribution is 1.04. Also, the power-law exponent $\alpha = (\zeta+1)/\zeta = 1.96$. As indicated above, the frequency of disease names in ophthalmology almost follows the power-law distribution[20]. The distribution of disease names has been found to follow a power-law distribution, not only in ophthalmology[48][49]. The frequency distribution of disease names will be discussed again in Chapter 4.



(A)

(B)

Fig. III-3-4 Ophthalmology Disease Frequency.
(A) Disease Frequencies
(B) Both logarithmic graph of disease frequency

Figures III-3-5 and III-3-6 show the distribution of length of hospital stay in acute and chronic care hospitals. There is a peak at 2 or 3 days in hospital stay, followed by a sharp decline, but the number of days does not drop to zero, but rather slowly and gradually declines, trailing a long tail. This distribution is also a "long tail distribution. Double logarithmic graphs are slightly more convex than the regression line and are similar to the lognormal and

---

[20] To be precise, the frequency distribution of many disease names deviates slightly from the perfect power-law distribution. To be more exact, it is a "truncated power-law distribution," which is explained in Chapter 2-3.

truncated power-law distributions. The mechanism of its occurrence is discussed in Chapter 3-4. Due to this long tail, many patients have a longer hospital stay than average[21]. Interestingly, although the average length of hospital stays differs greatly between the acute and chronic phases, both show a long tail distribution. The mechanism of an occurrence described below is a "random walk model with drift.” The difference between the acute and chronic phases is thought to be mainly a difference in the magnitude of this drift. The difference in the size of the drift appears in the difference in the average length of hospital stays, and the random walk mechanism may be responsible for the long tail. In any case, the long tail distribution makes it difficult to predict the length of hospital stay, making hospitalization management difficult.

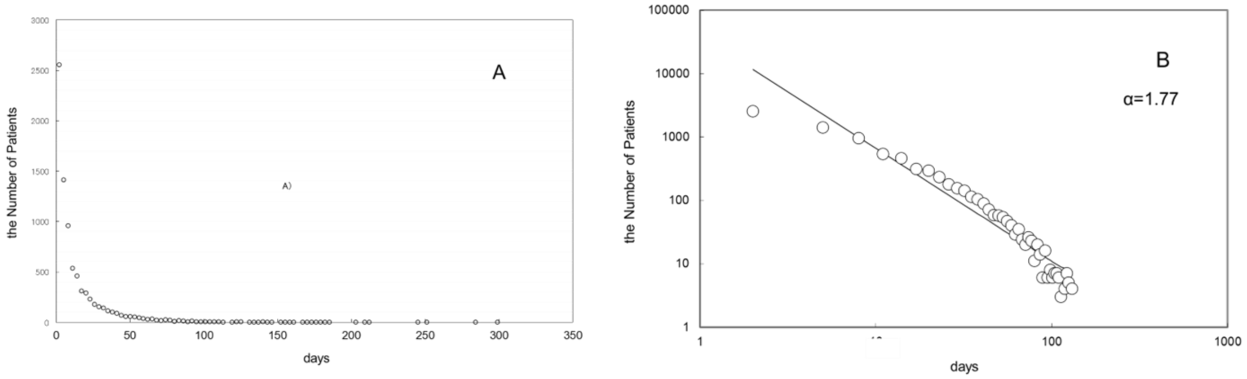
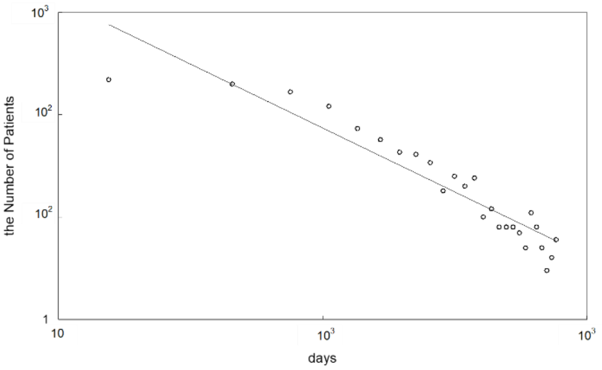


Fig. Ⅲ-3-5 Distribution of hospitalization days in acute care hospitals.
Average hospitalization days=15 days

A Distribution of hospitalization days.

B Double logarithmic graph of hospitalization days distribution.
α=1.77



---

[21] The average length of hospital stay here is the average of the actual length of hospital stay for discharged patients. The calculation method is different from the average length of hospital stay specified by the Ministry of Health, Labor and Welfare.

Fig. Ⅲ-3-6 Distribution of hospitalization days in long-term care hospitals.

α=1.24 Average hospitalization days=148 days

In the diagnosis distribution, the power-law exponent α is around 2 (1 in ζ). In the length of hospital stay, the power-law exponent is 1.77 for acute care hospitals and 1.24 for convalescent care hospitals, as shown in the figure. The fact that the power-law exponent is less than 2 has significant implications, which will be discussed in the following section.

From the author's research, power-law distributions or "long tail distributions" are not uncommon in medical care. Table III-3-1 lists events in medical care that exhibit long tail distribution. The events marked with a "?" in the table are only estimates due to insufficient data. However, given the occasional occurrence of significant bleeding during surgery and the failure of many hospitals to control operating time, it is reasonable to speculate that these distributions are "power-law distributions" or "long tail distributions.

- Frequency of disease
- Frequency of surgery
- Hospitalization days
- Intraoperative blood loss ?
- surgery time ?
- Medical bills ?
  probably
- medical error ?

Table Ⅲ-3-1 Power-law distribution in medical care

### 3 Impact of long tail distribution on medical care

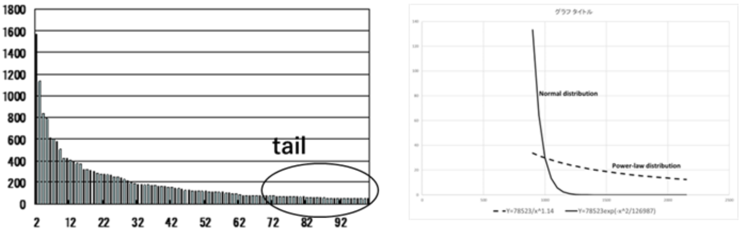
### 3-1 Abnormal values are not rare.

Long tail distributions have extraordinary characteristics for those accustomed to normal distributions. I will explain these characteristics in two parts: 1) outliers are not rare, and 2) the mean is not helpful. The following explanation is given as a pure "power-law distribution. As for the "long tail distribution" in general, please read words such as infinite and "huge value" to understand.

First, outliers are not rare or can be described as "rare but often occur." Let us discuss this using the example of the frequency distribution of disease names.

Figure III-3-7 (A) shows the frequency distribution of disease names in a hospital. The outliers are those in the tail marked with ○ in this figure. The very long tail is a characteristic of the power-law distribution. Figure III-3-7 (B) shows the power-law distribution superimposed on the normal distribution. The normal distribution shows only the right half of the symmetrical distribution. As can be seen from the figure, "the power-law distribution decreases rapidly in the small part of *x* (abscissa) but decreases more slowly than the normal distribution as *x* increases.

Therefore, positive values persist even in the tail, which would be almost zero in a normal distribution.



(A) (B)

Fig. Ⅲ-3-7　Power-low distribution and normal distribution

Let us consider the impact of this using the example of disease naming. Figure III-3-8 shows the frequency distribution of disease names in Figure III-3-7 (A), rewritten as cumulative percentages. That is, the value of the graph at rank *x* is the sum of all the frequencies of ranks from 1 to *x* in Figure III-3-7 (A). In other words, the value at rank *x* represents the total number of patients who can be named using all the disease names from rank 1 to *x* (but expressed as a percentage of the total number of patients in the graph).

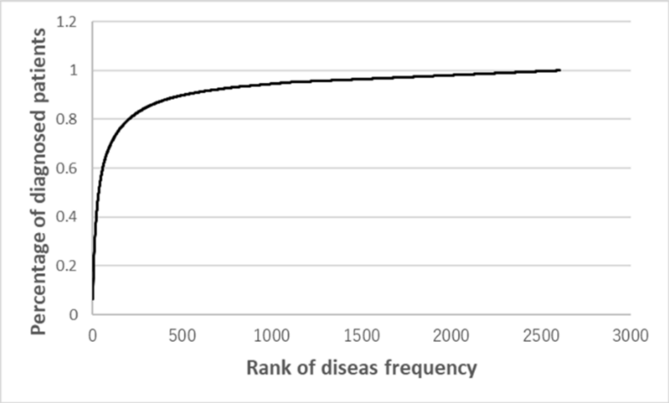


Fig. Ⅲ-3-8　How many disease names are needed ?

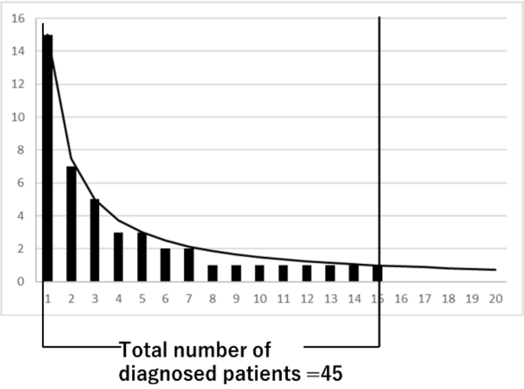
As seen in this figure, about 100 disease names can be used to label about 70% of all patients (tens of thousands of patients in this data). When looking at this figure, one might think, "It would only take a few hundred disease names to label everyone. However, the graph shows that this expectation is not valid. With 500 disease names, 90% of patients can be labeled. At 1,000, it is 95%, and after that, the increase in the graph becomes slower and slower. In the end, with this data, it would take thousands of disease names to label all patients. Contrary to initial expectations, it would not be easy to label everyone.

There is an ordering system in the hospital order entry system or electronic medical record called a disease name order. This is an order in which the physician enters the name of the patient's disease into the computer each time they examine the patient. The disease names are then entered into the medical record system and used for receipts and other purposes. If disease names are entered in free form, it is inconvenient to use the data later, so storing them in disease name codes is necessary. For this reason, in the past, a disease name master was created for each hospital.[22] When creating this master, the following procedure is used. First, prepare several disease names (thousands of them). At that point, when the system starts to operate, some disease names will be missing, so each year, those disease names will be added to the master. This process is repeated every year. The expectation is that if this process is repeated every year, the number of disease names to be added will decrease, and after a few years, there will be little need to add new disease names. However, what happened was as follows.

1) The exact number of disease names must be added to the disease master each year.

2) Many of the added disease names are never used again.

This phenomenon can be explained if the frequency distribution of disease names is a power-law distribution[50][51]. Figure III-3-9 shows a graph of the power-law distribution, Y=*A*/*X* (*A*=15 in the figure). The chart, *A/X,* generally has a decimal when *X is* varied. The actual frequency of disease names is an integer without a decimal. Therefore, as shown in the graph, the precise frequency is an integer obtained by truncating the decimal point of the value of *A*/*X*. Also, since the frequency must be greater than or equal to 1, the range *A*/*X*<1 is meaningless. Ultimately, this graph only makes sense in the range of *X* from 1 to 15. Adding up the frequencies of disease names in this range yields 45. This means that if the frequency of disease names follows the graph of *15/X,* we can label 45 patients with 15 different disease names. In the same way, if we set *A* to 30, we can label 111 patients with 30 different disease names. In other words, 40 disease names are needed to label 111 patients. Thus, the frequency of disease name distribution can be used to estimate the number of names needed to label patients. We then examined how the number of required disease names changes when *A* is larger and larger.



Total number of diagnosed patients =45

Fig. III-3-9 Calculation of the number of disease names needed

---

[22] Currently, the organization MEDS-DC provides a standard disease name master. Therefore, the discussion here may not already apply. We would like the discussion below to be read as an increase in new disease names within the ordering process. However, while this may not happen in each hospital, the same problem may occur when the standard disease name master is created, and it may be necessary to continue updating the standard disease name master indefinitely.

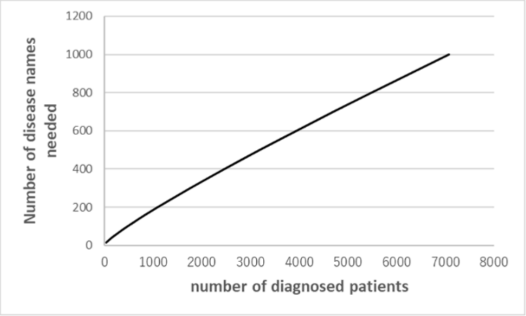


Fig. Ⅲ-3-10 Number of disease names needed for power-law distribution.

Figure III-3-10 shows the number of disease names required, as determined in this way. The horizontal axis indicates the number of patients who need to be labeled. The vertical axis shows how many disease names are required to label those patients. As can be seen from the graph, the number of disease names required increases forever as the number of patients increases. The number of necessary disease names increases indefinitely as the number of patients increases.

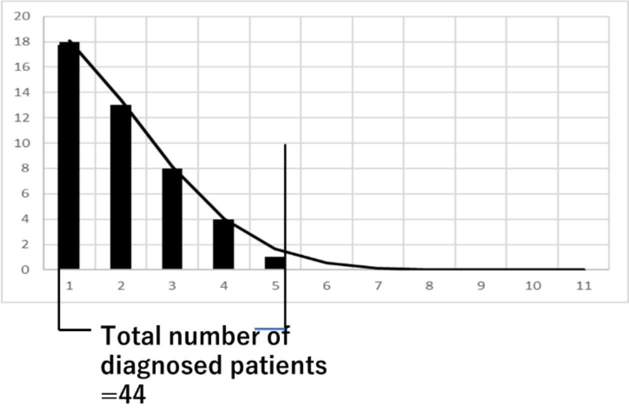
Let us do the same calculation if the frequency of disease names distribution has a normal distribution. Figure III-3-11 shows a schematic diagram of the case where the frequency of disease names is normal (the right half of the normal distribution is used). The number of disease names required for the total number of patients can be calculated using the same idea. The figure shows that the number of necessary disease names is 5 for patients 44.

Figure III-3-12 plots the number of required disease names against the total number of patients, assuming the distribution of disease names is normal. This figure shows that the number of necessary disease names does not increase above a particular value even if the number of patients increases. With such a distribution, if a certain number of disease names are prepared in advance, the names not included in it are exceptional and can be ignored or excluded[23]. Figure III-3-13 shows the results of calculating the change in the number of required disease names when the frequency of disease names is assumed to have a power-law distribution and when it is assumed to have an exponential distribution. In the exponential distribution, as in the normal distribution, the number of required disease names does not increase above a specific value. The number of necessary disease names increases rapidly because the power-law distribution has a longer tail (outlier) than the normal distribution, and outlier values are not rare. This is precisely what is meant by the feeling one gets from experience that many of the added disease names are never used again.

The sense that "outliers are not uncommon" is also obtained from other clinical experiences. The rare and unpredictable occurrence of significant bleeding during surgery is a phenomenon experienced by many medical care providers involved in surgery. Other phenomena, such as the occasional abnormally long operation time and the number of variances in the critical path does not decrease, can also be thought of as "outliers are not uncommon.

---

[23] The number of required disease names can be calculated theoretically (Figure III-3-13). See Appendix III-5 for the calculation formula.



Total number of diagnosed patients =44

Fig. Ⅲ-3-11　Calculation of the number of disease names needed for normal distribution

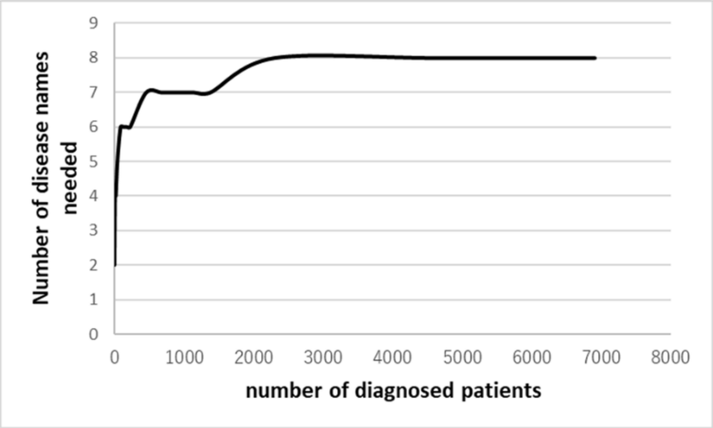


Fig. Ⅲ-3-12　Number of disease names needed for normal distribution.

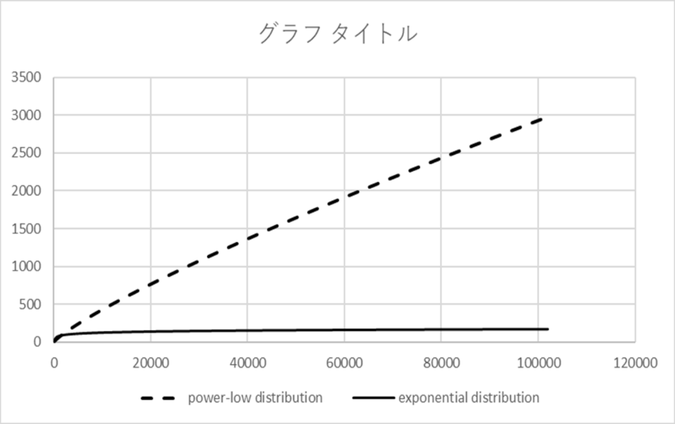


Fig. Ⅲ-3-13　Number of disease names needed
(theoretical calculation)

### 3-2 Mean value does not exist.

The second characteristic, "no mean value," is illustrated using the distribution of hospitalization lengths as an example. We have already mentioned that the length of hospital stay has a long tail distribution similar to a power-law distribution. Let us consider what happens if we assume the distribution is a power-law distribution. The mean length of hospital stay can be calculated from the distribution, but as we will see in a moment if the power-law exponent is less than 2, the mean is divergent (infinite). Also, the variance diverges if the power-law exponent is less than 3[24]. Thus, if the power-law exponent is less than 2, the mean and the variance of the length of hospital stay diverge. Mathematically, this is all good, but how should we interpret the divergence in the length of hospital stay? This point is explained below.

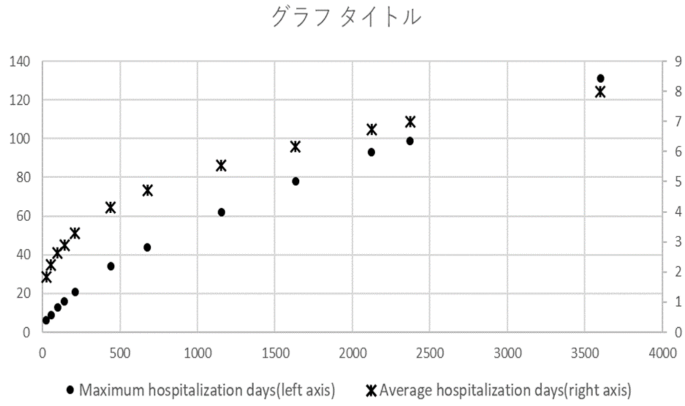


Fig. III-3-14 Change in Hospitalization Days by Number of Patients (power-law distribution $\alpha=1.5$).

Suppose 45 patients are admitted to the hospital, and their length of hospital stay distribution follows *A/X.* In this case, it is possible to calculate the average length of hospital stays. The distribution of length of hospital stay is like Figure III-3-9 with A=15[25]. In this case, the average length of hospital stays is (15(person)×1(day)+7×2+5×3+...+1×15)/45(person)=4.2(days). Similarly, if A=30, the number of hospitalized patients would be 111, and the average length of hospital stays can be 6.86 days. The average length of hospital stays increased from 4.2 to 6.86 days. Figure III-3-14 shows the results of calculating the average length of hospital stays when the number of hospitalized patients increases similarly. The figure shows the relationship between the

---

[24] In the case of a power-law distribution, the mean is calculated as follows.

$$\text{Mean} = \frac{\sum_{n=1}^{\infty} n \frac{A}{n^{\alpha}}}{\sum_{n=1}^{\infty} \frac{A}{n^{\alpha}}} \approx \frac{\int_{1}^{\infty} X \frac{A}{X^{\alpha}}}{\int_{1}^{\infty} \frac{A}{X^{\alpha}}}$$

. This equation is $\infty$ if $\alpha \leq 2$. Variance can be calculated in the same way.

[25] In Figure III-3-9, the horizontal axis is read as the length of hospital stay and the vertical axis as the number of persons in that length of hospital stay.

number of hospitalized patients and the average and maximum length of hospital stays. It is clear from the figure that as the number of patients increases, both the average and maximum lengths of hospital stays increase. This increase will not stop forever. The average length of hospital stays increases wherever the number of patients increases. This is what the divergence (infinity) of the average length of hospital stays means.

Incidentally, Figure III-3-15 shows the change in the average length of hospital stays when the length of hospital stay is assumed to have an exponential distribution[26], in which case the average length of hospital stays remains constant even as the number of patients increases. The same is true for a normal distribution.

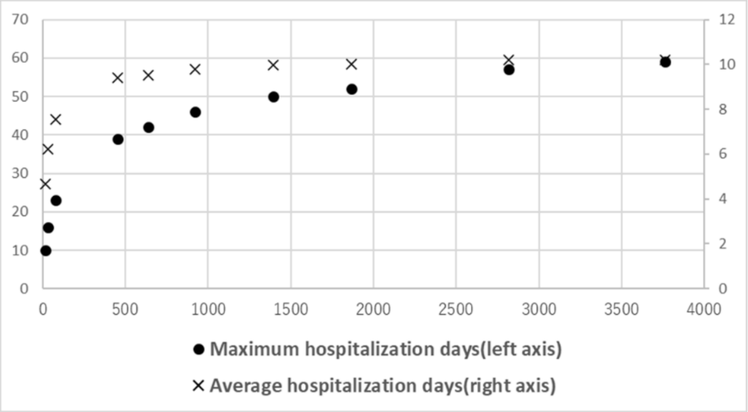


Fig. III-3-15 Change in Hospitalization Days by Number of Patients (exponent distribution ($Ae^{-0.1x}$)).

Thus, the practical meaning of the divergence in the average length of hospital stays is that the average value increases as the number of patients (or observation time) considered increases. Of course, such a mean that varies with the size of the population would not be usable as an indicator of the population.

Secondly, the variance divergence means the data's variability will be significant. Even if there exists a mean value ($2<\alpha\leq3$), the variation in the length of hospital stay for each patient will be too large to use the mean as an indicator of the population[27]. Outliers would need to be discarded to force it to be used as an indicator. For example, it would be necessary to discharge long-term hospitalized patients. Currently, hospitals are under pressure to reduce the length of hospital stay. This is a natural consequence of the nature of the power-law distribution.

Whether the length of hospital stay has a power-law distribution is fundamental in medical care economics. Recently, a patient classification method called DRG (Diagnosis Related Group) has been developed as a medical care management tool [52]. DRG is a classification method for hospitalized patients. DRG is a method of classifying hospitalized patients into approximately 500 categories based on disease names and other indicators. This classification method is based on the property that "patients in the same category are similar in the amount of medical care resources they require, such as length of hospital stay, medical supplies, and human resources." The amount of medical care resources used is related to the cost of treatment. Classifying patients into DRGs makes it possible to predict the amount of medical care resources used and their costs during hospitalization. This property is used for the

---

[26] Exponential distribution is the distribution expressed by the formula $Y = Ae^{-\alpha X}$. The properties of the distribution discussed in this section are similar to those of the normal distribution.

[27] Not only the mean, but other statistical indicators such as the median cannot be used as well.

flat-rate payment system for medical care of each DRG (DRG/PPS: Diagnosis Related Group - Prospective Payment System), hospital quality and cost management, etc.

However, for such a classification to be possible, resource use must have a normal distribution or at least a finite variance. If this assumption is invalid, then the DRG classification and its applications, such as the Diagnosis Procedure Combination (DPC) method itself, are invalid. This result is significant. Length of hospital stay is one of the most essential medical care resources for hospitalized patients, and a power-law distribution with an exponent of 2 or less will cause the variance and the mean to diverge. This indicates that the very premise of the DRG classification may break down. In fact, the length of hospital stay is not a power-law distribution in the strict sense. The tail is shorter than the power-law distribution (see Chapter 2-2). However, the mean and variance are still significant, so the interpretation described here would not be so wrong.

## 4 Causes of power-law distribution in medical care.

### 4-1 Three types of power-law distributions in medical care.

As we saw in the previous section, power-law distributions profoundly impact medical care. The issue of power-law distributions or "long tail distributions" in medical care cannot be avoided when considering medical policy and management. Investigating why power-law distributions occur in medical care is not merely a matter of interest. Once the mechanisms are understood, it will be possible to understand the possibilities of controlling and dealing with such power-law distributions.

The general causes of power-law distributions, including physical phenomena, have already been discussed in Chapters 2-2. Here, we briefly describe a theory that may help consider the causes of power-law distributions in medical care. The power-law distributions observed in medical care can be classified into three major types. The first is the frequency distribution of categories, such as the frequency of disease names and the frequency of surgical procedures. The second is the volume distribution, such as blood loss during surgery. The third is related to time distributions, such as length of hospital stay, duration of surgery, and time to recovery from illness. These properties are quite different, and the possible mechanisms are likely to be different as well. Since the frequency distribution of disease names will be discussed again in Chapter 4, we will consider the length of hospital stay distribution and the distribution of blood loss during surgery.

### 4-2 Distribution of length of hospital stay.

As mentioned in Chapter 3-2, the length of hospital stay distribution has a long tail distribution. It is generally said to have a lognormal distribution [24]. As can be seen from Figure III-3-5, both logarithmic graphs are slightly convex upward compared to the regression line. From this, it is said that the length of hospital stay distribution may be a lognormal distribution[53] (see Chapter 3-2). However, the cause of its formation is not clear.

The length of hospital stay is included in survival time analysis in a broad sense [25]. In medical science, it is used to test whether there is a difference in survival time between treated and untreated groups to see the effect of treatment. The survival time distributions assumed include exponential distribution, log-logistic distribution, log-normal distribution, and inverse Gaussian distribution. The lognormal and inverse Gaussian distributions are the best models for the length of hospital stay based on comparing data and these distributions. Whitmore has shown that the length of hospital stay follows an inverse Gaussian distribution using a 1-dimensional random walk model with a positive drift term [55].

Before considering the model with a Drift term, let us consider the random walk model. In this model, the patient's daily health status is represented by a single integer $S$. The health status at the time of admission is 0 and randomly changes by +1 or -1 each day. The patient is discharged from the hospital on the first day the health status $S$ reaches a specific value $z$(>0). The model can be expressed formally as follows.

◇ **Random walk model of length of hospital stays.**

Consider the state variable $S(n)$ as the patient's state at date $n$.

The patient is assumed to be in state 0 on admission. The discharge decision is made when the state reaches z (>0).

The state changes randomly each day, only improving (+1) or worsening (-1). That is, the following equation represents it.

$$S(n+1) = S(n) + X_n \ . \qquad \text{(III-3-2)}$$

$$X_n = \begin{cases} +1 & \text{probability } p \\ -1 & \text{probability } q = 1 - p \end{cases} \ .$$

The length of hospital stay is defined as the time *n* when S(*n)* = z.

For the above simple model, the probability $w_{z,n}$ that the length of hospital stay is *n* can be calculated theoretically and is as follows [20].

$$w_{z,n} = \frac{z}{n}\binom{n}{\frac{1}{2}(n-z)} p^{\frac{1}{2}(n-z)} q^{\frac{1}{2}(n+z)} \qquad \text{where, } \binom{n}{x} \text{ is the binomial coefficients} \ . \qquad \text{(III-3-3)}$$

If $p \neq q$ and *n is* large, equation III-3-3 asymptotically follows[28]

$$w_{z,n} \approx C\left(\frac{n-z}{2}\right)^{-\frac{3}{2}} b^{-\left(\frac{n-z}{2}\right)} \qquad C = \frac{z(2q)^z}{2\sqrt{\pi}} \quad b = \frac{1}{4pq} > 1 \ . \qquad \text{(III-3-4)}$$

Equation III-3-4 is a distribution that is a mixture of an exponential distribution and a power-law distribution. Where *n* is relatively small, it exhibits power-law distribution-like behavior and behavior similar to an exponential distribution in regions where *n* is large. In other words, this is the "truncated power-law distribution" described in Chapter 2. Therefore, it is possible to consider this model as one of the lengths of hospital stay models. However, we have not found any papers considering the length of hospital stay as a "truncated power-law distribution."
The asymptotic distribution for *p*=*q* is as follows.

$$w_{z,n} \approx \frac{z}{2\sqrt{\pi}}\left(\frac{n-z}{2}\right)^{-\frac{3}{2}} \ . \qquad \text{(III-3-5)}$$

This is a power-law distribution with an exponent 1.5. This exponent of 1.5 is an exciting number. According to actual data from hospitalization distributions [56], many exponents fall between 1 and 2. This indicates that the model has some validity. However, a slight variation of this model yields Whitmore's one-dimensional random walk model with a drift term[24][55].

The model is obtained by adding a positive drift term to the above random walk model. In essence, the idea is that the patient's condition improves at a constant rate due to treatment and other factors. The model is Equation III-3-2 with a constant positive *γ* (a positive integer). In other words,

$$S(n+1) = S(n) + X_n + \gamma \ . \qquad \text{(III-3-6)}$$

If the random term $X_n$ is absent, the length of hospital stay will always be $\frac{z}{\gamma}$. This is the model that the length of hospital stay varies with random fluctuations in the state. The distribution of this model is inverse Gaussian.

An inverse Gaussian distribution means that if the length of hospital stay is *x* (a real number), its probability density is expressed by the following equation using the above *z* and *γ*.

---

[28] $f(n) \approx g(n)$ (asymptotically equal) is when $\frac{f(n)}{g(n)} \xrightarrow[n\to\infty]{} c \ (0 < c < \infty)$ .

$$p(x) = \frac{z \cdot exp(z\gamma)}{\sqrt{2\pi}} x^{-3/2} \exp\left(-\frac{1}{2}(\gamma^2 x + z^2/x)\right) \quad \text{(III-3-7)}$$

The mean value is $z/\gamma$. Figure III-3-16 shows the density function of the inverse Gaussian distribution. Both logarithmic graphs are convex at the top and shift downward from the straight line at the long tail.

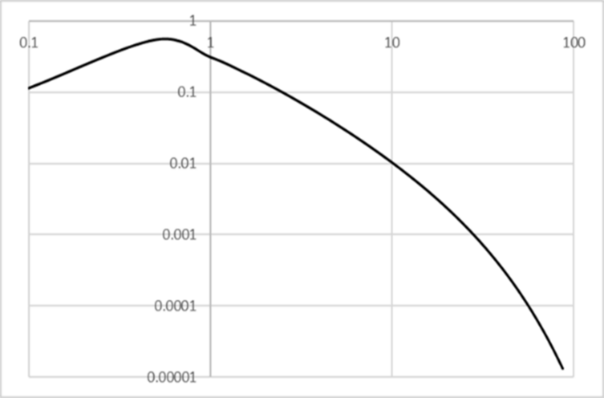


Fig. III-3-16 PDF of inverse Gaussian distribution
z=1 $\gamma$ =0.3
The graph decreases like a power-law distribution after the peak, but it is convex upward.

The three long tail distributions already mentioned, i.e., log-logistic, log-normal, and inverse Gaussian, are similar in shape, and it is difficult to distinguish between the three from actual data. It seems that which distribution to adopt depends on the purpose and convenience of use. However, as a criterion for determining which distribution is correct, the validity of the model in which the distribution appears should also be considered.

The distribution of length of hospital stay varies by hospital type. In Japan, hospitals are divided into acute care, convalescent care, and long-term care. All the distributions of the length of hospital stay are long-tail distributions. When a power-law distribution approximates them, all the power-law exponents are between 1 and 2. When the power-law distribution models the length of hospital stay distribution, all hospitals' distributions are represented by a single parameter, the power-law exponent. However, as shown in Kouchi et al.'s paper, the average length of hospital stay varies widely among different types of hospitals [56]. If all hospitals' distribution followed a power-law distribution represented by a single parameter, they should all have the same distribution function for more patients. However, the data show that this is not the case. This indicates that it is impossible to represent the length of hospital stay with only one parameter. All "long tail distributions" besides the power-law distribution described above have multiple parameters. If they have various parameters, the different parameters can represent different types of hospitals. From this point of view, it seems a bit unreasonable to say that the length of hospital stay is a perfect power-law distribution.

**4-3 Intraoperative blood loss**

Figure III-3-17 was created from data on blood loss during surgery at a particular hospital. A rough linearity is obtained by looking at the double-logarithmic graph of bleeding volume frequency (B). Thus, the bleeding volume frequency also appears to be a power-law distribution. However, the frequency graph deviates from the straight line downward at the points where the bleeding volume is large. In addition, although data are not shown, the distribution of blood loss varies greatly depending on the type of surgery. In minor surgeries such as thyroid surgery, blood loss is usually several dozen to 100 ml. In contrast, stomach surgery usually several hundred milliliters, and esophageal cancer and liver surgery are not uncommon to lose several liters. If all bleeding volumes could be expressed in a single power-law distribution, thousands of thyroid surgeries could result in several liters of blood loss. From experience, that is hardly possible. Given this, it is difficult to explain the amount of blood loss with a single power-law distribution. There must be different distributions for each type of surgery. Moreover, a sharp drop in frequency distribution, such as the "truncated power-law distribution," is needed where blood loss is significant.

By the way, blood vessels, the source of bleeding, have a very complex branching morphology. Such forms cannot be represented by simple geometry. Recently, fractal geometry has been developed to represent such forms[4][5]. Many papers have been published showing that the vascular system is fractal[57][58][59][60][61]. And Takayasu has created a model for such fractal flow rates [62].

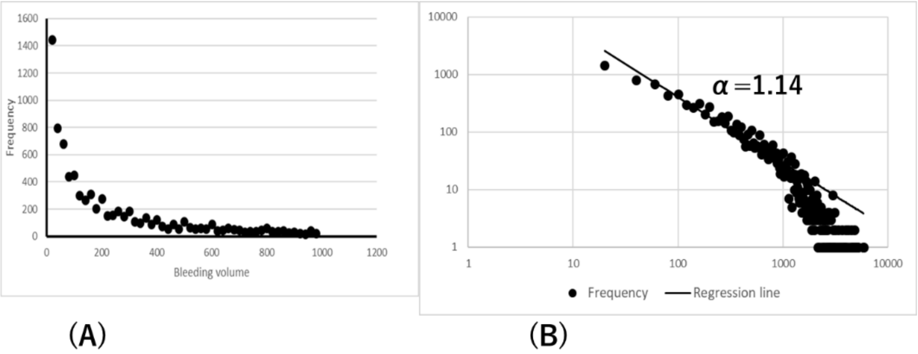


Fig. Ⅲ-3-17 Distribution of blood loss during surgery in an acute care. hospital
(A) Frequency distribution
(B) Double logarithmic graph
The line shifts from straight line to the bottom where the amount of bleeding is large.

Takayasu's model is as follows.

There is a particle of integer mass placed on a one-dimensional lattice. In discrete time, these particles aggregate in a random walk. Formally, it is defined as follows.

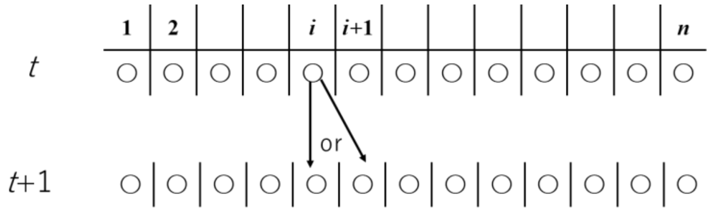
(1) There is a particle at every site of the lattice.

(2) They jump to the right site with a probability of 0.5 each time and stay at the original site with a probability of 0.5.

(3) After jumping, if two or more particles collide, they come together to form a larger particle.

(4) A unit particle is then added to every site. If a particle of size n exists in that site, it will be n+1.

Then, repeat steps (2) through (4) (Figure III-3-18).



図III-3-18 Schematic diagram of Takayasu model.
A particle at time t stays at i or moves to i+1 at the next time t+1.
The particle that moves to i+1 merges with the particle originally at i+1, increasing its mass.

In this model, after a sufficiently long time, the mass $s$ of each site particle has a power-law distribution. In other words, expressed in terms of the complementary cumulative distribution function (CCDF), it becomes $P(s \geq x) \propto 1/x^{\beta}$. Here, β is about 0.331. For the power-law exponent $\alpha$ in the case of the ordinary probability density function, $\alpha = \beta + 1$, so $\alpha$ is about 1.331. If we change the probability of jumping to the left or right in various ways, $\beta = \alpha - 1 \cong 1/3 \sim 1/2$.

The model can be viewed differently to arrive at the idea of a distribution of flow rates at each point of the river. Consider a one-dimensional site as a grid point in a two-dimensional plane. Consider each grid point as a water source spewing at a flow rate of 1. Consider a point $(x, y)$. From this point, a flow occurs. Assume that at the next time ($t$+1), this flow will flow to either the straight or right point, i.e., $(x,y+1)$ or $(x+1,y+1)$[29]. If, at the new point, it merges with another stream, the two flows are combined, plus the flow spewing at that point. If there is no merging, the water jet from that point is taken into the flow. At a point that does not flow in from anywhere, a new flow occurs at t+1. Thus, a flow diagram can be drawn.

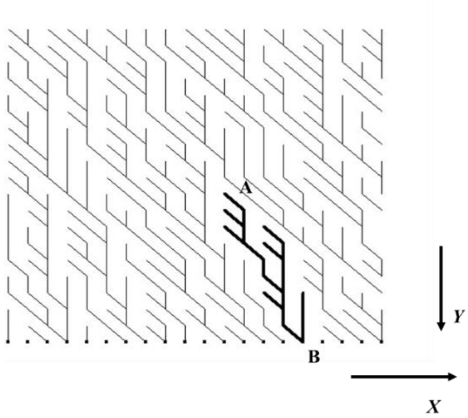


Fig. III-3-19 Takayasu model.
Figure modified by the author from reference [62].

---

[29] To be precise, it is not left or right, but either straight or right (x+1). The reason why it "does not flow to the left" is that if it flows to the left, water from an adjacent source will intersect it if it flows to the right.

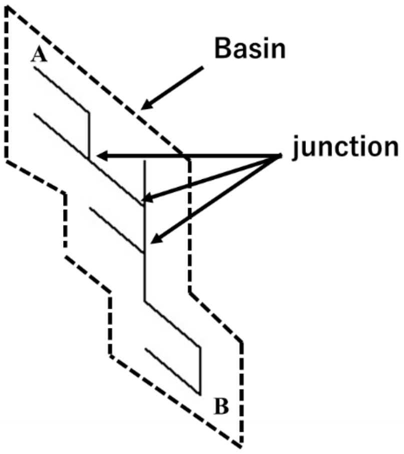


Fig. Ⅲ-3-20　Basin of the Takayasu model.
Figure modified by the author from reference [62].

For example, the bold lines from the starting point of A to B in Figure III-3-19 can be seen as a diagram of the river flowing into B. Figure III-3-20 is a redrawn version of this section. In this figure, the river is surrounded by a dotted line. This area is the watershed of the river.

Another way to look at this figure is to see it as a confluence of many tributaries. In Takayasu's model, in river terms, mass is the flow rate at a given point. When two branches merge, the flow at the confluence is the sum of the flows of their respective tributaries. Ultimately, the flow at issue is the sum of the flows of all branches upstream from that point. This is equal to the area of the watershed upstream from the end (or, more precisely, the sum of all the water gushing out in that watershed). The interpretation of the model's results is that this flow has a "power-law distribution. The area of a watershed depends on the geometrical properties of the river. Takayasu's model can be paraphrased as "if a fractal figure represents a river, then the area of its watershed follows a power-law distribution.”

Takayasu's theory is in its simplest form for rivers. This theory has been applied to more complex actual rivers by Rodríguez-Iturbe and others [63]. Let us briefly discuss the conclusions of that theory. Figure III-3-21 is a schematic diagram of part of a river map in Osaka Prefecture. A complicated branching diagram like this river is generally called a fractal. Please refer to References [4][5] for details, but a fractal is a figure in which the dimension of the figure is a number with a decimal point. Conventional geometric figures have integer dimensions. Straight lines and curves are one-dimensional, ordinary plane figures are two-dimensional, and three-dimensional figures are usually three-dimensional. However, complex winding lines, such as mountain ridges in nature, typically have a dimension represented by a decimal greater than one and less than two. The shapes of clouds, rivers, and coastlines in nature are said to be fractals. Attractors in chaos are also fractals.

Fig. Ⅲ-3-21 Map of rivers in Osaka Prefecture.
quote from the Osaka Prefectural Government website

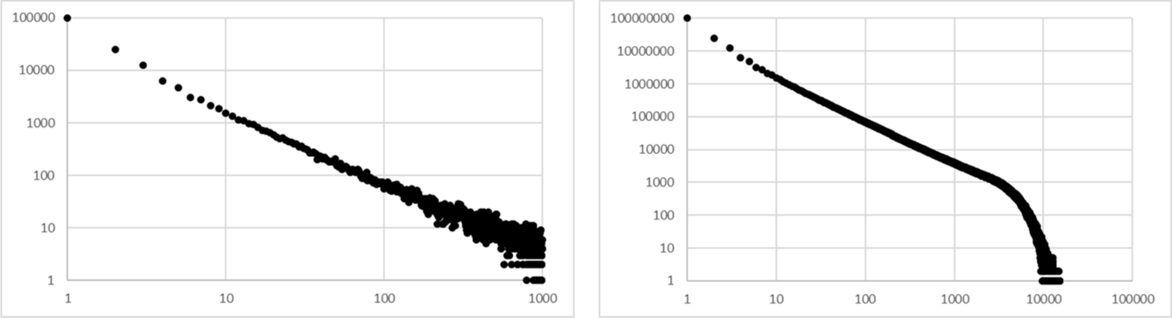
Rodríguez-Iturbe et al. [63] extended Takayasu's theory to derive the relationship between the fractal dimension of a river and the power-law exponent in Takayasu's theory. They conclude the following by using actual river maps and theoretical considerations.

(1) Actual rivers are also considered fractal.

(2) The fractal dimension $D$ of a river can be estimated from a photograph of an actual river.

(3) There is a relationship $\beta \simeq 0.5/\mathrm{D}$ between this $D$ and the (CCDF) power-law exponent $\beta$.

The fractal dimension obtained from the river photo was $D \simeq 1.1$. We ended up with $\beta \simeq 0.5/D \simeq 0.4545$ ($\alpha \simeq 1.4545$), agreeing with the measured value of 0.43.



(A) (B)

Fig. Ⅲ-3-22 Simulation of the Takayasu model with a grid size of 200
(A) Number of simulations $2\times10^{5}$回 Power-law distribution
(B) Number of simulations $2\times10^{7}$回 Truncated power-law distribution

If rivers are fractal, assuming that the vascular system would also be fractal is natural. Indeed, many papers have discussed fractal structures in living organisms[57][59][60][61][64]. Among them, the fractal nature of blood vessels has also been discussed. If the vascular system is fractal, Takayasu's theory could be adapted to blood flow. Since bleeding is a rupture of a blood vessel, it would be proportional to the flow rate at the location of the fracture. Therefore, Takayasu's theory can be used to consider the distribution of bleeding volume. However, when applying Takayasu's theory to bleeding volume, it is necessary to consider the lattice size in Takayasu's approach, which is not clearly stated in Takayasu's view. Still, there is an implicit understanding that the lattice size is infinite. When the lattice size is infinite, the blood flow distribution shows a perfect power-law distribution. However, when the lattice size is finite, the "truncated power-law distribution" described in Chapter 2 occurs. Simulations of the Takayasu model with a finite lattice size show that the frequency distribution approaches zero earlier than the power-law distribution in the tail (Figure III-3-22).

As can be seen from the figure, a finite lattice size results in a truncated power-law distribution with a steep decrease in the more significant portions of the distribution compared to the power-law distribution. This is consistent with the actual hemorrhage distribution being more convex on top than the power-law distribution, and changing the grid size results in different amounts of bleeding at the points that deviate from the power-law distribution. This indicates the possibility of obtaining other distributions for different types of surgeries and increasing the lattice size results in a hemorrhage model for extensive surgeries. By making the lattice size correspond to the scope of the surgery, it may be possible to explain the differences in the bleeding distribution for different types of surgeries.

## Chapter 4: The Nature of Complexity in Medical Care

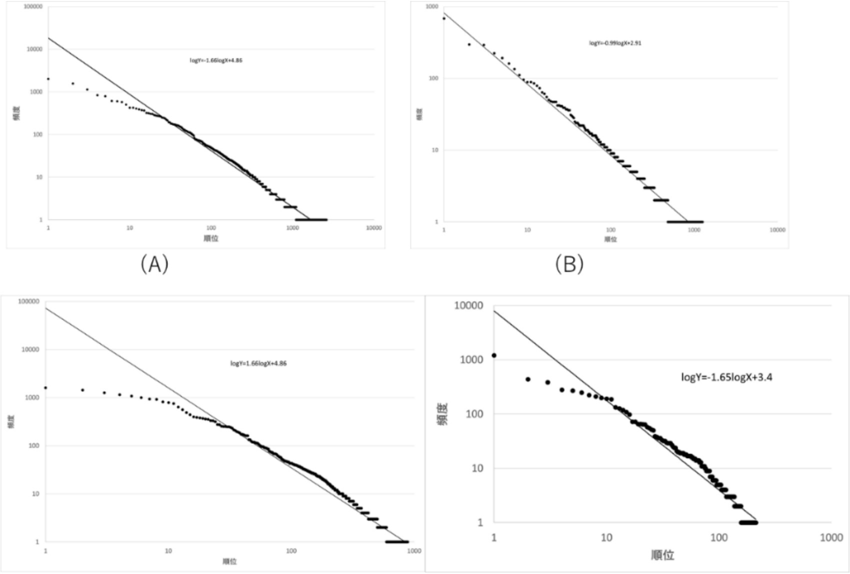
**Y. Tachimori**

### 1. Distribution of diseases

As mentioned in Chapter 3-2, the frequency distribution of disease names satisfies Zipf's law. Let us discuss it in more detail. Several studies have examined disease name distribution [48][49]. In all studies, the frequency distribution of disease names satisfies Zipf's law.

Figure III-4-1(A) is a Zipf plot of disease names in internal medicine at one acute care general hospital. When the disease names are listed in order of frequency, the rank is plotted on the x-axis and the frequency on the y-axis. The figure is presented as a double-logarithmic graph. As you can see from the figure, the chart fits well on a straight line at the tail ( the data are served well by regression lines .). In other words, the distribution of disease names follows Zipf's law (power-law distribution). Figure III-4-1(B) is a Zipf plot for ophthalmology. This plot also follows a power-law distribution. Moreover, the curve rides on a straight line better here.

Figures (A) and (B) in Figure III-4-1 are Zipf plots when the disease name is freely written in Japanese. In contrast, (C) and (D) show the frequency distribution for the same hospital (disease) when the disease name is coded in ICD-10[30,31] [49]. The data are also fitted well by regression lines. In internal medicine and ophthalmology, the slope of the graph is larger for ICD codes than for freely written names, and the deviation from the straight line tends to be more prominent in the areas where the ranks are more minor.



(A) (B) (C) (D)

Fig. III - 4 -1 Zipf plots of Disease frequency (freely written group and ICD coded group), double logarithmic graphs.
A: Internal medicine: freely written, Zipf exponent=1.32.
B: Ophthalmology: freely written, Zipf exponent=0.99.
C: Internal medicine: ICD coded, Zipf exponent=1.66.
D: Ophthalmology: ICD coded, Zipf exponent=1.65

---

[30] International Statistical Classification of Diseases and Related Health Problems [65].

[31] However, (C) and (D) are from a different period than (A) and (B).

Frequency distributions were also examined for the other departments, showing good fits to the straight lines on both logarithmic graphs. However, there is a tendency to deviate from the regression line in the areas with high rankings. Table III-4-1 summarizes the frequency of disease names freely written in Japanese and coded in ICD-10 for each department. The Zipf plots for all the departments fit the regression line well. For the frequency of disease names freely written in Japanese, the Zipf exponent $\zeta$ is around 1 for all departments, while for ICD codes, the Zipf exponent tends to be larger than that of freely written in Japanese. In addition to each department, we examined each doctor, and all satisfied Zipf's law[32].

| | free written group | | | coded group | | |
|---|---|---|---|---|---|---|
| | $\zeta$ | $r^2$ | N | $\zeta$ | $r^2$ | N |
| total | 1.20±0.002 | 0.99 | 93253 | 1.64±0.007 | 0.95 | 132621 |
| internal medicine | 1.32±0.006 | 0.98 | 29640 | 1.66±0.012 | 0.96 | 30788 |
| pediatrics | 1.30±0.009 | 0.98 | 8120 | 1.43±0.009 | 0.97 | 12164 |
| surgery | 1.14±0.007 | 0.98 | 7946 | 1.31±0.009 | 0.97 | 7278 |
| neurosurgery | 0.96±0.008 | 0.98 | 1995 | 1.07±0.014 | 0.96 | 1358 |
| ophthalmology | 1.04±0.005 | 0.99 | 5907 | 1.65±0.028 | 0.94 | 6440 |
| orthopedics | 0.83±0.004 | 0.98 | 5770 | 1.58±0.020 | 0.94 | 10350 |
| cardiac surgery | 0.97±0.023 | 0.94 | 831 | 0.97±0.014 | 0.96 | 791 |
| dermatology | 0.98±0.006 | 0.98 | 3623 | 1.41±0.012 | 0.96 | 6950 |
| urology | 1.25±0.009 | 0.99 | 4289 | 1.50±0.015 | 0.97 | 5227 |

**Table III-4-1 The result of regression analysis in free written group and coded group**

The set of disease names diagnosed by each physician, when collected for the entire department, equals the set of disease names aggregated for each department. Also, if all the disease names aggregated by the department are collected, it becomes the total disease name set of the hospital. The fact that these all obey Zipf's law suggests that merging two sets that obey Zipf's law (let's call this the Zipf's set) will create the Zipf set again.

Let us illustrate this with a simple example. Suppose we have two sets of the same size with no common disease names at all, each obeying Zipf's law (Zipf's exponent = 1) (for example, two sets of disease names from two departments with few common disease names, such as ophthalmology and urology, are close to this condition). Let $N$ be the number of disease names in each set; if the two groups are merged, the number of disease names becomes $2N$. If the two sets have the same frequency distribution and no two sets have the same disease names, the disease name with the $n$th frequency of one group will have the $2n$th (let $K$ *denote* this) or ($2n$-1)th frequency in the merged set. The frequency is $1/n$. The $K$th frequency of the disease names $Y_K$ *is* $Y_K = Y_{2n} = 1/n = 2/2n = 2/K = Y_{2n-1} = Y_{K-1}$. Therefore, it follows Zipf's law asymptotically with the same Zipf exponent at large $K$. However, at small $K$, it deviates from Zipf's law. Figure III-4-2 displays the graph of the case of the five sets together that do not contain the common disease name. At the low ranks (large $K$), the frequencies indeed asymptote to a straight line. However, at the low ranks (small $K$), they deviate from the expected straight line, and the frequencies become smaller.

---

[32] As already mentioned many times, the relationship between the power-law exponent $\alpha$ and the Zipf exponent $\zeta$ is $\alpha = (\zeta + 1)/\zeta$ .

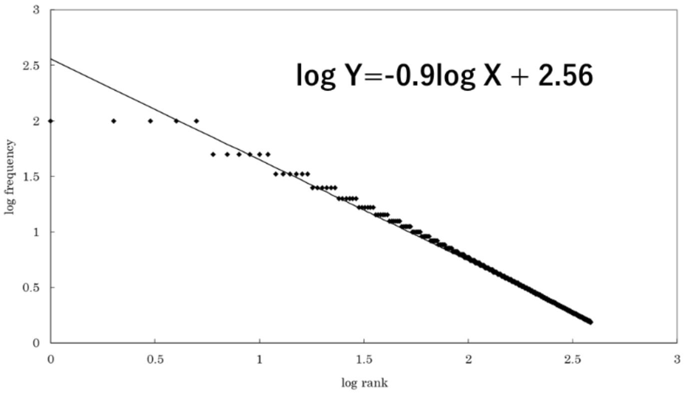


**Fig. Ⅲ-4-2 Zipf plot of the union of five Zipf sets (double logarithmic graphs).**

The above example shows that when several Zipf sets are merged, the frequencies asymptotically follow Zipf's law. However, where the frequency is significant, it deviates from Zipf's law. In the actual set of disease names, the data points are smaller than the expected straight line at high frequencies. This suggests that the set of disease names analyzed in this study (total disease names, departmental disease names, and physician names) may be a merged set of more minor disease name sets that follow Zipf's law.

The second characteristic of the data is that the value of the Zipf exponent for the coded group of disease names is more significant than that of the Zips exponent for the freely written group. Let us consider the influence of coding as the cause of this difference. Compared to the coded group, when disease names are freely written in Japanese, multiple names are given to the same disease.[33] This leads to a greater variety of disease names. In other words, coding has the effect of combining various disease names into a single disease name. Conversely, a single coded disease name is broken down into numerous, more subdivided disease names freely written in Japanese. In this case, we consider how the frequency distribution of disease names changes when the new classification is used. First, assume that a code disease is divided into a certain number of disease names and that the frequency of disease names is equal. In this case, the new set of disease names still forms a Zipf set, but $\zeta$ does not change. This can also be shown by merging two non-overlapping Zipf sets, producing a Zipf set with the same $\zeta$.

Next, let us assume that the individual disease name sets created by decomposing a single disease name are also a Zipf set. Although it is unclear why word distributions such as sentences follow Zipf's law, the act of free choice among many notations for the same meaning may lead to this distribution. In this sense, it is quite possible that the same principle is at work in the free reminder of disease names and that they follow the Zipf distribution.

Consider a set of disease names in code notation. The frequency of disease names with the *xth* frequency ($F_x$ ) is expressed as follows (where [] denotes rounding down to the nearest integer).

---

[33] For example, the disease name gastric cancer is differently expressed as gastric cancer, stomach cancer, gastric carcinoma, etc.

$$F_x = \left[\frac{A}{x}\right] \qquad x = 1,2,\ldots \quad . \tag{III-4-1}$$

Let us assume that each disease name is decomposed into several disease names and that the decomposed disease name set again has a Zipf distribution where the Zipf exponential distribution $\zeta$ is 1. The frequency of $x$th disease names is $F_x$, so let us assume this is decomposed into $n$ disease names.
Then, the following equation holds.

$$F_x = \sum_{j=1}^{n}\left(\frac{B}{j}\right) \quad . \tag{III-4-2}$$

Determine a hypothetical $B$, $n$ such that this is satisfied as follows.
Suppose the minimum frequency of disease names is 1, so B/n=1.
Substituting $B=n$ into (III-4-2) yields the following equation.

$$F_x = \sum_{j=1}^{n}\left(\frac{n}{j}\right) = n\sum_{j=1}^{n}\frac{1}{j} \cong n\int_{1}^{n}\left(\frac{1}{t}\right)dt = n\log n\,. \tag{III-4-3}$$

Then, $n$ is determined so that the following conditions are satisfied.

$n\log(n) \le F_x < (n+1)\log(n+1)$.

Next, find s satisfying the following equation.

$$F_x = \sum_{j=1}^{s}\left\{\left[\frac{n}{j}\right]+1\right\}. \tag{III-4-4}$$

In this way, the $x$th disease name is decomposed into $s$ disease names.

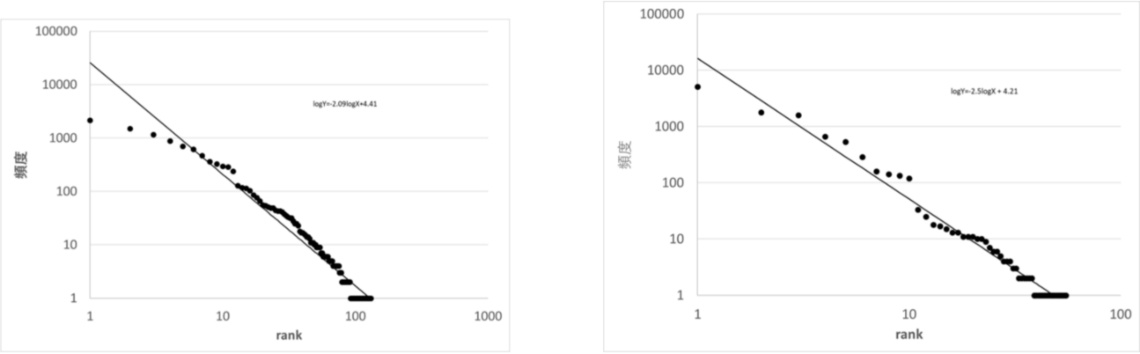
When this was done, we used simulations to determine the frequency distribution. In the simulation, we used data such that $F_1$ =A=10000 in equation (III-4-1). In this case, the Zipf exponent of the original set of disease names is 0.97. When each disease name was broken down into multiple names using the method described above, the number of disease name species was 19686, and $\zeta = 0.67$.

Thus, when each disease name set is decomposed into disease name sets with a Zipf distribution, a Zipf set with a smaller Zipf exponent is obtained. Conversely, the coding of disease names is the reverse of the above process and can be thought of as grouping a group of disease names that follow Zipf's law and treating them as a single disease name. If coding has such an effect, the Zipf exponent is expected to increase due to coding.

Actual data further examined this. The disease codes used in this study are six-digit, consisting of a two-digit code added to the original four-digit ICD code to accommodate more detailed classification. We then examined what happens when disease names are distinguished using only 4-digit codes out of the 6-digit codes. The result was the same Zipf set, but the Zipf exponent was more prominent for the 4-digit disease name set. This proves the correctness of the above assumption.

There was a significant deviation from Zipf's law for the internal medicine disease name set and the total disease name set in the analysis of code disease names. Let us examine the reasons for this deviation. First, internal medicine includes more disease names than other departments. This indicates that internal medicine disease names are a merged set of more basic disease names. As mentioned above, combining Zipf sets with no common parts asymptotically follows Zipf's law, but at high frequency, the frequency deviates from the straight line. In internal medicine, there is a possibility that many set mergers occur, and therefore, the frequency deviates from the straight line at high frequency. The same thing could be expected to happen with total disease. Another possibility is a saturation of disease names. With disease names by code, the number of disease names is limited to the number of codes registered. The master data of disease names used in this study has about 12,000 disease names registered. As the number of disease names assigned to patients increases and approaches the upper limit of the number of disease names on this master, there is a possibility of disease name saturation and a deviation from Zipf's law. In the case of total disease names, the number of disease names exceeds 3200, which may have had this effect.

In this study, the disease name set was subdivided by department, physician, and ICD-10 code, and Zipf sets were obtained for all these sets. The fact that Zipf's law holds for all these different classification criteria and that the sum of Zipf's sets again asymptotically follows Zipf's law suggests that Zipf's law may hold at a smaller set level than the classification set in this analysis. If this is the case, then all the classifying sets in this analysis are represented by the sum of these elemental Zipf sets, which are, of course, Zipf sets again. To prove the existence of such a basic set, it is necessary to construct and analyze a model using a constructivist method that shows how the set of disease names is created.



(A)　　　(B)

**Fig. Ⅲ-4-3　Disease frequency in hospitals specializing in thyroid**
A: free written
B: ICD coded

The results discussed so far are for general or internal medicine hospitals that treat many illnesses. The distribution of disease names in specialized hospitals with only certain patients or in more specialized hospitals such as university hospitals may differ. Therefore, we examined the distribution of disease names in hospitals specializing in thyroid diseases. The results are shown in Figure III-4-3. Figure (A) shows the frequency distribution of disease names freely written in Japanese, and (B) shows the frequency distribution of ICD codes associated with the same data. The Zipf exponent in graph (A) is 2.09, and that in graph (B) is 2.5. As in the case of the general hospital, the Zipf exponent for ICD codes is more prominent. However, both Zipf exponents are more significant than the general hospital's. Each patient usually has more than one disease, and the data from the general hospital counts the number of diseases when a patient has more than one disease. Patients coming to a specialized hospital may have multiple illnesses, but only diseases treated by that hospital may be registered. Therefore, the same phenomenon as coding may have occurred, resulting in a more prominent Zipf exponent.

## 2. Disease naming depends on context

The names of diseases are not all parallel to each other. Some disease names are strongly related to each other, while others are not so related. For example, a patient with common cold symptoms might have a common cold, upper respiratory infection, acute rhinitis, acute pharyngitis, acute bronchitis, etc. These names can easily be shifted to other diseases. These names are closely related because they can be easily transferred (or mistakenly given) to other diseases. However, since it is unlikely that a femur fracture and an upper respiratory tract infection could be transmitted, these two disease names could be considered distantly related. Of course, the relationship of two disease names to each other, close or distant, depends on the scale at which they are viewed. There are many possible scales, such as etiologically close, symptomatically close, close in severity, close in social impact, and so on. When a physician examines a patient, which scale they use depends on the situation (initial consultation, special examination, hospitalization, social status of the patient, beginning or middle of treatment, etc.). However, it is thought that the patient is examined with some scale each time. Depending on their scale at that time, the names of diseases closely related to each other may be shifted when making a diagnosis. When disease names with such a close relationship are collected, the disease name set may be considered to follow Zipf's law.

For example, suppose that the name of a disease given to a patient is one of a set of disease names that collect close relations to each other. Moreover, due to the ambiguity of the disease name, there may be no absolute standard for naming the disease. In such cases, choosing which names to use depends on the physician's free will. This situation is the same as selecting words when composing a sentence. Therefore, these disease name sets would follow Zipf's law. Such a set is considered to be an elemental set of disease names.

A disease is named because of observations made by a physician on a patient. The physician's knowledge influences this, the instruments they use to diagnose, the patient's environment, and their social circumstances. Thus, the name of a disease is not an objective fact from the beginning but is a product of the interaction between the patient, the physician, and the environment. Therefore, even if the patient is the same, the output disease name may change simply by changing the condition of the physician or the environment. This suggests that changing the observation method can change the distribution of disease names.

Disease names are created depending on a variety of observational situations. There are two significantly different observation situations. The first is the general medical examination. In the first situation, a physician selects a disease name from an almost infinite number of possible names based on the patient's complaints and condition. The other situation is the naming of diseases in cancer screening. In this case, the physician who examines the patient selects one disease name from a small number of disease names or the disease name "no abnormality." If these two observations were made for the same patient population, the frequency of disease names would be pretty different.

In sets that follow Zipf's law, an item's frequency varies with the set's size. The mean and variance of the frequency of the entire group also differ with the size of the set. This situation is not readily accepted by conventional thinking. However, the above discussion suggests that the frequency of disease names may vary depending on the observation situation for the population. This means there is no inherent measure of frequency of disease names that is extrapolated from the sample population. This is like the situation where the discovery of fractals led to the emergence of the concept of a figure with no curve length or area.

## 3. complex networks

### 3-1 Network Definition

In the previous section, we noted that the interaction between the patient, the physician, and the environment determines disease naming. The physician's knowledge of the disease is the most critical environment for disease naming. It is not surprising that knowledge structure is essential in disease naming. To examine the influence of knowledge structure on disease naming, it is first necessary to explore the knowledge structure of disease or medical care. There are many ways to examine knowledge structure, including the network structure of knowledge. The details of the network structure of knowledge will be explained in the next section, but first, knowledge of networks is necessary as a prerequisite. A "network" is mathematically called a graph. The word "network" is mainly used in the physical and social sciences, while "graph" is primarily used in mathematics, but both mean the same thing.

When dealing with a system, such as a complex system, it is often a question of some relationship between the elements of the system. When studying chemical reactions in a cell, the question of which chemicals (components) react with which chemicals is essential. In sociology, the question is what kind of relationships human elements have with each other. The relationships can be very diverse, such as friendships, business relationships, etc. A network (graph) is an abstraction of these elements and their relationships.

In network theory, elements are called vertices (nodes or vertexes), and the relationship between each vertex is called a branch (also called a link or edge). So formally, a network is defined as follows.

**Definition: A** network is a pair G ≡ ($V$, $E$) of two sets $V$ and $E$. Where,$V = \{v_1, v_2, \dots, v_N\}$ is the set of vertices and $E = \{e_1, e_2, \dots, e_N\}$ is the set of branches. A branch is a pair of two vertices $(v_i, v_j)$.

This is the mathematical definition, but the essence is shown in Figure III-4-4. In this figure, the numbered circles are vertices, and the lines connecting the vertices are branches. The number attached to the ring is the vertex number.

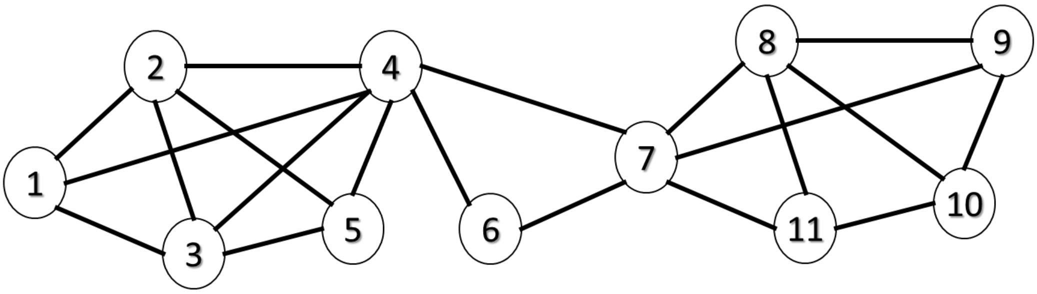


Fig. III-4-4 Example of Network

Network theory investigates various properties of such networks. For example, the shortest distance between two vertices (the minimum number of branches to connect two vertices), the diameter of the entire network, and so on.

### 3-2 Indicators of Network Theory

Some essential indicators are used in network theory, which are described below.

**[path]**

When branches connect two vertices, *i* and *j*, the sequence of branches is called a path.

For example, in Figure III-4-4, for vertices 1 and 4, the rows of branches (1, 2) and (2, 4) are paths between vertices 1 and 4. Other paths are (1, 3), (3, 5), and (5, 4). In general, there are many paths between two vertices.

**[connectivity]**

A network is connected when a path exists between any two vertices.

In short, all vertices are connected by branches. Figure III-4-4 is a connected network. If there is no branch between vertex 4 and 6, 7 in this figure, it is not connected.

**[path length]**

The number of branches in each path.

**[Distance $l_{ij}$ ]**

The minimum path length of vertex $i, j$ *is* called the distance between two vertices $i, j$ and denoted as $l_{ij}$.

**[Network Diameter]**

The maximum distance between any two vertices in the network.

Determine that the diameter is infinite when the network is not connected.

**[average shortest length $\langle l \rangle$ ]**

The average shortest length of a network is defined as the average distance of all pairs of vertices.

That is, if the total number of vertices *is* $N$, then,

$$\langle l \rangle \equiv \frac{1}{n(n-1)} \sum_{i,j} l_{ij} . \qquad \text{(III-4-5)}$$

**[degree $k_i$ ]**

The degree of vertex $i$ is the number of branches coming out of $i$ and is denoted as $k_i$.

The vertices connected to vertex $i$ *are* called neighbors of $i$.

**[degree distribution $p(k)$].**

It is defined as the probability of a vertex having degree $k$ in the network.

$$p(k) \equiv \frac{\text{number of vertices having dgree } k}{N} . \qquad \text{(III-4-6)}$$

**[Average degree $\langle k \rangle$ ]**

The average of the degrees of all the nodes in the network.

$$\langle k \rangle \equiv \frac{1}{N} \sum_{i=1}^{N} k_i = \sum_{k=0}^{\infty} k p(k) . \qquad \text{(III-4-7)}$$

**[Cluster coefficients $C(i)$ ]**

The cluster coefficient $C(i)$ for vertex $i$ is defined as follows.

$$C(i) \equiv \frac{e_i}{k_i(k_i-1)/2} \qquad ( k_i > 1)$$

$$\equiv 0 \qquad (k_i \leq 1) , \qquad \text{(III-4-8)}$$

Where $e_i$ is the number of branches between neighborhoods of $i$.

It is the ratio of the actual number of branches between neighbors of vertex $i$ to the maximum possible number of branches between neighbors of vertex *i*. Or it is the ratio of the number of triangles formed by the actual edges between vertex i and its neighbors and the number of triangles that would be created if branches connected all neighbors. Intuitively, it is an indicator of the extent of connection between neighbors. Socially, it is the probability that a person's friends are friends.

**[Average cluster coefficient $\langle C \rangle$ ]**

It is the average of the cluster coefficients for all vertices.

$$\langle C \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} C(i) \ . \tag{III-4-9}$$

Intuitively, it measures how closely individuals in a group are related to each other. A high average cluster coefficient means that when you meet someone new, there is a significant probability that you have a common acquaintance with that person.

### 3-3 Network Example

The following three are examples of typical networks (Figure III-4-5)

1. complete graph

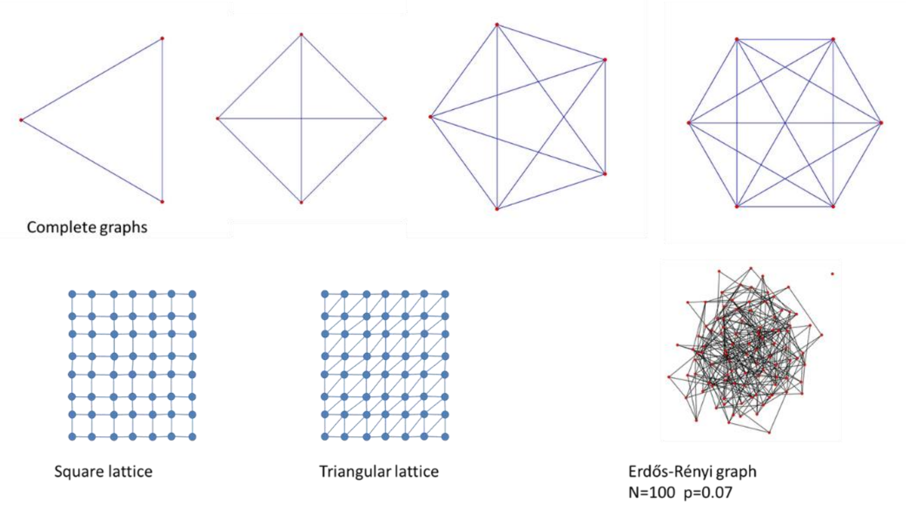
   Network with all vertices connected by branches.

2. lattice

   Vertices and branches form a lattice.

3. random graph (Erdős-Rényi graph)[67][70][71]72][73][74][75][76][77]

   Network with *N* vertices such that the probability of having a branch between any two vertices is *p*, and it is represented as $G_{n,p}$.



Complete graphs

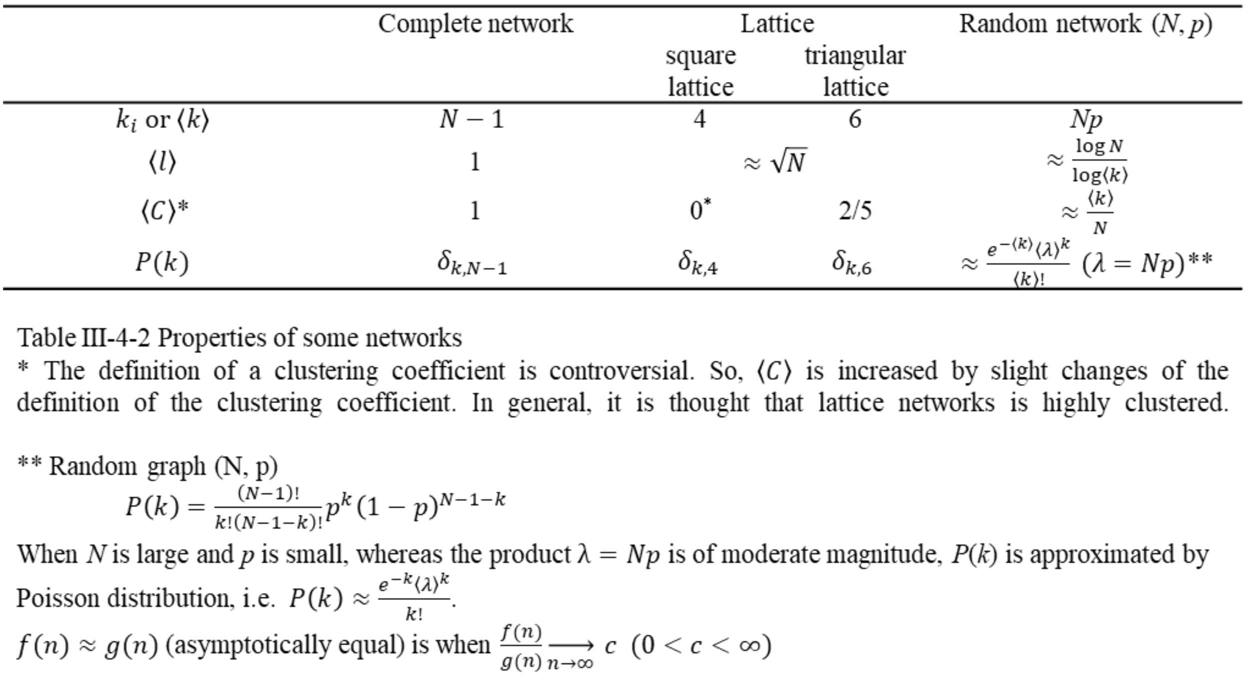
Square lattice

Triangular lattice

Erdős-Rényi graph
N=100 p=0.07

Fig. III-4-5 Examples of networks

Table III-4-2 summarizes the properties of each network. Here, $\langle l \rangle$ (average shortest length) and $\langle C \rangle$ (average cluster coefficients) of the Random graph are essential.



| | Complete network | Lattice | | Random network $(N, p)$ |
|---|---|---|---|---|
| | | square lattice | triangular lattice | |
| $k_i$ or $\langle k \rangle$ | $N-1$ | 4 | 6 | $Np$ |
| $\langle l \rangle$ | 1 | $\approx \sqrt{N}$ | | $\approx \frac{\log N}{\log \langle k \rangle}$ |
| $\langle C \rangle$* | 1 | $0^*$ | 2/5 | $\approx \frac{\langle k \rangle}{N}$ |
| $P(k)$ | $\delta_{k,N-1}$ | $\delta_{k,4}$ | $\delta_{k,6}$ | $\approx \frac{e^{-\langle k \rangle} \langle \lambda \rangle^k}{\langle k \rangle !}$ $(\lambda = Np)$** |

Table III-4-2 Properties of some networks

* The definition of a clustering coefficient is controversial. So, $\langle C \rangle$ is increased by slight changes of the definition of the clustering coefficient. In general, it is thought that lattice networks is highly clustered.

** Random graph (N, p)

$$P(k) = \frac{(N-1)!}{k!(N-1-k)!} p^k (1-p)^{N-1-k}$$

When $N$ is large and $p$ is small, whereas the product $\lambda = Np$ is of moderate magnitude, $P(k)$ is approximated by Poisson distribution, i.e. $P(k) \approx \frac{e^{-k} \langle \lambda \rangle^k}{k!}$.

$f(n) \approx g(n)$ (asymptotically equal) is when $\frac{f(n)}{g(n)} \xrightarrow[n \to \infty]{} c$ $(0 < c < \infty)$

Network theory first gained attention in complex systems when the concept of the small-world nature of networks and scale-free networks emerged.

**Small world network**

A network is a small world [78] if it has two properties: 1) the average shortest length increases only about $\log N$ as $N$ increases, and 2) the average cluster coefficient does not go to zero as $N$ increases. These two properties are defined in comparison with a random graph. Table III-4-2 shows that the average cluster coefficient becomes zero in the Random graph s $N$ increases. Also, the average shortest length between vertices increases only about $\log N$ with increasing $N$. The term "small world network" means that the average shortest length is about the same as that of the Random graph, but the average cluster coefficient is more significant than that of the Random graph.

Small world[79][80][81] is based on an experiment conducted initially by Milgram and others. In the experiment, letters were relayed from a randomly selected person in the U.S. to a target person in Boston. Each person sends a letter to someone close enough to call each other by first name. The recipient sends a letter to a person who is also close to them. The experiment was conducted to see if the letters would eventually reach the target person. Then, the letter was delivered through an average of six people. This was described and called “the sixth degree of separation.” This led to the conclusion that any two people in the U.S. are connected within the sixth degree of separation. This phenomenon was called the "small world" in the sense that anyone can be connected to anyone else through only six people, a surprising result in a country with a population of over 200 million.

By the way, the logarithm of 100 million is $\log 100000000 = 8$. Though this is a little larger than the 6th order above, it can explain the phenomenon of 6th order separation if the network roughly satisfies the small world property.

**Scale-free network**

A scale-free network is a network such that the degree distribution of the network follows a power-law distribution[82][83][84]. Namely,

$$p(k) = Ak^{-\alpha} \qquad (k > k_{min}). \qquad \text{(III-4-10)}$$

As explained in Chapter 2, the power-law distribution is a distribution that follows the power-law above a particular value of $k_{min}$. Here, α is called the power-law exponent. There are many networks whose degree distributions follow the power-law. Typical examples are WWW (World Wide Web)[85], language networks (networks that connect words in a sentence. See Chapter 4-4), actor networks, etc. Several models for creating scale-free networks have been proposed; the most important one is called the preferential attachment model by Barabási and Albert [82][83][84]. This is a network version of the Yule process (rich-get-rich) described in Chapter 2. Specifically, the model is as follows.

1) Create a complete graph consisting of $m_0$ vertices.

2) Add one vertex with $m$ ($\leq m_0$) branches at each point in time. These branches are connected to $m$ different vertices. The probability of connecting to vertex $i$ is proportional to the degree $k_i$ of the vertex. That is if the probability is $\pi_i$,

$$\pi_i = \frac{k_i}{\sum_{j=1}^{N} k_j} \qquad (1 \leq i \leq N). \qquad \text{(III-4-11)}$$

After the $t$ steps in this model, we have a network with $t+m_0$ vertices and $mt$ branches. Simulation of this model produces a scale-free network. The power-law exponent in this case was $\alpha$=2.9. Theoretically, $\alpha$=3.0. Since the power-law exponent of scale-free networks in the real world is 2 to 3, the power-law exponent of this model is a little larger. However, this index can be varied by changing the model slightly, so the size of the power-law exponent is not a significant issue [80][87][88][89][90]. Importantly, preferential attachment occurs during the network growth in generating scale-free networks.

The degree distribution is a perfect power-law distribution in the Preferential attachment model. However, actual scale-free networks often deviate from the power-law distribution at large degrees. Such a distribution that deviates from the power-law distribution is the "truncated power-law distribution" described in Chapter 2. By applying some restrictions to the preferential attachment model, such a truncated power-law distribution can be obtained [18][91][92][93].

The average shortest length of a scale-free network increases only about $\log N$. This is a necessary condition for a network to be a small world. However, cluster coefficients are generally not satisfied. Several network models have been proposed that satisfy the small-world requirement for cluster coefficients [58]. An important one among them is the hierarchical model [89]. A hierarchical model is a model created by hierarchical or recursive rules. Recursive is also a construction method used to create fractals, which we will not discuss in depth here [4].

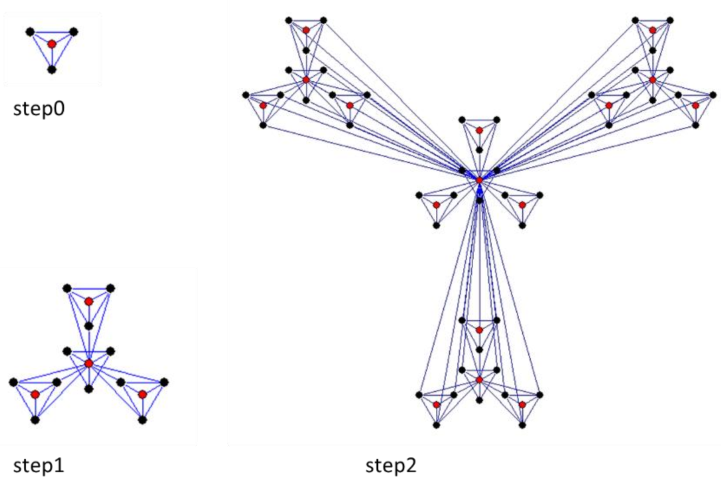
Figure III-4-6 illustrates how to create a hierarchical network using Ravasz's model. In Step 0, connect three red vertices to one vertex. Next, combine three of these step0 figures to the center of step0 (step1). This process is repeated. This network is a deterministic model, but if we include the factor of preferential selection when connecting new diagrams, we can create the preferential attachment model. The hierarchical model has the following two features in addition to being scale-free.

1) The cluster coefficient is a constant value regardless of the network size.

2) The average cluster coefficient $C(k)$ of a vertex of degree $k$ satisfies the power-law as in the following equation.

$C(k) \approx \frac{1}{k}$ (Where $\approx$ refers to the description in Table III-4-2).

Property 1) above indicates that this network is a small world. These two properties can also be used to check whether a network has a hierarchical structure. WWW, language networks, actor networks, etc., have these characteristics and are considered to have a hierarchical structure.



**図III-4-6 Creation method of the hierarchical model by Ravasz.**
**Figure in Ravasz's paper [89] modified by the author.**

All large networks in the real world, such as the World Wide Web, are scale-free and small-world in nature. Furthermore, many of them have a hierarchical structure. The concept of complex networks was created to handle these networks collectively. The exact concept of complex networks is still ambiguous, but in general, it often means networks that are scale-free and have small-world properties [66][94][95]. However, since the term is only for studying large-scale networks found in the real world, the definition itself is subject to change in the future.

I could not find a clear definition of the word "complex" in the term "complex network. However, in general, the systems subject to complex networks are usually large and complex systems. In this sense, the word "complex" in complex networks can be interpreted as meaning "complex system.

Many scale-free networks have been found in biology and medical care. Since medical knowledge networks are discussed in the next section, we describe some examples of scale-free networks in the biological and medical care fields.

In the cell, numerous chemical components undergo chemical reactions (metabolism) with each other. The network of chemical reactions among these chemical components (metabolic network) has been investigated by Jeong and others[88][90]. The results showed that metabolic networks are scale-free and small-world in nature. The same was examined for many species, which showed similar structures in those species, indicating that these properties are general. Furthermore, concerning small-worldness, the radius is around 3, even as the network size increases. This is contrary to the fact that in broad small-world networks, the radius increases according to $\log N$. This is a stronger characteristic than small-worldness. It has been stated that organisms may have kept the radius from increasing during evolution [88]. This holds for networks of medical care knowledge, which will be discussed in the next section. Ravatz also examined the metabolic networks of different species in the same way [90]. The results showed that in all species, the metabolic network consists of many modules, each of which is a small, tightly connected network. In a hierarchical structure, these modules consist of large units that lack closeness. The power-law governs the number of units and cluster coefficients. At this time, as a proof of hierarchy, they use the result $C(k) \approx \frac{1}{k}$, which is a characteristic of networks with hierarchical structures shown above.

Another important scale-free network in medical care is infectious diseases. Initially, Lilijeros et al. examined the sexual behavior of 4,781 randomly selected Swedish individuals [96]. This was done to answer the question of how sexually transmitted diseases such as HIV spread. The results showed that the number of sexual relations one person had within 12 months of the survey date satisfied the power-law. The power-law exponent at that time was approximately 2.4. In other words, the sexual behavior network proved to be scale-free. This study provided new insights into the transmission of sexually transmitted diseases such as HIV. Sexually transmitted diseases spread through sexual behavior are not equally distributed by all people to other people but rather by certain people to many people. The spread of infectious diseases through scale-free networks spreads at a different speed from random networks, in which people spread equally to others. In addition, different methods are needed to control the spread of the disease in a scale-free network than in a random network. For such infectious diseases, focused education for those who can infect many people is more effective than education for all people equally to control infection.

The properties of scale-free networks can explain the above infection control measures. Albert and others have investigated the vulnerabilities of scale-free networks [97]. A connected network loses its connectivity and breaks down somewhere when the vertices and branches within it are excluded. According to Albert and others, a scale-free network is strong against attacks on randomly selected vertices. Still, the network is frail if the attacks are selective against vertices with many branches. This is precisely in line with the strategy for prevention of sexually transmitted diseases described above.

In the new coronavirus infections that have occurred since 2019, the term "super-spreader" has become famous. This term was born from the fact that in SARS and the new coronavirus, each person does not transmit the disease equally to others, but some people infect a vast number of people, while others do not infect so many. As for sexually transmitted infections, the term super-spreader is also used in the new corona and SARS to mean a person who sends the disease to many people. This suggests that infectious diseases other than sexually transmitted diseases may also have scale-free transmission networks. Taking individual preventive measures against super-spreaders in such scale-free networks is more effective. This method is known as "cluster crushing."

## 4. medical care knowledge networks and disease distribution[98][99].

As mentioned earlier (Chapter 4-2), the name of a disease is not an objective name of the disease from the beginning. Still, it is a product of the interaction between the patient, the physician, and the environment. Within this environment, medical care knowledge is essential. Therefore, it is necessary to examine the structure of medical care knowledge. But what does medical care mean in the first place? In the case of mathematics, for example, first, there is a precise definition of a word, and then there is a clear logic based on that definition. Physics shows basically the same structure. First there is a definition of mass, and then it is organized as equations of motion governed by that definition. On the other hand, medical science or medical care is very vague. The words used are somewhat open-ended, and new words are often created. When they are no longer needed, they disappear. The definitions of many words are also ambiguous and are often defined vaguely in the context of a whole sentence. Even the names of diseases, once defined, change over time.

As an example, let us consider one of the symptoms: pain. Pain is such a common symptom that it is said that pain and fever account for most new outpatients. This concept of pain is a good illustration of the ambiguity or diversity of concepts in medical science. Pain is classically a type of perception. Pain, touch, cold, and warmth are part of perception. When an injury or other disturbance occurs, the corresponding receptors are stimulated, and the stimulus travels through the nerves to the nerve center. At the nerve center, pain is felt. This is the classical concept of pain. Descartes has already described this concept of pain. Descartes described not only physical pain but also the concept of mental pain. People with mental illnesses, such as depression, complain of pain even when no specific injury exists. He classified such pain as mental pain.

If pain can be categorized into physical and mental, then at least the coping strategies for physical pain are simple. To remove the pain, one can immediately think of eliminating the causative injury, numbing the pain receptors, blocking the nerves transmitting the pain, and so on. With such actions, it should be easy to control the pain symptoms. Analgesic and antipyretic drugs such as aspirin exert their analgesic effects by raising the pain threshold of pain receptors. Local anesthetics such as xylocaine inhibit the transmission of pain stimuli by blocking nerve transmission. Substances called opioids, such as morphine, exhibit analgesic effects by suppressing pain transmission in the spinal cord and inhibiting the excitation of pain information conduction pathways in the brain, such as the thalamus and cortical sensory areas[34].

With the development of these numerous analgesics, pain was thought to be easily controlled. However, even with the development of multiple analgesics, the number of patients suffering from pain (usually physical) has not decreased. Furthermore, several pains have been discovered that are unlikely to be caused by the classical pain mechanisms.

There is a procedure for pain relief called a cordotomy. This is a surgical procedure in which the pain-conducting pathway, called the lateral spinothalamic tract of the spinal cord, is surgically severed. This disconnection interrupts the transmission of pain signals and is supposed to provide analgesia below the disconnected area. However, many patients undergoing this surgery have experienced recurrent or increased pain.

Herpes zoster (Singles) is a disease that causes blisters and pain on the skin. Herpes zoster is a disease that almost always heals spontaneously without an immunodeficiency. However, pain may occur in the same area more than one month after the zoster has cured. This is a disease called postherpetic neuralgia. The zoster is believed to have healed, but the pain persists due to nerve damage during the infection.

---

[34] Opioids have many other effects besides their central effects.

Some trauma and persistent pain may occur a month or more after the trauma has healed. This disease used to be called causalgia and is now called complex regional pain syndrome (CRPS). This is also pain after the causative trauma that has completely healed. A more curious type of pain is phantom limb pain, which occurs after a limb's amputation. This is also pain for which no cause for pain can be found. These are diseases in which the cause of the pain is unclear. In each case, ordinary analgesics and nerve blocks have little effect. In some cases, sympathetic nerve blocks and antidepressants are more effective.

Thus, even regarding pain that is considered physical pain, there is pain that cannot be dealt with by conventional methods. Furthermore, there are cases in which people do not feel pain, even when there is an apparent cause of pain. One famous example is pain on the battlefield. Some soldiers do not complain of pain even though they are seriously wounded. It is also said that if a dog is repeatedly given a painful stimulus and food simultaneously, it will not show any pain behavior in response to the painful stimulus. In the battlefield example, being wounded means being able to return home from the battlefield, which is said to suppress pain. In dog experiments, the pleasure of receiving food is also said to stop pain.

How did human beings learn the sense or word "pain" in the first place? When one person complains of "pain, " is the same as the sensation of "pain" in another person? Let us follow Maruta on this matter [100]. The question is how a person acquires the ability to feel the sensation of pain in the first place or to learn the word pain. A baby falls and falls on its back. At this time, the baby experiences the sensation of pain, but learning is necessary to interpret this sensation as pain. When the baby falls, the mother may hold the baby in her arms and say, "Pain, Pain, Go away.” If this kind of voice is repeated many times, the child learns that the sensation of the fall is the sensation of "pain.” The child is thought to acquire the sensation of "pain" by repeatedly undergoing this learning for pain sensations other than falling.

Pain is a purely subjective sensory experience. Therefore, we should not be able to understand the pain of others. However, many people can feel pain the same way when one person cries out. This is because many people repeat similar experiences. When everyone is taught the word "pain" by their mothers when they fall, the feeling of pain is shared among people, and the sensation of pain is understood as a common language.

However, not all mothers respond the same way. For example, consider the following situation. Suppose another child pushes a toddler, falls, and scrapes her knee. In such a situation, the infant may feel surprised, sad, anxious, frustrated, and in pain. Different mothers may respond differently to this situation. Mother A is surprised by the injury, does her best only to treat it, and says, "It doesn't hurt anymore.” Mother B says, "You must have been so frustrated and sad when you were pushed away," and hugs her. An infant who has always received responses like mother A's may perceive all feelings of surprise, anxiety, sadness, and frustration as pain.

The following example could be considered. Suppose a toddler does something wrong and is beaten. If this happens repeatedly, the guilt of doing something bad might become a pain.

In other words, the sensation of "pain" is not innate but is learned. Learning from a shared experience, such as hitting or falling, gives us the label "pain" as the sensation we experience. However, each person knows a little differently. Therefore, for some people, "pain" is mixed with feelings of anxiety, sadness, guilt, surprise, and so on. If such a person complains of pain, it may be a sign of guilt. Using analgetic on such a person will not improve the pain.

This process defines pain in the pain clinic world as follows.

> “Pain is a symptom. Pain is a subjective perceptual experience.”

This definition, in a sense, states nothing. It is a definition that seems as if it has abandoned the idea of defining. However, it shows that it cannot be determined without saying this.

I have discussed the concept of pain at length to illustrate the ambiguity of the idea in the world of medical science and medical care. Unlike mathematics and physics, concepts in medical science and medical care are rarely used in a strictly defined manner. The early classical concept of pain and the modern concept of pain, which includes mental problems, are quite different. In mathematics and physics, a new term is usually coined when a concept undergoes a significant change like this. It is possible to think of creating a new concept of pain in the form of old and modern pain. However, this is not usually done in medical science and medical care. This is because concepts in medical science and medical care change daily. Nor can medical scientists alone change the idea. It is the patient who complains of pain. It is no use telling a patient who comes to the hospital complaining of "pain" that it is not pain. The world of medical science and medical care is a discipline in constant contact with the real world of patients. Changing the word "pain" at will is impossible for academic convenience.

How, then, should we recognize this ever-changing world of concepts? The various concepts used in medical science and medical care cannot be defined in a mathematical sense. A concept (word) is determined by the terms related to that concept. The classical concept of "pain" is defined by injury, receptor, stimulus, nerve, and transmission. However, the modern concept of "pain" is defined by words such as mind, anxiety, emotion, sympathy, guilt, antidepressant, and society, in addition to the above terms.

The idea that a collection of words defines certain concepts and terms is not an idea unique to medical science. Human speech, the natural language system itself, is imbued with such properties[101][102]. Thus, Medical care knowledge is considered a system described by natural language.

How can we examine knowledge systems that are not systematically assembled? Recently, methods have been developed to analyze the structure of such natural language descriptions. Some criterion connects the words used in the system to form a network. The network structure is then analyzed using network theory methods. This analysis has shown that the small-world and scale-free characteristics also appear in language [103][104][105][106][107][108][109]. Using these methods, it is possible to analyze medical science systems. The methods of natural language analysis are used for descriptions of medical science systems, e.g., medical textbooks.

The authors used Harrison's "Internal Medicine," a standard textbook of internal medicine, as a representative description of medical knowledge [110]. First, the words in the text were organized by hand, and a term was defined as a single word or a group of words with a medical science meaning [98]. The reason for not dealing with the words themselves is not to know the structure of the language but to know the structure of medical science knowledge. Each term consists of one or several words. Each term is classified into disease (diagnostic) name (d), subjective symptoms (s), objective signs (f), and other medical science terminology (x). For example, stomach cancer is a term classified as a disease name consisting of two words. The sentences also contain articles, conjunctions, verbs, adverbs, adjectives, etc., which are not included in terms, but these words were excluded from the analysis. Each term extracted from a sentence in this way is a vertex (node) of the network. Each term in a sentence is connected to each other by branches. In this way, it is possible to construct a network by connecting terms in a sentence with each other by branches (Figure III-4-7). This network will be called MKN (the medical (care) knowledge network).

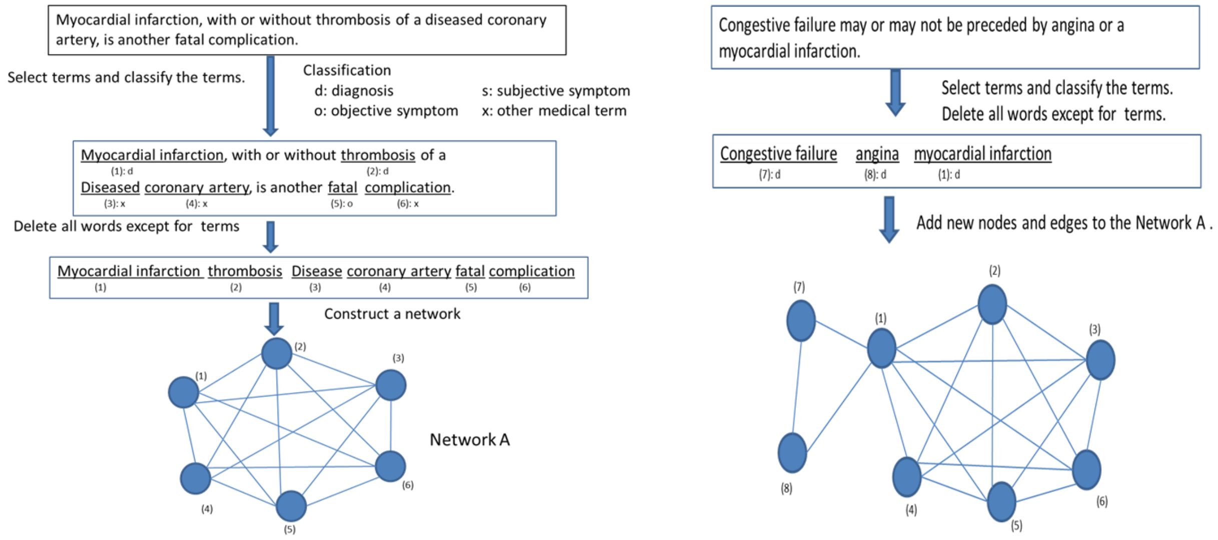


**Fig. III-4-7 Construction of the MKN**
**Reference[99]**

Table III-4-3 summarizes the network analysis. The average degree in the MKN network is 37, the average distance is 4.317, and the cluster coefficient is 0.86. The short average distance and large cluster coefficients are characteristic of small-world networks. The corresponding random graph with the same number of vertices and branches as the MKN has an average distance of 3.07 and a cluster coefficient of about 0.0008. This indicates that the constructed MKN has an average distance almost as large as the corresponding random graph and a cluster coefficient much larger than the random graph. In other words, the MKNs were shown to be small-world.

| *network* | *nodes* | *edges* | $\langle k \rangle$ | $\langle C \rangle$ | $\langle C \rangle_{random}$ | $\langle l \rangle$ | $\langle l \rangle_{random}$ |
|---|---|---|---|---|---|---|---|
| MKN | 47769 | 884613 | 37.0 | 0.86 | $7.75 \times 10^{-4}$ | 4.317 | 3.07 |
| DDN | 44997 | 505565 | 22.5 | 0.83 | $4.99 \times 10^{-4}$ | 3.894 | 3.44 |

**Table III-4-3 Profiles of the MKN and the DDN.**
**Ref. [98][99]**

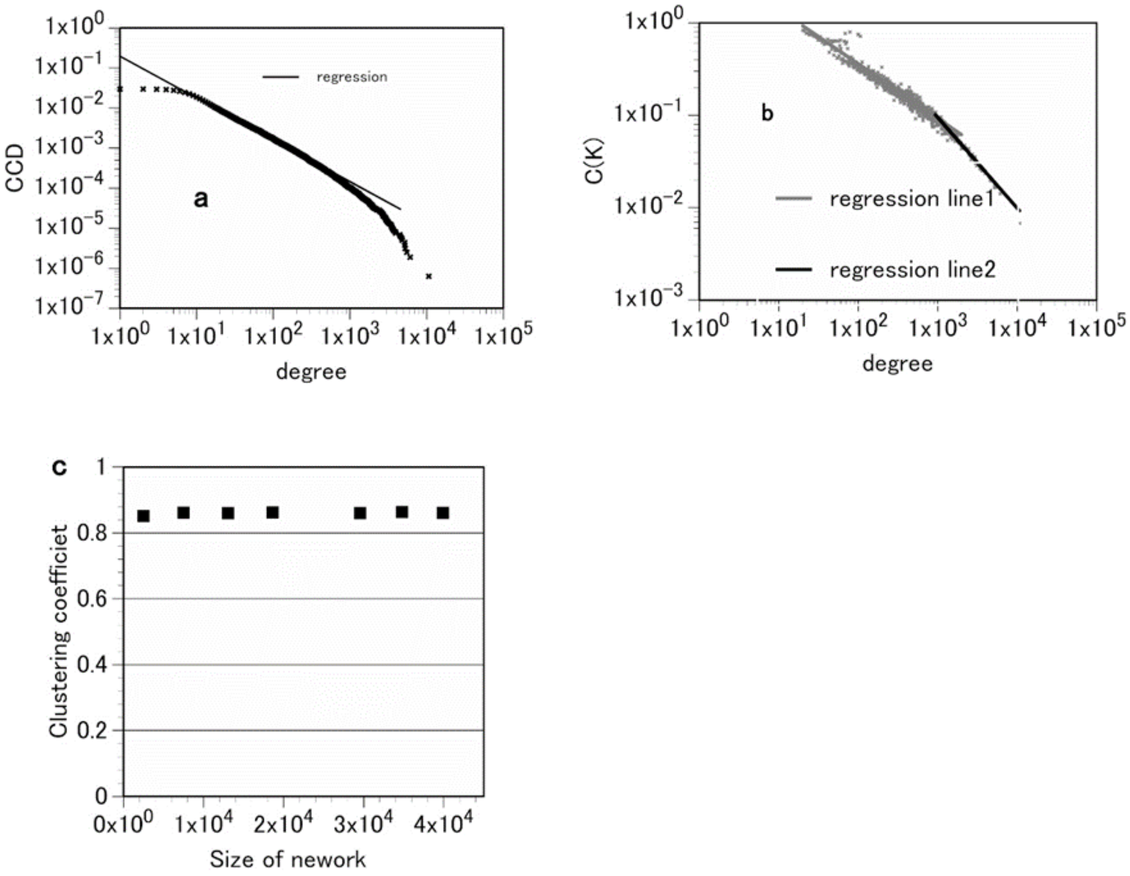


**Fig. III-4-8 Attributes of the MKN**
**a: complementary cumulative distribution (CCD) of degrees of the MKN.**
**b: average clustering coefficient of nodes with k edges (C(*k*)).**
**c; The average clustering coefficient vs. the size of the network.**
**Ref. [98][99]**

Figure III-4-8 shows the complementary cumulative distribution of degrees (CCD) for MKN, the distribution of cluster coefficients $C(k)$, and the average cluster coefficients for different network sizes. As can be seen from the figure, the degree distribution for MKN exhibits the scale-free characteristic of a power-law distribution (or, more precisely, a truncated power-law distribution). Its power-law exponent is 2.045[35]. Compared to the full power-law distribution, it deviates from a straight line where the degree is large, but this may be due to the finiteness limitation of the network, as described in Chapter 4-2. Furthermore, the hierarchical nature of the network was also examined. As mentioned in the previous section, Ravasz et al. have shown that when a network has a hierarchical structure, the following two properties of the network are valid: (1) the cluster coefficients are independent of the size of the network, and (2) the average cluster coefficient $C(k)$ of a node of degree k satisfies the power-law with exponent 1 [54][55]. This is illustrated in Figures III-4-8 b. and c. As shown in the figure, it has a hierarchical structure since MKN satisfies Ravasz's two conditions.

---

[35] (a) is the complementary cumulative distribution. From the slope of the regression line, $\beta=1.04417$. The power-law exponent of the degree distribution is $\alpha=\beta+1=2.045$ (see Appendix 4).

The above facts indicate that MKN is a small-world network with a scale-free and hierarchical structure. The small-world nature is advantageous in terms of efficiency of knowledge and information retrieval. When physicians examine patients, they select an appropriate disease name from vast knowledge. The physician must quickly choose the proper disease name from tens of thousands of disease names based on the patient's symptoms and test results. The small-world structure of medical care knowledge indicates that the appropriate disease name can be selected in as few as five steps from the patient's symptoms and test results. The small-world nature of medical care knowledge is beneficial for making such associations.

On the other hand, scale-free and hierarchical structures may be related to the mechanisms by which medical care knowledge is generated. Medical care knowledge is constantly changing. Disease names are added, definitions vary, new knowledge is added, and old knowledge is deleted. New knowledge corresponds to the addition of new vertices in MKN. The newly added vertex is connected to the already existing vertex by a branch. The preferential selection principle may be at work in the connection. In addition, medical science knowledge has a complex hierarchical structure. It is necessary to grow while maintaining the old hierarchical structure when adding a new node. These two principles will grow a network of medical care knowledge.

Is a medical knowledge network merely a knowledge network? Or does it have something to do with the actual behavior of a medical practitioner? If the medical knowledge network simply represents a medical text's structure, its use and application in actual medical care will be limited. However, the implications are significant if this network structure influences medical care behavior. To investigate this, we used a database of disease names in a hospital information system that reflects physicians' behavior. The disease name database is created by entering the names of diseases of the patients that physicians examine in their daily practice into a computer. The disease name in the database does not necessarily mean a confirmed diagnosis. The presumed disease name of a patient may change as the patient undergoes medical treatment and examination. Many different disease names are registered for the same patient in such cases. In other words, the disease name registered changes according to the physician's examination behavior. For each disease name in the database, information such as patient number, department code, disease code (ICD code), physician code, date of entry, and entry terminal code are recorded simultaneously. These data are simply the network's vertices and the Diagnosis Database Network (DDN) can be constructed by connecting the vertices in the same record with branches.

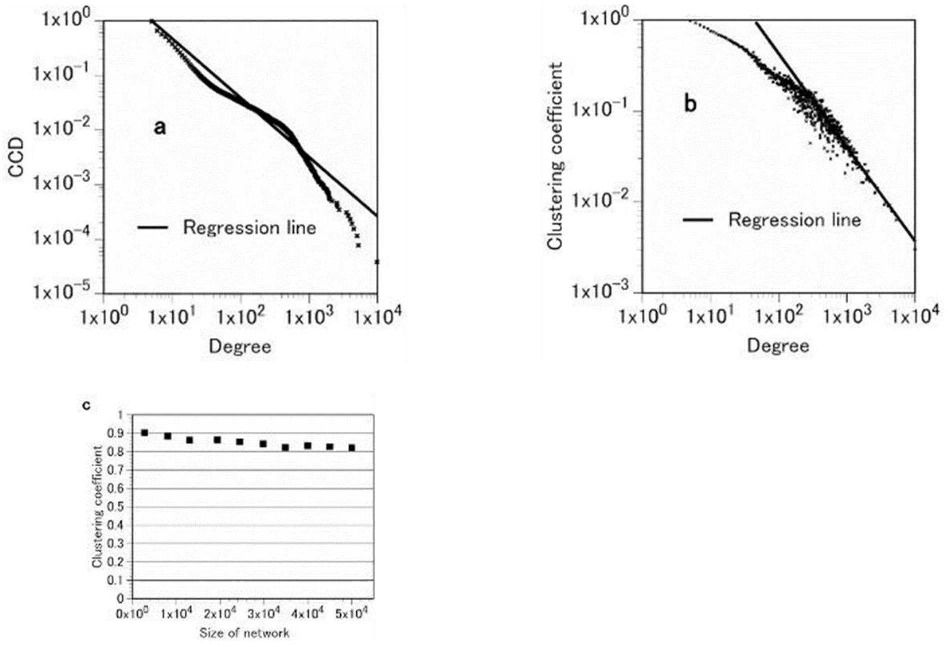


**Fig. Ⅲ-4-9 Attributes of the DDN.**
**a: The CCD distribution of the DDN. b: The average of the clustering coefficients of nodes with k edges. (C(k))**
**c: The average clustering coefficient vs. the size of network.**
**Ref. [98][99]**

As shown in Table III-4-3, DDN exhibits small-world properties. Furthermore, DDN is shown to be a scale-free and hierarchically structured network like MKN (Figure III-4-9). Moreover, MKN's average cluster coefficient and power-law exponent are 0.86 and 2.045, respectively, and those of DDN are 0.83 and 2.084, respectively. These are very close values. These facts suggest that DDN has a very similar network structure to MKN.

It is a remarkable fact that the DDN, which represents the network of the physician's act of naming a disease, and the network of medical knowledge have very similar characteristics. This fact allows us to consider the following possibilities.

1) The knowledge structure in the physician's mind to perform medical care is like that of MKN.

2) The knowledge structure is used to diagnose the disease.

3) The small-world nature of the knowledge structure allows physicians to use a vast amount of medical science knowledge.

4) The resulting disease names have a particular relationship (network-like ties) to each other.

5) The medical care practices of physicians affect the knowledge structure of general medical science.

To investigate the relationship between physician behavior and MKN, let us compare the decision-making behavior of a physician to the movement of a particle on an MKN. In MKN, consider a particle that moves randomly on the network from any vertex. Suppose that the first disease name visited by this particle is the doctor's diagnosis. If this operation is repeated many times, the frequency distribution of disease names is obtained. The frequency of disease names thus acquired would be proportional to the degree of the vertex representing the disease name. Therefore, let us examine the degree distribution of the vertices of the disease name in MKN.

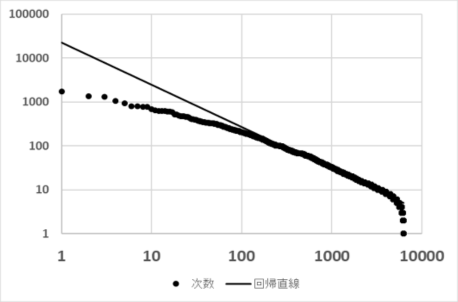


**Fig. Ⅲ-4-10 Zipf plot of degree of diagnosis nodes in MKD and the regression line.**
**Zipf exponent $\zeta$ =0.96**
**Ref.[98][99]**

Figure III-4-10 shows a Zipf plot of the degree of the disease name vertices. As shown in the figure, the Zipf plot shows that the power-law is valid. Since the Zipf exponent $\zeta$ *is* 0.96, the power-law exponent $\alpha$ is 2.04 (Appendix III-4). As Chapter 4-1 mentions, the power-law also holds for the frequency of disease names in hospitals. Moreover, its Zipf exponent is around 1 (around 2 for the power-law exponent), as shown in Table III-4-1, which is very similar to the Zipf exponent of the MKN power-law distribution. It is very suggestive that the actual frequency of disease names and the degree of MKN diagnoses follow a similar power-law. In MKN, a higher degree indicates that the term is more frequently referenced. A large degree means that the term is connected to many words, which is more likely to be used and selected. This is the reason why the present results were obtained. This is like the fact that the word frequency distribution of a sentence strongly correlates with the degree distribution of the network created from the sentence. However, in the case of sentences, it is the frequency of words in the sentences. At the same time, the present results are not the frequency of disease names in the sentences but the frequency of disease names in actual practice, which means slightly different and more meaningful results.

Why are the frequency distribution of disease names and the medical knowledge network degree distribution so similar? The possibilities are as follows.

1) Texts will always reflect new knowledge of diseases in actual medical care. In other words, diseases that appear frequently in medical care are often described in the text.
2) Diseases frequently appearing in the text are accessible for physicians to associate with.
3) The knowledge structure in the mind of the diagnosing physician has a network structure like the knowledge structure shown by the text.
4) Conversely, the text itself is constantly changing, influenced by the medical care practices of physicians.

## 5. medical care is a complex system

In Chapter 3 and this chapter, I have described many complex system-like properties that emerge in medical care. We have precisely described the power-law distributions in medical care commonly found in complex systems. As mentioned in Chapter 3, three main types of power-law distributions appear in medical care. The first is the "category frequency distribution," such as the frequency of disease names and surgical procedures. The second is a "quantity distribution," such as the distribution of blood loss during surgery. The third is "time distribution," such as length of hospital stay, operation time, and recovery time.

The latter two of these distributions were considered in Chapter 3-4. For the third length of hospital stay, random walk is the most critical factor. This is the idea that the patient's condition fluctuates randomly. The cause of the random fluctuations can be interpreted as the interaction of the various organs randomly oscillating the average state of the patient. As for the second, bleeding volume, its key factor is the fractal nature of the vessels. From a complex systems perspective, the complex interactions between tissues and cells during the development of the human body can be considered to result in fractal vascularity. In this way, the power-law distribution in the second "quantity distribution" and the third "time distribution" can be regarded as a reflection of the complexity of each organism, just like the chaos in the organism described in Chapter 3-1. On the other hand, the power-law distribution in the "category distribution," such as disease names described in this chapter, reflects the "complexity of the social structure" of medical care rather than the complexity within individual organisms.

In Chapter 4-4, we analyzed a network created from a medical science textbook. The results showed that the MKN was small-world, scale-free, and hierarchical. Furthermore, we showed that the same structure is applied to the database of disease names, which is thought to reflect actual medical care practices. Thus, the fact that MKN reflects existing medical care practices indicates the importance of examining MKN. Furthermore, this fact suggests that medical care itself is a complex system. To consider this, let us restate the definition of a complex system (see Chapter 2-1).

1) A system is composed of many elements (agents).
2) Each agent interacts locally with each other.
3) This local interaction results in the emergence of overall properties, states, and behaviors that cannot be predicted from the properties of each agent.
4) These overall properties, state, and behavior affect individual agents' behavior and change it.

Looking at this definition, we vaguely assume that biological groups, human social groups, etc., are complex systems. This is true. However, the significant difference between human society and biological groups is that humans have language. The information that controls biological groups is behavior, such as gestures and eye movements. These behaviors can only have local effects in time and space. Therefore, the biological population is precisely the group that engages in the local interactions described in 2) above. Language, however, can influence the world far away in time and space. In this sense, human societies do not only engage in local interactions. Political groups, social action groups, corporate organizations, etc., are united by language, a method (rules, laws) that has a global impact. Of course, individual members interact with each other not only through such global influences but also through local influences. Therefore, their interactions can change the group's direction and cause it to move contrary to its own rules. In this respect, it can be considered that complex systemic influences are genuinely occurring.

The global rules (language) that govern human society are like a strictly defined language, as can be seen by considering laws and the like. In other words, it is a mathematical language. As mentioned above, mathematical language is characterized by the fact that it starts from a strict definition and is based on logic deduced from it. Unlike in the past, the language that defines modern society seems to have evolved into this type. Take factory production, for example. In a factory, the product is strictly defined. Starting with the blueprints, a bill of materials that specifies even the product's contents is just such a thing. Furthermore, the production process itself is also defined. In the past, many products were produced by so-called "artisanal techniques," but now these techniques are becoming precisely manualized so anyone can make the same product. This trend is becoming stronger and is spreading to the world outside of factory production. For example, the restaurant industry has been a world of artisan craftsmanship until recently. The art of serving customers was also a skill of the artisan, a way to "win the hearts and minds" of customers. However, the fast-food industry, which spread from the U.S. to the rest of the world, introduced a process chart called a manual to this food industry. Very detailed behavior manuals have been created regarding what to say to customers when they arrive, how to recommend products, and so on. Through such manualization, the behavior of employees has become prescribed, and local, free interaction with customers is being lost.

Similar efforts are underway in medical care. The DRG/PPS (DPC in Japan) is a typical example of such an attempt. It is an attempt to classify groups of patients according to DRG codes and to apply the same medical care (called a clinical path, or path for short) to the same group. This has led to a more manual approach to medical care. The individualized treatment of each patient by each physician is gradually being lost, and the ideal is to produce the same result regardless of which physician performs the procedure. At an academic conference, a speaker said, "Nowadays, there is no need to be excited or thrilled about what we do to patients. We live in an era in which we do what we are supposed to do in a routine manner". Having a great doctor is no longer necessary, and it is becoming ideal that every doctor is the same.

These methods are progressing rapidly. However, there are still many patients who fall outside the clinical path. Hospitals refer to patients who fall outside the clinical path as "variances," and when they occur, they modify the clinical path so that it can be adapted to those patients. It is believed that patients who fall outside the path are because it is still incomplete and that with repeated modifications, all patients can eventually be treated by the path. But is this true? Perhaps the number of patients who fall outside the clinical path will not decrease, no matter how much the clinical path is modified. As discussed in Chapter 3-3, this may be a phenomenon in which outliers are not rare.

Language is originally a function of information contraction. Humans refer to all dogs collectively by the name “dog.” How many dogs are there in the world? All of them are represented by a single word, “dog.” This function allows us to think of the world simply. If language did not have this function, we would have to call each dog by a different word. The world would then be much more complex. Of course, this is not a function of language alone. It is possible that animals that do not have language also know what kind of animal they are. Pet dogs clearly distinguish between dogs and cats. They seem to think, "Dogs recognize each other as dogs, and cats as something else.” However, expressing this recognition by a single symbol and the simple fact that they recognize each other is another matter. Using language as a symbol, we can think about dogs and discuss dogs with other people even when dogs do not exist. At any rate, language has enabled humans to express and communicate a diverse world with a few symbols.

Language thus has the function of representing the world with less information. However, the world represented by language and the actual world are different. This causes problems. Even if dogs are collectively described as "dogs," sometimes it becomes necessary to distinguish one dog from another. One dog is friendly and does not bite. But other dogs are aggressive toward humans. Then, it becomes necessary to distinguish between the two. Therefore, it is necessary to create a new term to distinguish between the two. This kind of work will inevitably occur if we try to express the actual world in language. One after another, new words are created, obsolete words are deleted, and sometimes the meaning of words is changed. It is the language that continues to be created through these constant changes. In this sense, language itself can be considered a complex system. In this way, the original "language" constantly changes as it adapts to changes in the world and human beings.

However, amid all this, some languages have emerged that represent things that do not change. Mathematics is a typical example. Mathematics is a language, but its object, what it represents, does not change. Or rather, mathematics creates what does not change and expresses it. Straight lines and triangles are concepts that do not exist in the real world. The structure of mathematics begins with a definition. First, one defines a line, then one discusses and expresses its properties according to that definition. To define is to create. It is thought that the original "word" began by expressing what already exists. In the case of mathematics, however, it does not exist but is created. Since it is created, its nature does not change in the creation process. This is not possible in the case of representing something that actually exists, but it is possible in the case of creation.

Human society is not only a society that expresses itself through "words" but also a society that creates, as in mathematics. Since ancient times, we have created plants, houses, carriages, towns, and countries. For this world of creation, the language of mathematics must have been very advantageous. The mathematical world does not change the meaning of words along the way. It works according to the first definition. This trend has led to the standardization and manualization of the industry.

Now, let us return to medical care. As mentioned earlier, medical care has also become more formulaic and manualized. However, this has not always worked. I have argued in this book that the reason is the complexity of medical care. Why, then, does medical care, unlike other industries, maintain such complexity? In a word, because medical care is a system that expresses disease rather than a system that is created. In Western medical science, diseases are categorized by the name of the disease. This is the same as naming a dog "dog.” Initially, Western medical science recognized diseases by their terms and tried to deal with each disease according to its name. However, it is impossible to describe an actual disease accurately. Therefore, the same phenomenon as in the original "language" occurred in medical care. The subject always influences medical care. As mentioned in the section on pain, no matter how pain is defined in medical science, if the nature of the patient's pain no longer matches, the definition must be changed. The subject is not something man has created but exists initially. Therefore, its nature cannot be defined as in mathematics. If the subject changes, medical science itself must change accordingly. By describing these changes, medical science books are created. It is thought that the accumulation of these changes has built a medical science book, which has resulted in MKN having the properties of a complex system. The complex nature of the MKN is an expression of the complexity of medical care itself. At first glance, medical care appears to be a human-made system, like the automobile industry. However, it seems to be a complex system that cannot be tightly defined. Without recognizing this, we may be unable to solve the medical care problem. Or perhaps medical care is an inherently insoluble subject. Perhaps medical care problems are not to be solved but rather to be the subject of continuous efforts to solve them.

## Part III References

［1］ Kunihiko Kaneko, Takashi Ikegami, Ichiro Tsuda,
Evolutionary Scenarios of Complex Systems - Developmental Style of Life (Complex Systems Series, in Japanese),
Asakura Publishing Co., Ltd. 1998.
［2］ Kunihiko Kaneko, Ichiro Tsuda,
Chaotic scenarios of complex systems (Complex Systems Series, in Japanese),
Asakura Publishing Co., Ltd. 1996.
［3］ Kazuyuki Aihara,
Fundamentals and Applications of Chaotic Time Series Analysis ( in Japanese),
Sangyo Tosho Publishing Co., Ltd. 2000.
［4］ J. Feder,
“Fractals,”
Plenum Press, New York, 1988.
［5］ B. Mandelbrot,
The Fractal Geometry of Nature,
Freeman 1983.
［6］ Ayumu Yasutomi,
Complexity of Money (in Japanese),
Sobunsha Co., Ltd. 2000.
［7］ Bak P,
How Nature Works: The Science of Self-Organized Criticality,
New York: Copernicus. ISBN 0-387-94791-4, 1996.
［8］ P. Krugman,
The Self – Organizing Economy,
Wiley Blackwell 1996.
［9］ Hideki Takayasu, Misako Takayasu,
Critical Fluctuations in Economics, Information, and Life: Near Future in Complex Systems Science (in Japanese),
Diamond, Inc. 2000.
［10］ Hideki Takayasu,
Discovering Econophysics ( Kobunsha Shinsho) (in Japanese),
Kobunsha Publishing Co., Ltd. 2004.
［11］ Naoki Masuda, Norio Konno,
Complex Network: From Fundamentals to Applications (in Japanese),
Kindai Kagaku Co. 2010.

［12］ Mandelbrot BB,
The stable Paretian income distribution when the apparent exponent is near zero,
Int. Econ. Rev. (1963) 4, 111-115.
［13］ Mantegna RN, Buldyrev SV, Goldberger AL,
Linguistic features of noncoding DNA sequences,
Phys. Rev. Lett. (1994) 73, 3169-3172.
［14］ Nicolis JS, Tsuda I,
On the Parallel between Zipf's Law and 1/f Processes in Chaotic Systems Possessing Coexisting Attractors,
Prog. Theor. Phys. (1989) 82, 254-274.
［15］ Okuyama K, Takayasu M, Takayasu H,
Zipf's law in income distribution of companies,
Physica A (1999) 269, 125-131.
［16］ Sato A, Takayasu H,
Dynamic numerical models of stock market price: from microscopic determinism to macroscopic randomness,
Physica A (1998) 250, 231-252.
［17］ Zipf GK,
Human Behavior and the Principle of Least Effort,
Cambridge Mass, Addison-Wiesley, 1949.
［18］ Clauset A, Shalizi CR, Newman MEJ,
Power-law distributions in empirical data,
SIAM Rev. 51 (2009), 661-703.
［19］ Newman MEJ, Power laws,
Pareto distributions and Zipf's law,
Contemporary Physics (2005) 46, 323-351.
［20］ W. Feller,
An Introduction to Probability Theory and Its Applications Vol. 1, Vol. 2,
John Wiley & Sons 1968.
［21］ Bak P, Tang C, Wisenfeld K,
Self-organized criticality,
Phys. Rev. A. (1988)38, 364-374.
［22］ Yasuo Kawai, Osamu Suzuki,
The Folded Bar Model in Population Genetics (in Japanese),
RIMS Kokyuroku 1597(2008), 209-213.
［23］ Kei Tokita, Haruyuki Irie,
The island biogeography and the Zipf's law (in Japanese),
RIMS Kokyuroku 1409(2006), 1-6.

［24］ Manabu Takayama,
Statistical models for the length of stay in hospitals (in Japanese),
Japanese journal of insurance medicine 108 (2), 120-134, 2010-06.
［25］ Kouhei Akazawa, Takashi Yanagawa,
Survival Data Analysis: Biostatistics Series 3 (in Japanese),
Kindai Kagaku Co. 2010.
［26］ Kunihiko Kaneko,
What is Life ?, 2nd ed Introduction to Complex systems Biology (in Japanese),
University of Tokyo Press, 2009.
［27］ Hiroto Kuninaka, Naoki Kobayashi and MitSugu Matsushita,
Statistical Properties Hidden in Complex Systems − On the Basis of the Lognormal Distribution – (in Japanese),
The Physical Society of Japan (2011), 66(9), 658-667.
［28］ Sano M, Sawada Y,
Measurement of the Lyapunov Spectrum from a Chaotic Time Series,
Phys. Rev. Lett. (1985), 55(10), 1082-1085.
［29］ Wolf A, Swift JB, Swinney HL, Vastano A,
DETERMINING LYAPUNOV EXPONENTS FROM A TIME SERIES,
Physica D (1985), 14, 285-317.
［30］ Kilborn KM, Barger AC, Shannon DC, Cohen RJ, Benson H,
Assessment of autonomic function in humans by heart rate spectral analysis,
Am J Physiol. (1985) 248, H151-H153.
［31］ Goldberger AL, Rigney DR, Mietus J, Antman EM, Greenwald S,
Nonlinear dynamics in sudden cardiac death syndrome: Heartrate oscillations and bifurcations,
Experientia(1988), 44(11-12), 983-7.
［32］ Eckmann JP, Kamphorst SO,
Liapunov exponents from time series,
Phys. Rev. A. (1986), 34(6), 4971-4979.
［33］ Tsuda I, Tahara T, Iwanaga H,
CHAOTIC PULSATION IN HUMAN CAPILLARY VESSELS AND ITS DEPENDENCE ON MENTAL AND PHYSICAL CONDITIONS,
International Journal of Bifurcation and Chaos(1992), 2(2), 313-324.
［34］ Ichiro Tsuda, Takashi Tahara, Hiroaki Iwanaga ,
Capillary Chaos and its Significance in Information Processing,
Holonics, 3(1992), 163 – 176.
［35］ Takashi Tahara,

Applications of Chaos in Clinical Practice (in Japanese),
Society of Biomechanisms Japan (1992), 19(2), 105-116

［36］ Sumida T, Tahara H, Iwanage H,
A physiological significance of Shilnikov phenomenon in the focal accommodation systems of human eyes,
Int. J. Bifurcation and Chaos(1994), 4(1), 231-236.

［37］ Babloyantz A,
Evidence of Chaotic Dynamics of Brain Activity During the Sleep Cycle,
Physics Letters A(1985), 111(3), 152-156.

［38］ Report on Promotion of Database Construction and Technological Development (in Japanese),
Database Promotion Center (DPC), Inc. 1989

［39］ Report on Promotion of Database Construction and Technological Development - Medical Evaluation Database and Communication System with Semantic Information at the Core – (in Japanese),
Database Promotion Center (DPC), Inc. 1990.

［40］ Report on Promotion of Database Construction and Technological Development - Health self-management and disease prevention database Construction (in Japanese),
Database Promotion Center (DPC), Inc. 1991.

［41］ Imanishi A, Oyama-Higa M,
The Relation between Observers’ Psychophysiological Conditions and Human Errors during Monitoring Task.
2006 IEEE Conference on Systems, Man, and Cybernetics(2006), 2035-2039.

［42］ Oyama-Higa M, Miao T, Mizuno-Matsumoto Y,
Analysis of dementia in aged subjects through chaos analysis of fingertip pulse waves,
2006 IEEE Conference on Systems, Man, and Cybernetics(2006), 2386-2867.

［43］ FUJITA E, OGURA Y, OCHIAI N, MIAO T, SHIMIZU T, KAMEI T, MURATA K, UENO Y, KANEKO S,
Development of the measurement method of the prediction of sleep by finger plethysmogram data (in Japanese),
The Japanese Journal of Ergonomics 2005 41(4), 203-212.

［44］ Yamazoe M, Kushida J, Kamei K,
An After Exercise Exhilaration Index Using Acceleration of Pulse for Emotional Fitness (in Japanese),
26th Fuzzy System Symposium (Hiroshima, September 13-15, 2010)

［45］ same as [46].

［46］ Ishiyama, S., Iwanaga, H., Tahara, T., Kiyooka, Y., & Ohashi, K,
Fetus and mother are a deterministic chaos-Analysis of surrogate data method- (in Japanese),
The Society for Nursing Sci. and Eng. 5(1), 74-79(2018).

［47］ Ishiyama, S., Tahara, T., Iwanaga, H., Kiyooka, Y. & Ohashi, K,
Emotional Changes in Fetuses and Mothers over the Course of Pregnancy,
Chaos Analysis of Heart Sounds, Int. J. caring sci. 13(3), 1545-1554(2020).

［48］ Fink W, Lipatov V, Konitzer M,
Diagnoses by general practitioners: Accuracy and reliability,
International Journal of Forecasting (2009) 25, 784-793.

［49］ Tachimori Y, Tahara T, Clinical Diagnoses Following Zipf's Law,
Fractals (2002) 10, 341- 351.
［50］ Tahara T, Tachimori Y,
Considering the merits and demerits of hospital information systems and their ideal state through practice - The Ideal Medical Information System for the Future: Introduction to Chaotic and Complex Medicine (Part 1) (in Japanese),
Hospitals (2001) 60 (2) , 128-134.
［51］Tahara T, Tachimori Y,
Considering the merits and demerits of hospital information systems and their ideal state through practice - The Ideal Medical Information System for the Future: Introduction to Chaotic and Complex Medicine (Part 2) (in Japanese),
Hospitals (2001) 60 (2) , 128-134.
［52］Kawabuchi K,
The whole picture and problems of DRG/PPS (in Japanese),
Yakugyo Jiho (1997).
［53］ MORIYAMA O, ITOH H, MATSUSHITA S, MATSUSHITA M,
Long-Tailed Duration Distributions for Disability in Aged People,
J. Phys. Soc. Jpn. (2003) 72(10), 2409-2412.
［54］ Whitmore GA, Neufeldt AH,
An application of statistical models in mental health research,
Bulletin of Mathematical Biophysics (1970) 32, 563-579.
［55］ Whitmore GA,
The Inverse Gaussian Distribution as a Model of Hospital Stay,
Health Serv Res (1975) 10(3), 297–302.
［56］ M. Kouchi, Y. Tachimori, M. Hoshi,
Difference in distribution of duration of hospitalization between general hospitals, long-term care facilities and sanatorium type medical care facilities (in Japanese),
Japan Journal of Medical Informatics (2005) 25 (Suppl.), 245-248.
［57］ Barry R,
FRACTAL ANALYSIS OF THE VASCULAR TREE IN THE HUMAN RETINA,
Annu. Rev. Biomed. Eng.(2004) 6, 427–52.
［58］ Cohen R, Havlin S,
Scale-Free Networks are Ultrasmall,
Phys. Rev. Lett. 90 (2003), 058701.
［59］ Johannes HGM, James BB,
Stephen AR, FRACTAL NETWORKS EXPLAIN REGIONAL MYOCARDIAL FLOW HETEROGENEITY,
Adv Exp Med Biol. (1989) 248, 249–257.
［60］Takashi Matsuo,

Fractals structure in living body (in Japanese),
Society of Biomechanisms Japan (1995) 19(2), 125-131.
［61］ Takashi Matsuo, Kyoko Ohno-Matsui,
Fractal Analysis of Human Retinal Blood Vessels (in Japanese),
MEDICAL IMAGING TECHNOLOGY(1997) 15(5), 592-596.
［62］ Takayasu H., Nishikawa I, Tasaki H,
Power-law mass distribution of aggregation system with injection,
Phys. Rev. A (1988) 37, 3110-3117.

［63］ Rodríguez-Iturbe I, Ijjasz-Vasquez E, Bras RL, Tarboton DG,
Pwer-law distributions of mass and energy in river basins,
Water Resour. Res.(1992) 28(4), 1089-1093.
［64］ Family F, Masters BR, Patt DE,
FRACTAL PATTERN FORMATION IN HUMAN RETINA, VESSELS,
Physica D.(1989) 38, 98-103.
［65］ Statistics and Information Department, Minister's Secretariat, Ministry of Health and Welfare,
ICD-10 - Mortality and Morbidity Statistics vol 1~vol 3, 1995 (in Japanese).
［66］ Barrat A, Barthélemy M, Vespignani A,
Dynamical Processes on Complex Networks,
Cambridge University Press, Cambridge, 2008.
［67］ Bollobás B,
Random Graphs (Second Editon),
Cambridge University Press (2001).
［68］ Newman MEJ,
Networks An Introduction,
Oxford University Press, Oxford, New York, 2010.
［69］ Norio Konno, Yusuke Ide,
Introduction to Complex Graphs and Networks (in Japanese),
Kodansha Scientific LTD. 2008.
［70］ Bollobás B, Riordan O,
The diameter of a scale-free random graph,
Combinatorica, 24 (2004), 5-34.
［71］ Bollobás B,
The Diameter of Random Graph,
Trans. Amer. Math. Soc. 267 (1981b), 41-52.
［72］ Erdös P, Rényi A,
On random graphs,
Publ. Math. 6 (1959), 290-297.

［73］ Erdös P, Rényi A,
On The Evolution of Random Graphs,
Publ. Math. Inst. Hunga. Acad. Sci. 5 (1960), 17–61.
［74］ Gilbert EN,
Random Graphs,
The Annals of Mathematical Statistics (1959) 30(4), 1141—1144.
［75］ Chung F, Lu L,
The Diameter of Random Sparse Graphs,
Advances in Applied Mathematics 26 (2001), 257–279.
［76］ Klee V, Larman D,
Diameters of Random Graph,
Can. J. Math. 33(3) (1981), 618-640.
［77］ Newman MEJ,
Random graphs as models of networks,
arXiv:cond-mat/0202208［ cond-mat. stat-mech］.
［78］ Watts DJ, Strogatz SH,
Collective dynamics of ‘small-world’ networks,
Nature (1998) 393, 440-442.
［79］ Amaral LAN, Scala Barthélémy AM, Stanley HE,
Classes of small- world networks,
PNAS 97 (2000) 11149-11152.
［80］ Milgram S,
The Small-world problem,
Psychology Today (1967) 1(1), 61-67.
［81］ Travers J, Milgram S,
An Experimental Study of the small World Problem,
Sociometry (1969) 32(4), 425-443.
［82］ Barabási AL, Albert R,
Emergence of scaling in random networks,
Science(1999) 286, 509-512.
［83］ Barabási AL, Ravasz E, Vicsek T,
Deterministic scale-free networks,
Physica A (2001) 299, 559–564.
［84］ Barabási AL, Albert R, Jeong H,
Mean-field theory for scale-free random networks,
Physica A 272 (1999), 173-187.
［85］ Albert R, Jeoung H, Barabási AL,
Diameter of the World Wide Web,

Nature 401 (1999) 130-131.
［86］ Dorogovtsev SN, Goltsev AV, Mendes JF,
Pseudofractal scale-free web,
Phys. Rev. E (2002) 65, 066122.
［87］ Dorogovtsev SN, Mendes JFF,
Evolution of reference networks with aging,
Phys. Rev. E (2000) 62, 1842.

［88］ Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási AL,
The large-scale organization of metabolic networks,
Nature (2000) 407, 651-654.
［89］ Ravasz E, Barabási AL,
Hierarchical organization in complex networks,
Phys. Rev. E (2003) 67, 026112.
［90］ Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL,
Hierarchical organization of modularity in metabolic networks,
Science (2002) 297, 1551-1555.
［91］ Lee J, Park NK,
Scale-free Networks in the Presence of Constraints: An Empirical Investigation of the Airline Route Network,
Seoul Journal of Business (2007) 13(1), 3-19.
［92］ Maschberger, Kroupa TP,
Estimators for the exponent and upper limit, and goodness-of-fit tests for (truncated) power-law distributions,
Mon. Not. R. Astron. Soc. (2009) 395, 931-942.
［93］ Mossa S, Barthélémy M, Stanley HE, Amaral LAN,
Truncation of power law behavior in “scale-free” network models due to information filtering,
Phys. Rev. Lett. (2002) 88, 138701.
［94］ Lipsitz LA and Goldberger AL,
Loss of 'Complexity' and Aging Potential Applications of Fractals and Chaos Theory to Senescence,
JAMA (1992) 267, 1806-1808.
［95］ Song C, Havlin S, Makse HA,
Self-similarity of complex networks,
Nature (2005) 433, 392-395.
［96］ Liljeros F, Edling CR, Amaral LAN, Stanley HE, Åberg Y,
The web of human sexual contacts,
Nature (2001) 411, 907-908.
［97］ Albert R, Jeong H, Barabási AL,
Error and attack tolerance of complex networks,

Nature (2000) 406, 378-382.

［98］ Tachimori Y, Iwanaga H, Tahara T,
The networks from medical knowledge and clinical practice have small-world, scale-free, and hierarchical features,
Physica A 392 (2013) 6084-6089.

［99］ Tachimori Y,
Network analysis of medical knowledge to investigate the structure of medical service: Toward a mathematical approach to medical service,
Ph. D. Thesis (2015 HokkaidoUniv.).

［100］ Toshihiko Maruta,
Psychology of Pain (in Japanese),
Chuko Shinsho 1989

［101］ Mutsumi Imai,
Solving the Mysteries of Language Development (in Japanese),
Chikumashobo Ltd. 2013

［102］ Mutsumi Imai, Etsuko Haryu,
How we learn languages - from mother tongue to foreign languages,
Chikumashobo Ltd. 2014

［103］ Cancho RF, Solé RV, Köhler R,
Patterns in syntactic dependency networks,
Phys. Rev. E 69 (2004) 051915.

［104］ Cancho RF, Solé RV,
Least effort and the origins of scaling in human language,
PNAS (2003)100, 788-791.

［105］ Cancho RF, Solé RV,
The small-world of human language,
Proc. R. Soc. Lond. (2001) B 268, 2261-2266.

［106］ Markošová M,
Network model of human language,
Physica A 387 (2008) 661-666.

［107］ Motter AE, Moura APS, Lai YC, Dasgupta P,
Topology of the conceptual network of language,
Phys. Rev. E (2010) 65, 065102(R).

［108］ Solé R,
Syntax for free?,
Nature (2005) 434, 289.

［109］ Solé RV, Murtra BC, Valverde S, Steels L,
Language networks: their structure, function and evolution,

Complexity (2010) 15, 20-26.

［110］ Dennis L. Kasper et al.,
Harrison’s Principles of Internal Medicine 16th ed.
The McGraw-Hill Companies, Inc. 2005.

［111］ T. Tahara,
Development of Chaos Attractogram Processor (in Japanese),
New Medicine in Japan (1992), 19(5), 35-38.

［112］ Ichiro Tsuda,
Mathematics and Medicine - On Chaos Analysis of EEG and Pulse Rate (in Japanese),
J. Math. Soc. Japan (1999), 52(1), 91-96.

［113］ Hideyuki Uchimura, Takashi Tahara,
Present state of transmitter research on schizophrenia etiology (in Japanese),
Clinical Psychiatry 1994, 36(6), 564-570.

［114］ Hideyuki Uchimura,
Somaticity in Psychotherapy (in Japanese),
The Japanese Society for Morita Therapy 1996, 8(2), 195-199.

［115］ Takashi Tahara,
Chaotic and Complex Systems for Life Rhythm and Health (in Japanese),
Education and Medicine 2001, 49, 334-347.

［116］ Hideyuki Uchimura,
Morita therapy as dialectic-From contradiction to "as is” (in Japanese),
Japanese Journal of Psychotherapy 2013,39 (3),409-416.

［117］ No target magazine.

［118］ Takashi Tahara,
Living organisms and chaos,
Nikkei Inc. 1994 (11).

## Appendix

### 1 Differential equations as a dynamics system

A dynamical system can be considered an equation that predicts the future from the present state.

$$x_{n+1} = f(x_n) \qquad \text{( III -A -1 )}$$

is an expression that predicts the state $x_{n+1}$ at time $n+1$ from the current state $x_n$. This formula is for the case of discrete-time, but if time is continuous, how can we predict the future? If time is continuous, there is no such thing as the next time after a specific time. This is where the idea of differential equations comes in, to be used to predict the future.

First, we will use $t$ instead of *n to* represent time and $x(t)$ and $y(t)$ instead of $x_n$ and $y_n$[36]. Consider predicting the future with the current time as *t.* Since $t$ is a real number, the next does not exist as the "next" future. If $h$ is a small real number, the time after $h$ is expressed as $t+h$. Here, $h$ can be any small number. Let us represent the future state $x(t+h)$ using the present state $x(t)$. To describe the future, in other words, we can define the difference between the future state $x(t+h)$ and $x(t)$.

Namely, $x(t+h) - x(t) = f\big(x(t)\big)$ . ( III-A-2 )

Is it possible to have a representation of the right-hand side that has nothing to do with $h$ for any h? Since this $h$ can be anything, it does not matter how small $h$ is. So, let's make $h$ on the left-hand side of ( III-A-2 ) smaller and smaller. As you can easily imagine, as $h$ is smaller, $x(t+h)$ will naturally become closer to $x(t)$. This means that the left-hand side of (III-A-2) approaches zero. For sufficiently small *h,* the left-hand side is almost zero. Since we are trying to express the right-hand side of (III-A-2) in terms of the current value of *x,* it does not matter what $h$ is. This means that the value of $h$ must be the same as that of $h$ when $h$ is sufficiently small. Since the left-hand side of (III-A-2) is almost zero for sufficiently small $h$, $f(x(t))$ of the right-hand side must also be zero. This means that the future cannot be predicted using expressions like (III-A-2) (this is the significant difference from the discrete case).

So what is to be done? An old genius found the answer. He wondered what would happen if the left side of (III-A-2) was divided by *h.* Dividing by $h$ means that if $x(t)$ represents the position at time *t*, then the value divided by $h$

$$\frac{x(t+h)-x(t)}{h}\ , \qquad \text{(III-A-3)}$$

is the velocity (or the average velocity between time t *and* $t+h$). Unlike (III-A-2), this value can be a constant non-zero value even when *h is* small (zero means the object is stationary). The idea is that if it becomes a constant value, then this equation can predict the future.

A differential equation is an expression like the following.

---

[36] *n* (*i,j,*k, *etc.*) is generally used to represent integers and natural numbers, while *t* (*x,y,*z, *etc.*) is used to represent real numbers. The same is true for $x_n$ and $x(n)$. When writing $x_n$ , there is usually an implicit understanding that *n is an* integer (as in a number sequence). $x(t)$ usually implies that *t is* a real number. In any case, since these are customs, it is not wrong to write either of them, but it is usually easier to make a mistake if you follow the customs.

$$\frac{dx}{dt} = ax + b \,. \qquad \text{(III-A-4)}$$

In the case of two variables, the next is true.

$$\begin{cases} \frac{dx}{dt} = ax + by \\ \frac{dy}{dt} = cx + dy \end{cases} \quad . \qquad \text{(III-A-5)}$$

Where *a*, *b*, *c*, *d* are ordinary numbers (constants). However, in general, the right side can be any expression (function) for *x* or *y* (e.g., $x^2$, sin(*x*), etc.). So, in general

$$\frac{dx}{dt} = f(x). \qquad \text{(III-A-6)}$$

Also, *x* is precisely written *x*(*t*), and y *is y*(*t*), but since the equation is dirty, we usually omit t and write *x*.

Here, the definition of the left side is written as follows[37].

$$\frac{dx}{dt} = \lim_{h \to 0} \frac{x(t+h) - x(t)}{h} \quad .$$

Here, “lim” means limit. In the function $\frac{x(t+h)-x(t)}{h}$, $\frac{dx}{dt}$ is the value when a small value of *h* is brought closer and closer to zero. Initially, the limit is taken, but if *h* is fixed here (“lim” is eliminated), then, the right side is $\frac{x(t+h)-x(t)}{h}$.

If *h* is small enough, fixing *h* will not make much difference from the limit value. So, for example, *h*=0.01 (second) is determined. Then, equation (III-A-6) becomes as follows.

$$\frac{x(t+h)-x(t)}{h} = f(x) \quad h = 0.01 \quad . \qquad \text{(III-A-7)}$$

Once replaced this way, the differential equation becomes essentially the same as the discrete equation (III-A-1).

Let's transform equation (III -A-7) a little. Multiplying both sides by *h*

$$x(t+h) - x(t) = hf(x).$$

Further,

$$x(t+h) = x(t) + hf(x) \quad . \qquad \text{(III-A-8)}$$

This equation (III-A-8) is almost the same as the equation ( III-A-1 ) if *n* is considered *t*, and *n*+1 is *t*+*h*.

The equation ( III-A-8) predicts *x* at the future *t*+*h*, with the current value of *x* (*x*(*t*)). This is the same as (III-A-1). Furthermore, by repeating this process, the value far in the future can also be determined.

Consider the following concrete equation.

$$\frac{dx}{dt} = 2x. \qquad \text{(III-A-9)}$$

At the first time (*t*=0), let *x*=1. At this time, consider the value of x after 1 second. Let us assume that *h*=0.01. Then,

---

[37] The "d" in "dx" is the symbol for difference, not d multiplied by *x*. The difference is the difference between *x* at *t*+*h* and *x* at *t*, *x*(*t*+*h*) -*x*(*t*). It is a bit different strictly speaking, but this understanding is sufficient. *dt* is strange, but it means *t*+*h*-*t*, *which* is *h* after all.

the equation ( III-A-9) can be transformed as in equation (III-A-8) to give

$$x(t+h) = x(t) + h2x.$$

Putting in concrete numbers (t=0, $h=0.01$, $x(0)=1$),

$$x(0.01) = x(0) + 0.01 \times 2 \times x(0) = 1 + 0.02 = 1.02 \ ,$$

The same is true after 0.02 seconds. Similarly, after 0.02 seconds

$$x(0.02) = x(0.01) + 0.01 \times 2 \times x(0.01) = 1.02 + 0.01 \times 2 \times 1.02 = 1.0404 \ .$$

The same will apply below.

$$x(0.03) = x(0.02) + 0.01 \times 2 \times x(0.02) = 1.0404 + 0.01 \times 2 \times 1.0404 = 1.061208 \ ,$$

.....

$$x(1) = x(0.99) + 0.01 \times 2 \times x(0.99) \ ,$$

$$= 7.10259423358087 + 0.01 \times 2 \times 7.10259423358087 = 7.24464611825249.$$

This means that the value of $x$ after 1 second can be determined. In the same way, 1000 calculations will give the value after 10 seconds, and 10000 calculations will provide the value after 100 seconds. This is a complicated calculation by hand, but it can be done quickly on a computer.

If we apply this, we can solve any differential equation. Strictly speaking, there are a few problems, but it means that middle school students can solve differential equations[38].

There are a couple of issues here. The first one is: "I decided to set $h$ to 0.01 at random, but I am not allowed to do that". It is not good. If it were okay to do anything, there would not be a limit or anything like that in the original definition. If we decide $h$ to be a concrete number instead of an extreme limit, the solution is slightly different than if we took the limit. To obtain a perfect solution, $h$ must be made smaller and smaller; as $h$ is made smaller and smaller, it will gradually approach the proper solution. Therefore, if $h$ is kept small to some extent, the solution will be very close to the real one, and there will be no problem in practical use. Then there is the question of what specific value of $h$ should be used, but there is no answer. Empirically, if $h$ is reduced little by little and the solution does not change in practical use, it is considered good enough[39].

By the way, differential equations are usually classified as a part of higher mathematics. You may think that junior high school students can solve differential equations. Therefore, I will briefly explain what mathematics is doing. First, consider the case where the previous example is solved mathematically. That is, the following differential equation.

$$\frac{dx}{dt} = 2x.$$

In mathematics, they do not want a specific number, but rather $x$ as a function of $t$. The solution to this differential equation can be written in terms of a function as follows.

$x(t) = Ae^{2t}$ $\qquad$ $A$ is an arbitrary constant.

---

[38] It may not be solvable depending on the shape of the right-hand side, but it is usually solvable.

[39] The actual computer calculation method is slightly different from the method shown here. The method shown here is almost correct when $h$ is very small, but when $h$ is a little larger, the error becomes large. Therefore, computer calculations are devised to reduce the error (Runge-Kutter method).

In mathematics, they are trying to find a functional form and prove that there is no other answer.

In addition, they also theoretically study questions such as "Do differential equations have answers in the first place? If so, how many are there, and are the answers uniquely determined?” In fact, because of such theoretical research, the correctness of numerical calculations such as the one we are doing here is guaranteed. Such mathematical studies are challenging to understand. However, the primary or original idea of differential equations is relatively simple. What I want you to know is the basic idea.

Now, a change-equation like (III-A-1) or a differential equation like (III-A-4) or (III-A-5) is called a dynamical system. Initially, the term dynamical system comes from Newton's mechanics. Newton was the first to use differential equations to describe the motion of particles and celestial bodies. Therefore, we call such a type of differential equation a dynamical equation system. The fact that the solution of a differential equation is unique gave rise to determinism, which states that if the present state is known, the future can all be known. Descartes advocated elemental reductionism, but there is no idea of determinism there. Determinism is an idea derived from Newton's equations of motion. Dynamical systems were initially used in the dynamics of physics. Still, they are now used in economics, biology (mathematical ecology), and many other fields besides physics and have become the most powerful method for describing (studying) objects.

## 2 Phase space and attractors

### 2-1. phase space

Given a dynamical system consisting of multiple variables, the phase space is the space in which the various variables are the coordinate axes. This dynamical system's motion (change) can be expressed as an orbit in this space. The trajectory described in this phase space is sometimes called a phase portrait.

In mechanics, the motion of a particle moving in a straight line is described by a dynamical system (differential equation) related to the position and velocity of the particle. Therefore, the phase space of this motion is a plane with position and velocity as two coordinates. For example, the following equation expresses the movement of a particle pulled by a spring.

$$\begin{cases} \dfrac{dx}{dt} = y \\ \dfrac{dy}{dt} = -x \end{cases} , \qquad \text{(III-Attachment-10)}$$

where $x$ is the position and $y$ is *the* velocity of the particle. In solving this equation, the coordinates $x$ and the velocity $y$ are expressed as a function of time $t$. That is, $x=sin(t)$ and $y=cos(t)$ (Figure III-A-1). The phase space shows the x-coordinate position and the y-coordinate velocity at each point in time (Figure III-A-1B).

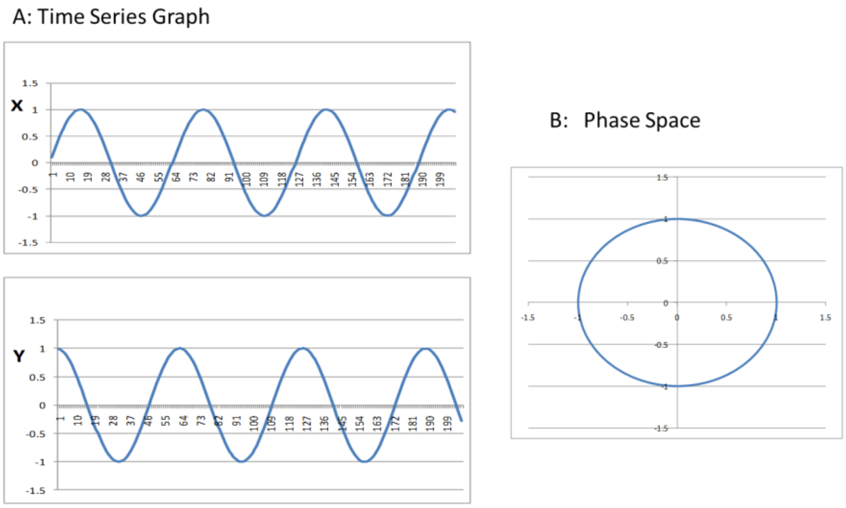
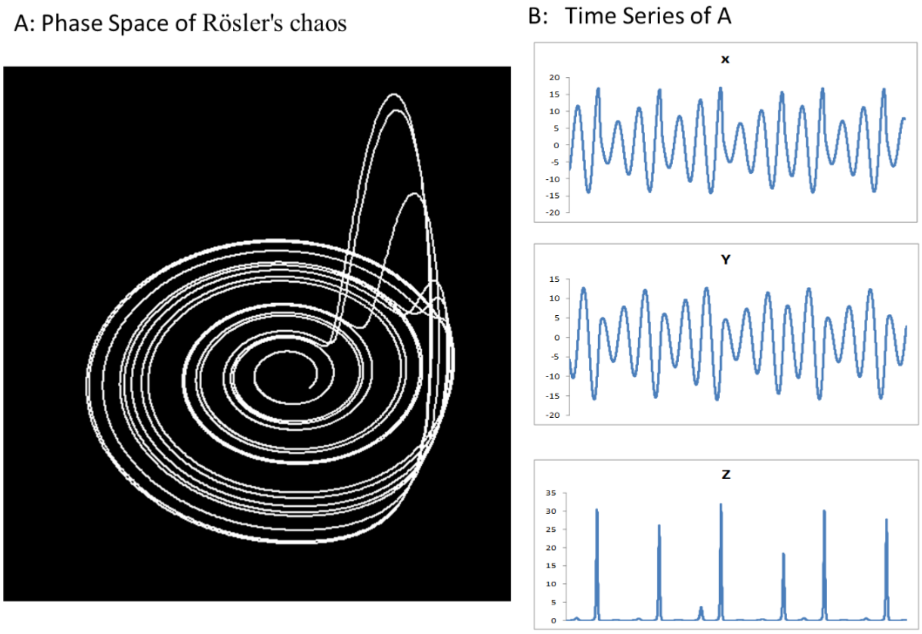


Fig. III-A-1

Suppose the equation representing the dynamical system is expressed in two variables. In that case, it becomes a plane diagram (two-dimensional phase space), and if it is expressed in three variables, it becomes a three-dimensional diagram (three-dimensional phase space). Since the number of variables can be any number depending on the subject, the phase space also has four or five dimensions, as many as the number of variables. The phase space dimensions are called the degrees of freedom of the dynamical system.



A: Phase Space of Rösler's chaos

B: Time Series of A

Fig. III-A-2

Figure III-A-2 shows the dynamical system, which is called Rösler's chaos. *A* is a trajectory in phase space, and *B* is a time series graph. The time series graph of each variable is very complicated and does not seem to have any regularity. Still, a beautiful figure emerges when it is represented in phase space, as shown in *A*. This is an example of how a change in representation can reveal regularity in something that appears to have no regularity at first glance. Generally, a very beautiful figure emerges when a phenomenon called chaos is represented in phase space.

## 2-2. Attractor

As the text mentions, a linear dynamical system has a single final stable state. Figure III-A-3 shows a two-dimensional linear dynamical system (differential equation) solution in phase space.

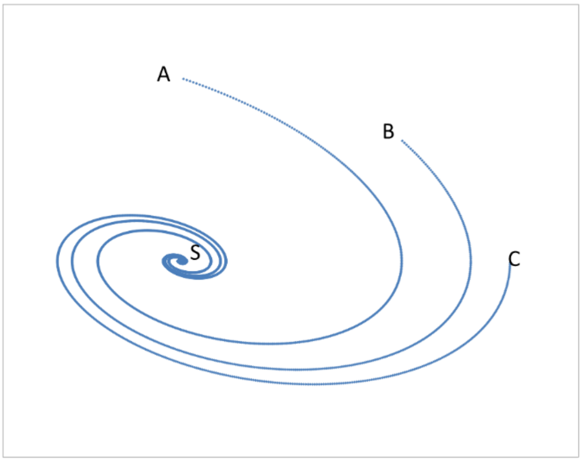


**Fig. III-A-3**

The figure depicts three trajectories starting from three different starting points (A, B, and C in the figure), and although the three points are other, they eventually approach point *S* as time passes. In fact, in this dynamical system, the final state is the point *S,* no matter where one starts from in the plane. All points approach this point, so this point *S* can be considered to attract other points. In this sense, such a point (generally, a figure consisting of a collection of points) is called an attractor.

In a linear dynamical system, the attractor consists of a single point. That is, the final state of a linear dynamical system is a single fixed point. However, in nonlinear dynamical systems, it is different.

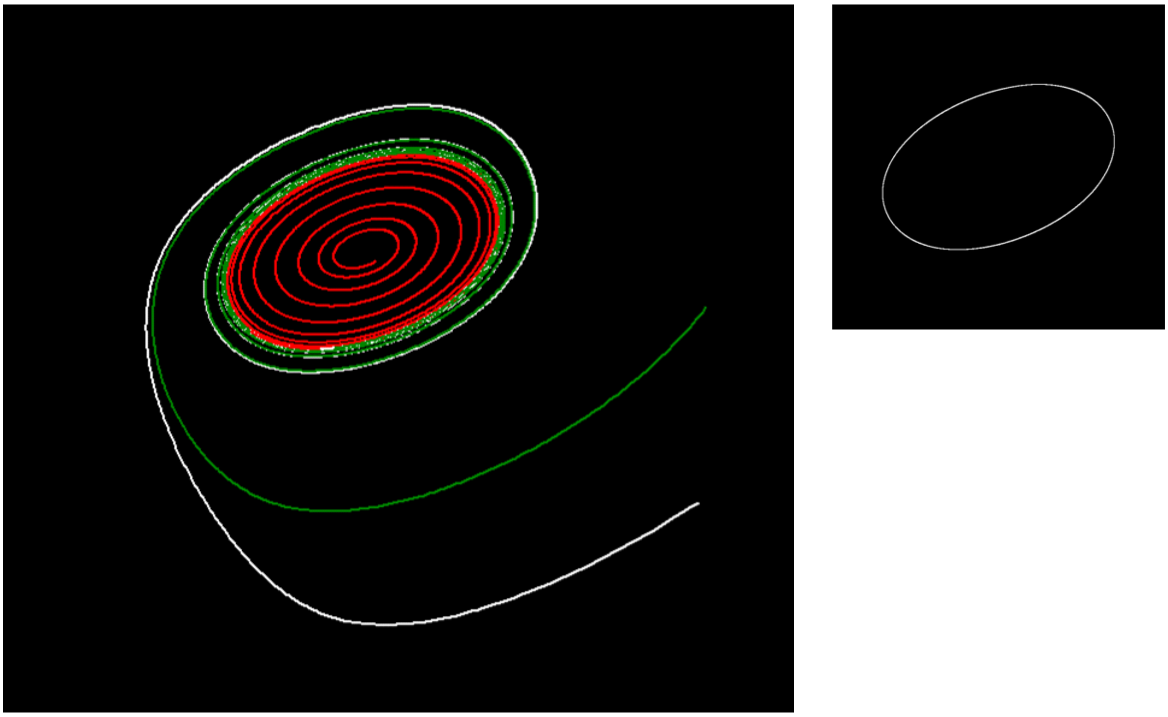


**Fig. III-A-4**

Please see Figure III-A-4. This figure is a two-dimensional phase space of a particular nonlinear dynamical system. In the figure, white, red, and green curves can be seen. Each of these three curves starts from a different initial value. However, you can see that eventually, all curves wind up into an ellipse-like curve (only that part is illustrated on the right side). The white and green curves wrap around from the outside, while the red curve wraps around from the inside. Eventually, it will continue to wrap permanently around the ellipse-like curve. This ellipse-like curve (closed curve because it is closed) is called a limit cycle. Thus, in a nonlinear dynamical system, the final state is not a fixed point but an ever-moving state that continues indefinitely in periodic motion. This limit cycle is an attractor because it attracts surrounding points in the same way as before.

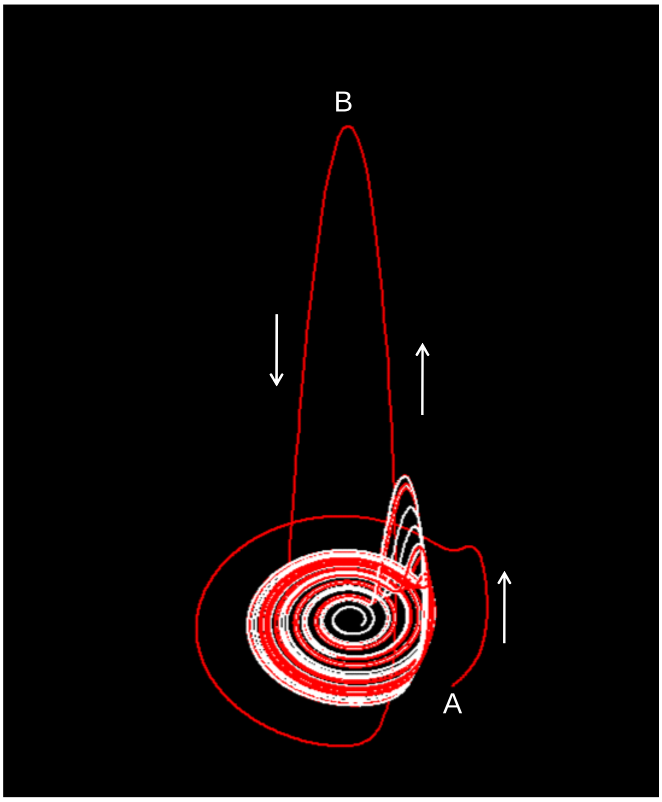


**Fig. III-A-5**

Figure III-A-5 is the Rösler chaos described above. This is a nonlinear dynamical system in three dimensions. Figure III-A-2 (A) showed only the middle clump (white curve in the figure) in this figure, but Figure III-A-5 shows what happens when starting from point A, which is far from this clump. Notice the red curve in the figure. As you can see in the figure, the point starting from A goes around the clump and once moves away from it (point B), but then changes direction rapidly and is finally absorbed by the clump. After being absorbed, it continues to move permanently within this mass. This clump, like the previous limit cycle, attracts surrounding points. In other words, this lump is also an attractor. However, this attractor has a very complicated and strange shape compared to a linear point or the previous limit cycle. For this reason, this strangely shaped attractor is called a strange attractor. Strange attractors are a characteristic of chaos.

## 3 Logistic Map

A logistic map is a dynamical system expressed by the following equation.

$$x_{n+1} = ax_n(1 - x_n) \ . \qquad \text{(III-Attachment-11)}$$

The properties of the solution of this equation change drastically with the value of *a.* And chaos occurs at 3.569945...<a<4. Figure III-A-6 is the graph when $a = 3.8$. The dotted line *A* is the graph for the initial condition $x_1 = 0.1,$ and solid line B is the graph for the initial condition $x_1 = 0.101$. The two curves are almost identical at first, but after n=10, the two curves begin to deviate from each other, and after n=15, the two curves behave as if they are two different curves. This is the sensitivity to initial conditions of chaos. Even if the initial values are almost the same (0.1 and 0.101), the two motions become far apart as time goes by.

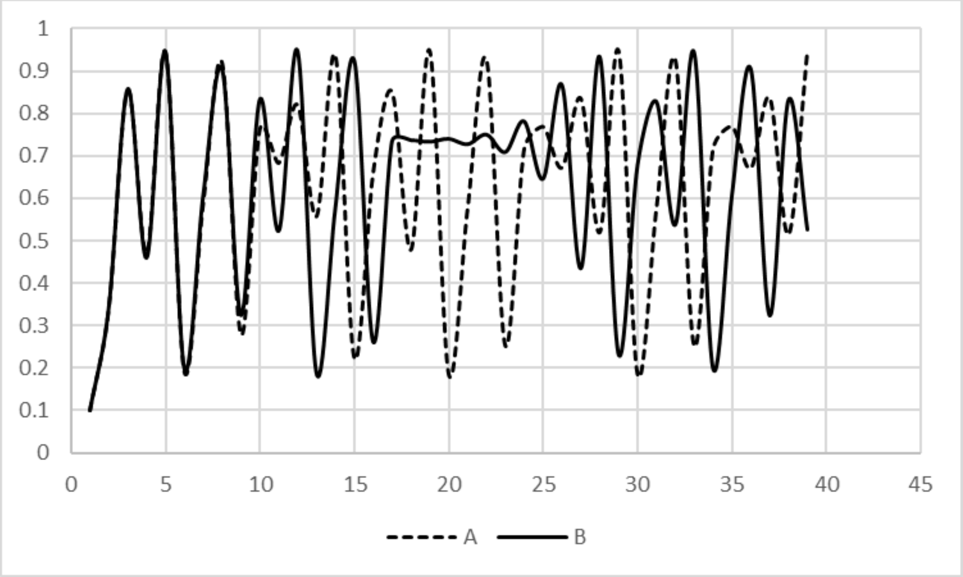


**Fig. III-A-6**

In addition, $y_{n+1} = 1 - ry_n^2$ is also called a logistic map and shows chaos as in the above equation. In this case, if $r = \frac{a(a-2)}{4}, x = 4(y-2)/(a-2)$ in the above equation, they are the same as each other. In this case, however, chaos occurs at 1.4011… <*r*<2.

## 4 Three expressions for frequency distribution

### 4-1 Three Expressions of Distribution

### A. Size-Frequency plot (Figure III-A-7A )

This is a commonly used distribution chart. The size (measured size) is written on the horizontal axis, and the frequency is on the vertical axis. It is the same as a histogram. It is the same as the probability density function $p(x)$. In the case of word frequency within a sentence, the number of words with the same frequency of occurrence is plotted against frequency. For the frequency distribution of disease names, the frequency of occurrence is plotted on the horizontal axis, and the number of disease names with that frequency is plotted on the vertical axis.

In the case of a power-law distribution, the frequency distribution $f(x)$ of objects with size $x$ is expressed as

$$f(x) = \frac{B}{x^{\alpha}} \quad (x > x_{min}) \quad , \qquad \text{(III-A-12)}$$

where $B$ and $\alpha$ are constants. If the total number of targets is T, $p(x) = f(x)/T$. This $\alpha$ is called the "power-law exponent.”

### B. Rank-Size plot (Figure III-4-7B )

This is also called a Zipf plot. Now, the data are ordered. The criteria for ordering can be size, weight, frequency, or anything else. In this case, a graph with the order (rank) on the horizontal axis and the criteria (size, weight, frequency, etc.) on the vertical axis is called a Rank-Size plot. There are many kinds of order: words in order of frequency of disease names, rocks in order of their size when smashed, nodes in order of the number of edges in the graph (network), and so on.

However, in the case of the frequency of words or disease names, the criteria are the frequency (the frequency of specific disease names is the criteria for ordering), so we call it a rank-frequency plot.

For a power-law distributed object, the size of the *nth* object, $S_n$, is expressed as

$$S_n = \frac{A}{n^{\zeta}} \quad , \qquad \text{(III-A-13)}$$

where $A$ and $\zeta$ are constants. In particular, we refer to $\zeta$ as the Zipf exponent.

### C. Complementary Cumulative Distribution Function (CCDF) and Complementary cumulative frequency plot (Figure III-A-7C)

The complementary cumulative distribution function (CCDF) of the random variable[40] $X$ is defined as follows.

$$P(x) = \Pr\,(X \geq x) \ .$$

In ordinary language, it means the probability that the value of an event is greater than or equal to a specific value $x$.

P(x) is expressed as follows in the power-law distribution case.

---

[40] By a random variable, we mean a variable that takes a value that is assigned to a possible occurrence with respect to a certain event. For example, in the case of dice, there are six possible values: 1, 2, ..., 6. Let X be the variable that takes these six possible values, Pr(X=6) means the probability of taking the value 6 (1/6), and Pr(X≥4) means the probability of taking 4 or more (4, 5, 6) (1/2).

$$P(x) = \Pr(X \geq x) = \int_x^{\infty} p(t)dt = \frac{D}{x^{\beta}} \quad , \tag{III-A-14}$$

where D and $\beta$ *are* constants, $\beta = \alpha - 1$ relationship between α and β.

Furthermore, there is the following relationship between α *and ζ,*

$$\alpha = \frac{\zeta+1}{\zeta} \quad . \tag{III-A-15}$$

Thus, no matter which data representation is expected to have a power-law distribution, the power-law exponent *α* can be obtained.

**2. Evaluation of the power-law exponent**

As shown above, a set of data can be represented by three types of plots. If the data set follows a power-law distribution, then the three types of plots above, displayed on both logarithmic scales, will be approximately linear. Then, we can estimate the exponent *α* from the slope. Any of the three plots can be used to estimate *α*, but for practical purposes, the CCDF provides the most accurate estimate.

Figure III-A-7 shows three types of plots drawn from the same data. Table III-A-1 shows the regression analysis of these three plots. These results show that the slope and *α* obtained from the size-frequency plot have a more considerable variance than those estimated from the other plots. Newman states that the *α* obtained from the size-frequency plot is imprecise compared to the other plots [49].

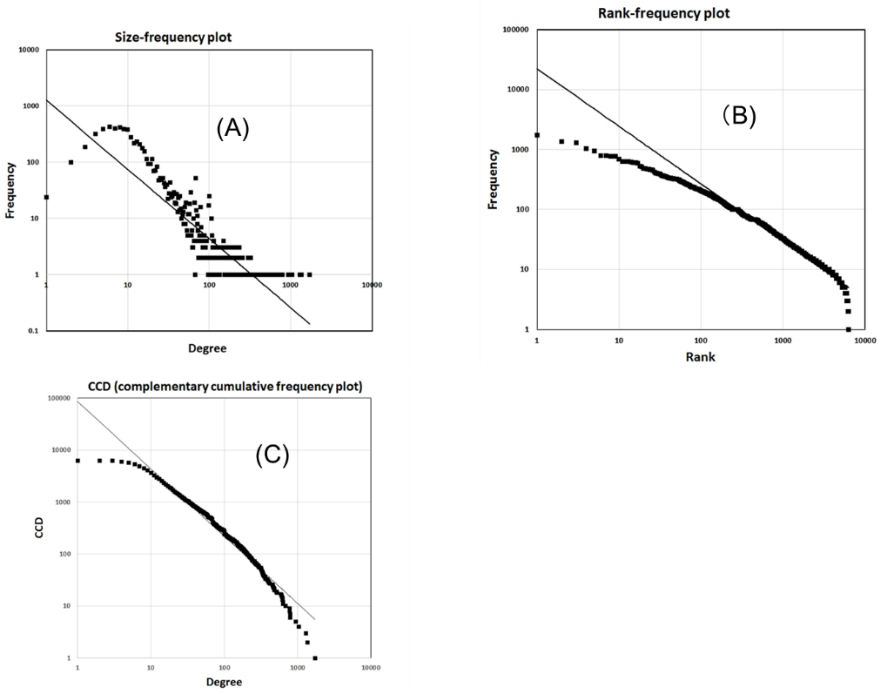


**Fig. III-A-7 Three plots of Frequency distribution**
(A) Frequency
(B) Zipf plot
(C) Complementary Cumulative Distribution Function

|  | slope | R2 | α derived from slope |
|---|---|---|---|
| Rank-Frequency plot | 0.96±0.003 | 0.94 | 2.001 |
| Size-Frequency plot | 1.23±0.04 | 0.76 | 2.23 |
| CCDF | 1.29±0.018 | 0.95 | 2.23 |

**Table III-A-1** Power-law exponent *α* derived from the Plots.
R2: Determination Coefficient

## 5 Number of diseases required.

### 5-1 Number of disease names required when the distribution follows a power-law.

Suppose the frequency distribution of disease names is as follows.

$$f(x) = \frac{A}{x^{\alpha}} .$$

Since the frequency of disease names is an integer at this time, the only part that makes sense is the part where *f(x)* ≥ 1. In other words, once *A* is determined, the number of disease names available is determined by the largest $a$ for which *f(a)* ≥ 1.

Then, the number of patients (*s*) for which the disease can be named is as follows.

$$s = \sum_{k=1}^{a} \left[\frac{A}{x^{\alpha}}\right] \approx \int_{1}^{a} \frac{A}{x^{\alpha}} dx . \qquad \text{( III-A- 16)}$$

Calculate this.

If α > 1,

$$s = \left[-\frac{A}{\alpha-1} x^{-\alpha+1}\right]_{1}^{a} = \frac{A}{\alpha-1}\left(1 - a^{1-\alpha}\right) .$$

Here, $\frac{A}{a^{\alpha}} = 1$ So, substituting $A = a^{\alpha}$, $s = \frac{1}{\alpha-1}\left(a^{\alpha} - a\right) \approx \frac{a^{\alpha}}{\alpha-1}$. Solving for $a$, we get

$$a = \sqrt[\alpha]{(\alpha-1)s} . \qquad \text{(III-A-17)}$$

This indicates that the number of disease names required when naming a disease for *s* patients is *a*.

For α = 1, then,

$$s = [A \log x]_{1}^{a} = A \log a = a \log a . \qquad \text{(III-A-18)}$$

Here, we used *A*= $a$.

### 5-2. Number of required disease names in case of exponential distribution

If $f(x) = A\exp(-bx)$ , as in the power-law distribution, we approximate as follows.

$$s = \sum_{k=1}^{a} A\exp(-bn) \approx \int_{1}^{a} A\exp(-bx)\, dx . \qquad \text{(III-A-19)}$$

However, since $A\exp(-ba) = 1$, $A = \exp(ba)$ or, $a = \log A / b$ .

If we perform an integral

$$s = \left[-\frac{A}{b}\exp(-bx)\right]_{1}^{a} = \frac{A}{b}\exp(-b) - \frac{A}{b}\exp(-ba) .$$

Since $A = \exp(ba)$ , $ab = \log A$. Then, we have Substitute

$$s = \frac{1}{b}\left(\frac{e^{ba}}{e^{b}} - 1\right) \approx \frac{e^{b(a-1)}}{b} .$$

Solving for *a, we* obtain

$$a = \frac{\log bs + b}{b} \quad . \qquad \text{(III-A-20)}$$

Here, *b* is appropriately matched to the actual measurement.

For example, if b=0.05, then s=1000 and *a*=80. To name a disease for 1000 people, we only need 80 disease names. This is about the same as the actual measured data. See Figure III-3-13 in the main text for a graph of the results.