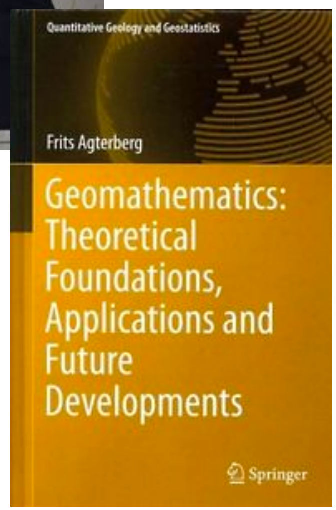
The **shape** of the frequency distribution allow us to find indications about the mechanisms generating randomness and the factors controlling it.



Agterberg, F.P., 2014. *Geomathematics: Theoretical foundations, applications and future developments*, Springer, 553).

Many **random processes** occur in nature. To make accurate predictions it is necessary to construct models that include “random components”.

Random: A haphazard course ***– at random***: without definite aim, direction, rule or method

The concept of **"randomness"** as used in common English, is different from its meaning in statistics.

To emphasize this difference, the word **stochastic** commonly is used in statistics for **random**, and ***stochastic process*** is a ***random process***.

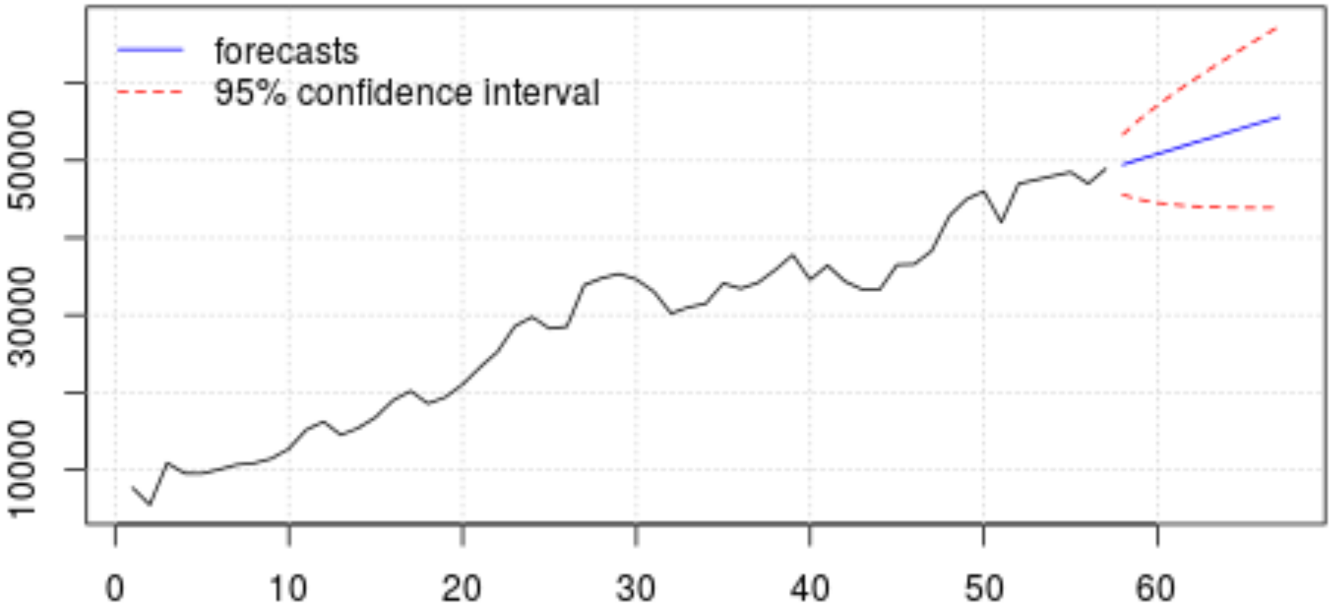
A **stochastic process** is one that includes any **random components**, and a process without random components is called ***deterministic***.

Because environmental phenomena nearly always include random components, the study of stochastic processes is essential for making valid *environmental predictions*.

The world we live in consists of many identifiable **cause-effect relationships**.

A **cause-effect relationship** is characterised by the certain knowledge that, if a specified action takes place, a particular result always will occur, and there are no exceptions to this rule.

Such a process is called **deterministic**, because the resulting outcome is determined completely by the specific cause, and the outcome can be predicted with certainty.

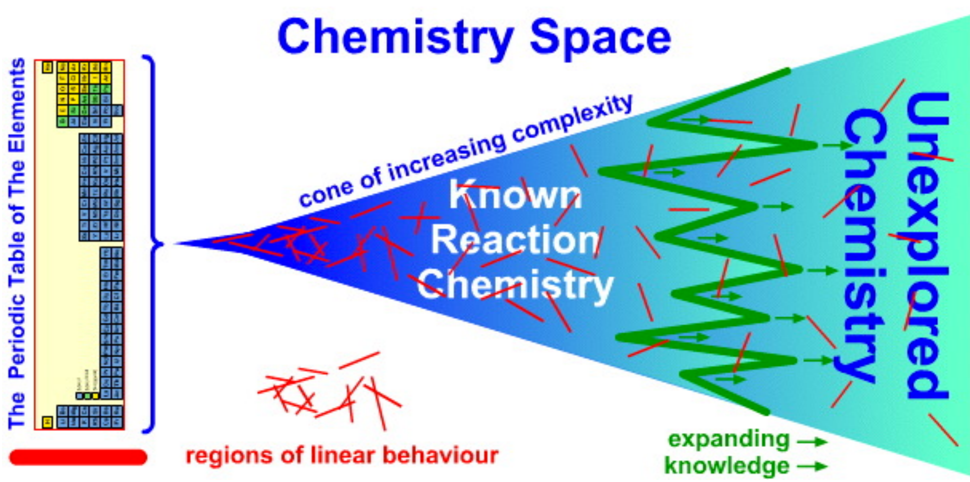


To make accurate predictions about the system's future behaviour

Two independent sources of information

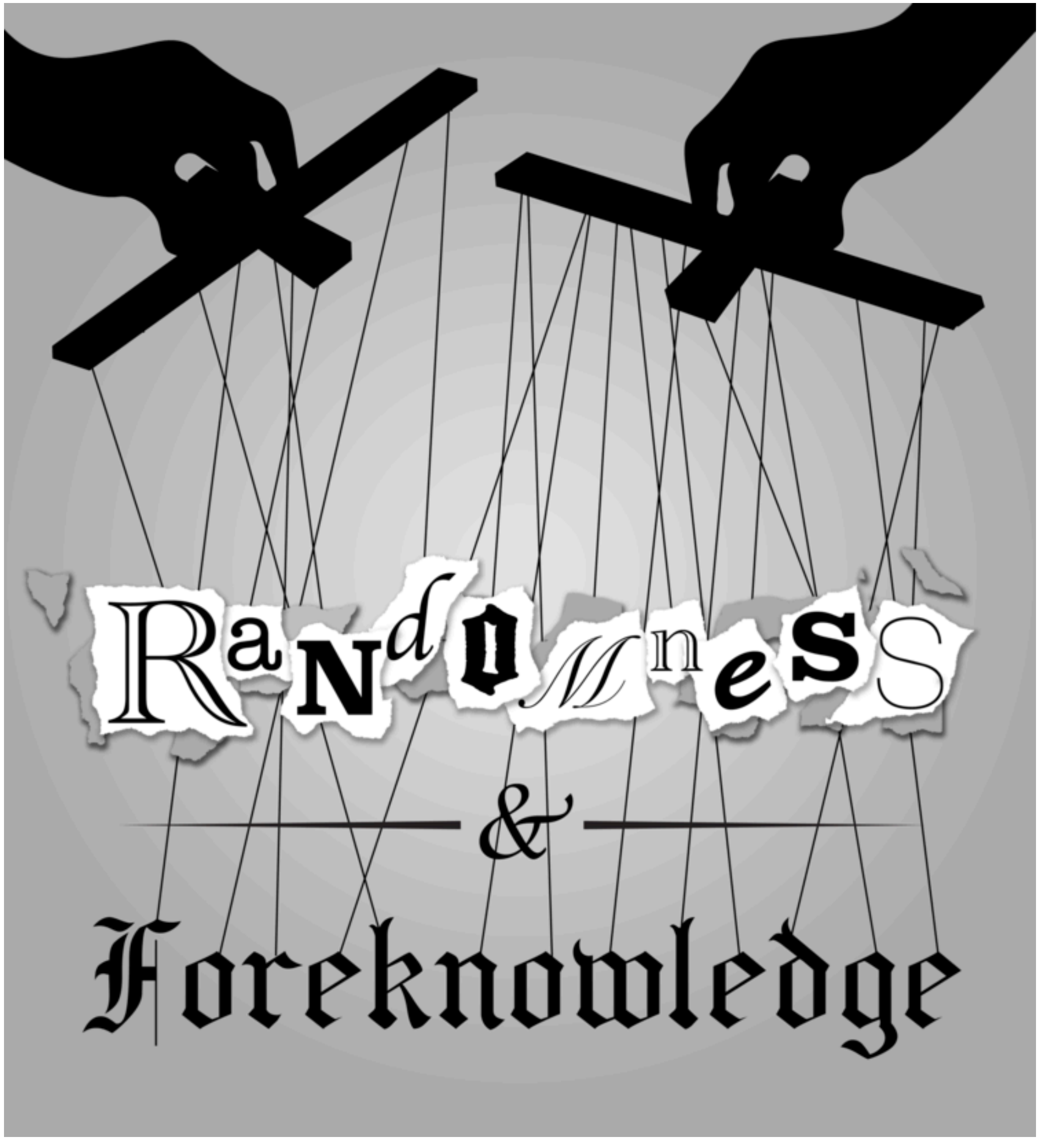
Observational knowledge about its behaviour

Physical (chemical) knowledge of the structure of the system



Conceptual model (theoretical information)

Validation of the conceptual model with real observations (empirical information)

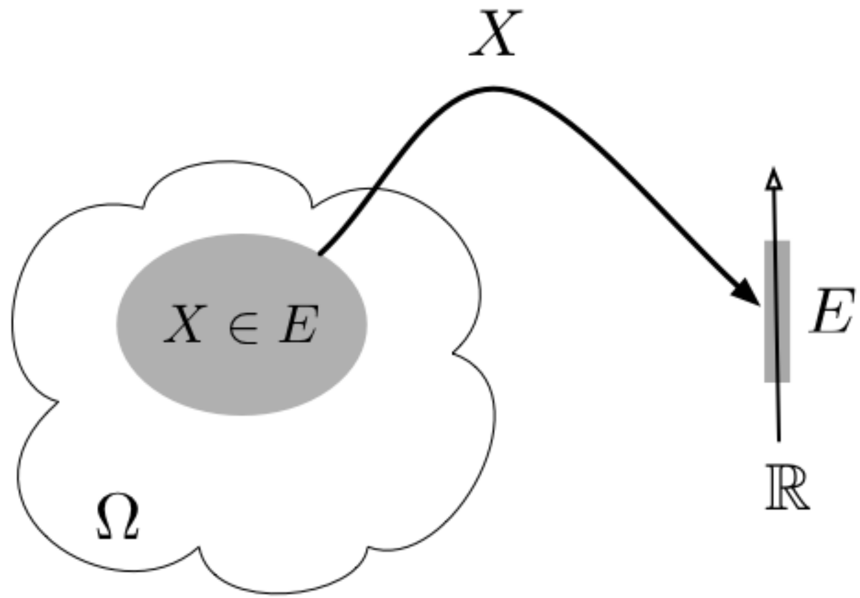


A theoretical model validated by empirical observation usually provides a **powerful tool** for predicting future behaviour of a system or process.

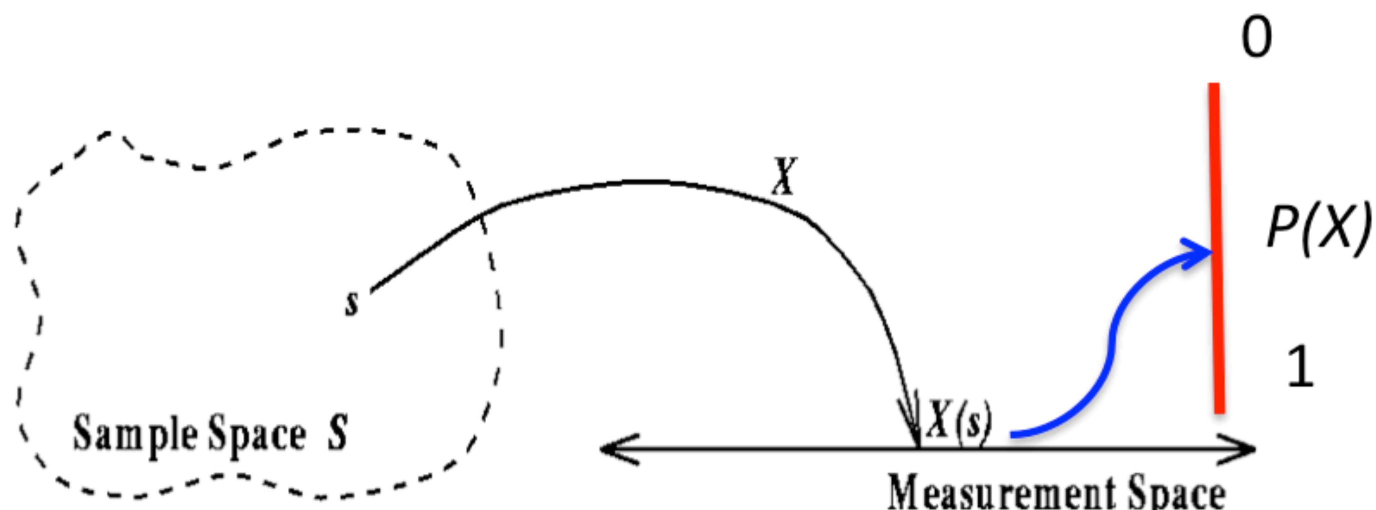
However sometime our knowledge will be vague and uncertain or very limited.

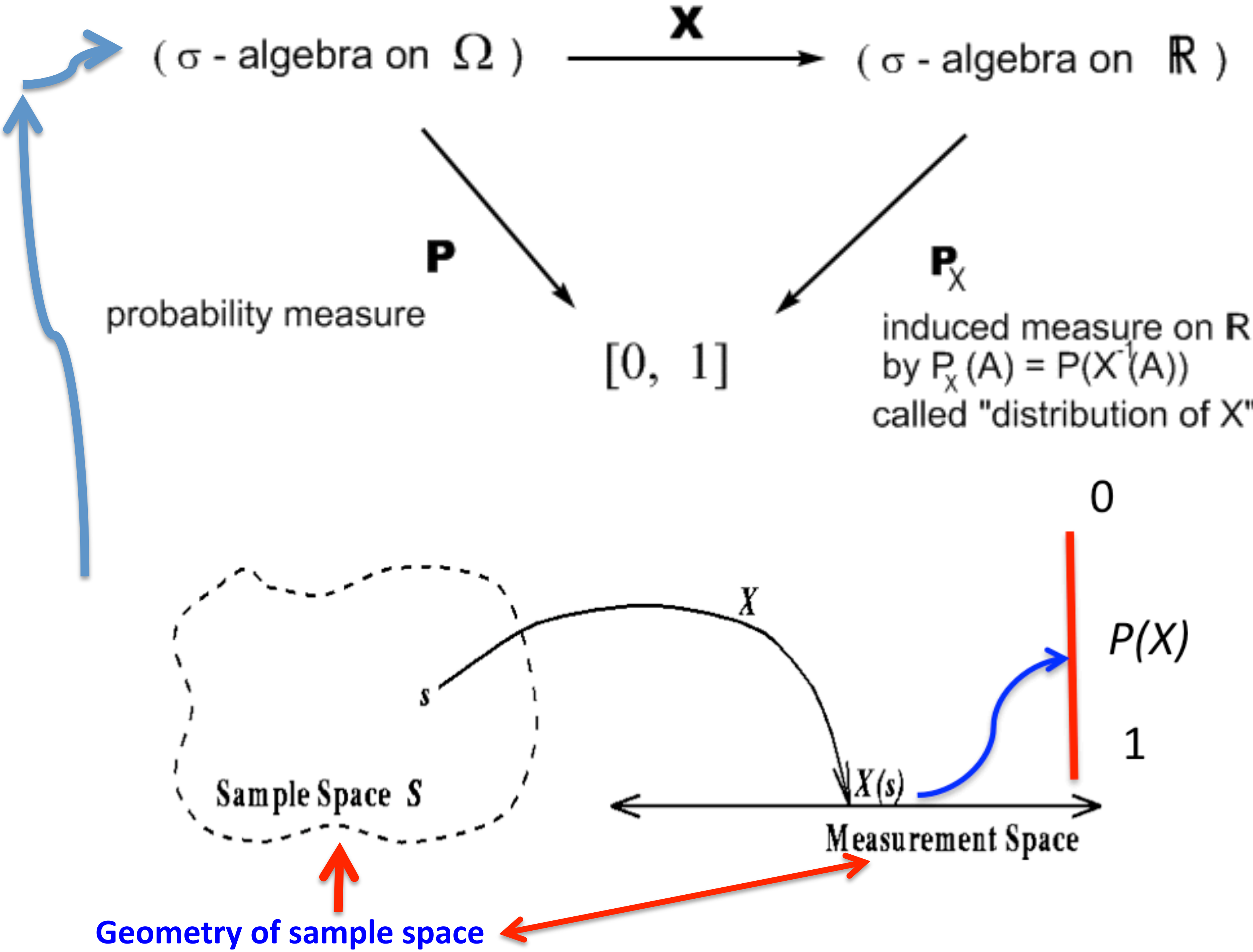
Despite our lack of information it may be necessary to make a prediction about the future behaviour of the system.

Introduction to the concept of **random variable**.

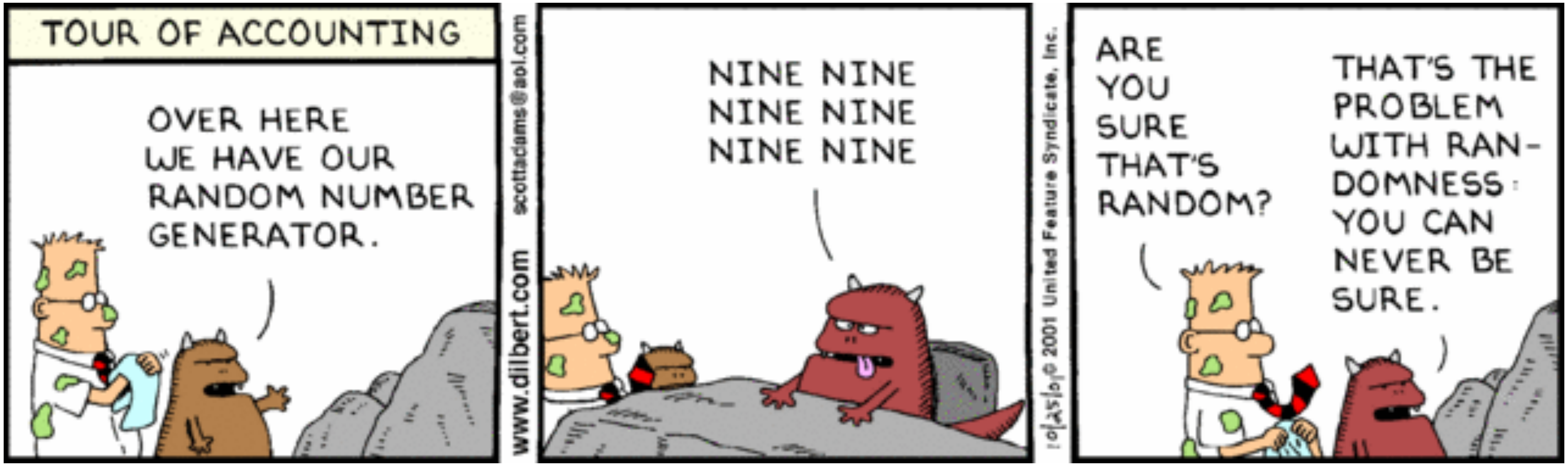


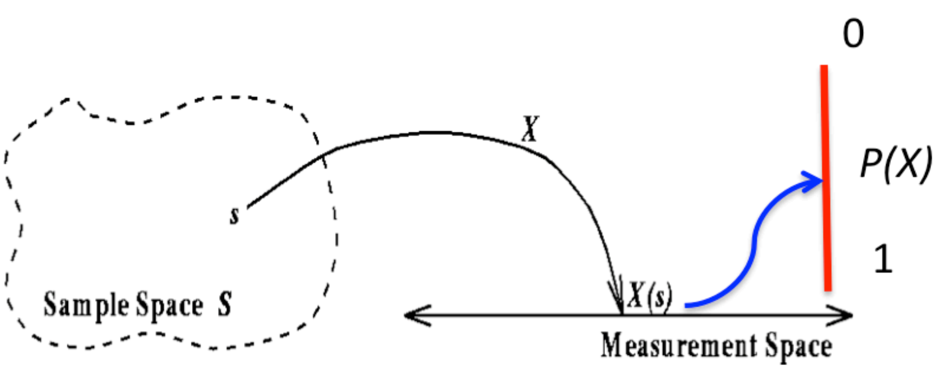
A random variable X is a mapping from Ω to $\mathbb{R}$.

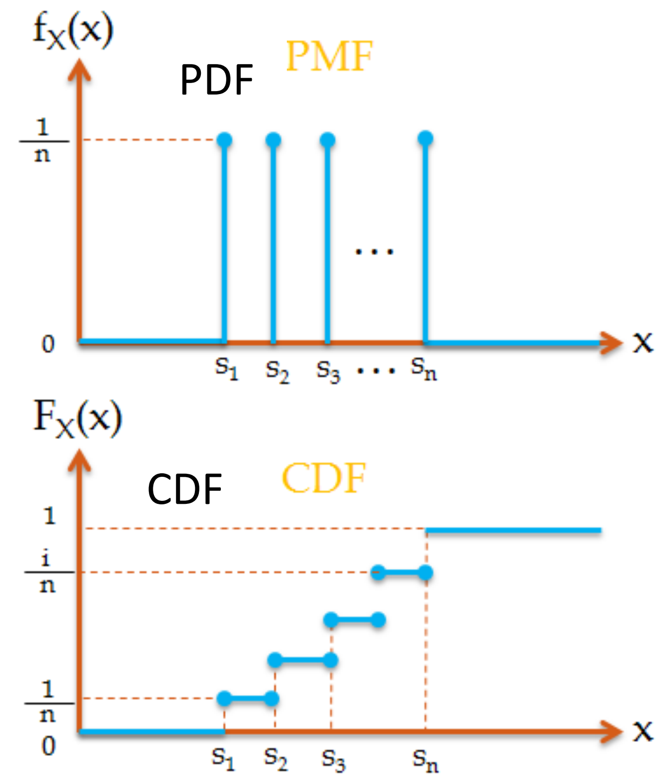


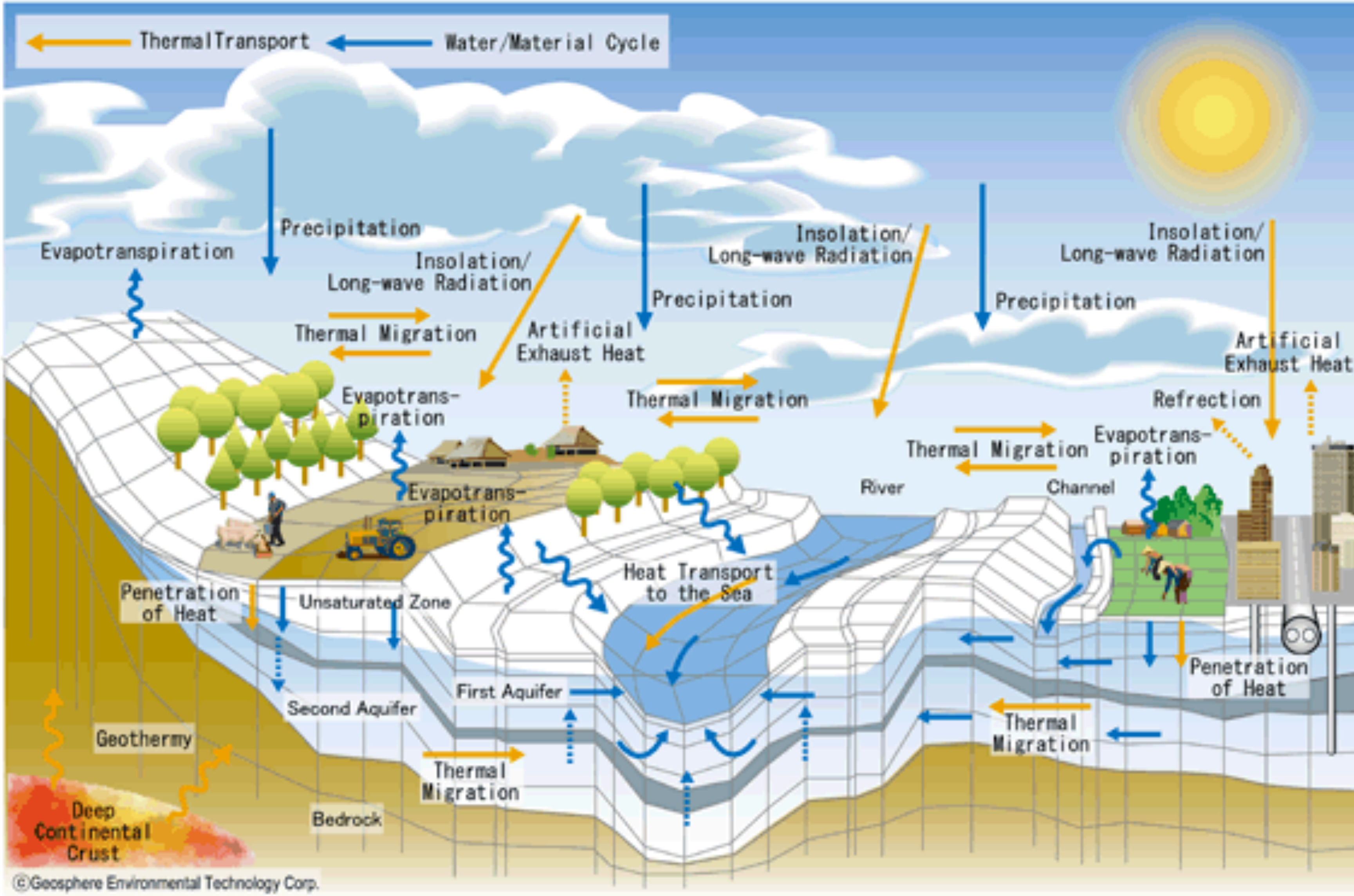


(σ - algebra on Ω)
X
(σ - algebra on ℝ)
P
probability measure
[0, 1]
P_X
induced measure on R
by P_X (A) = P(X^-1(A))
called "distribution of X"
0
X
P(X)
s
1
X(s)
Sample Space S
Measurement Space
Geometry of sample space

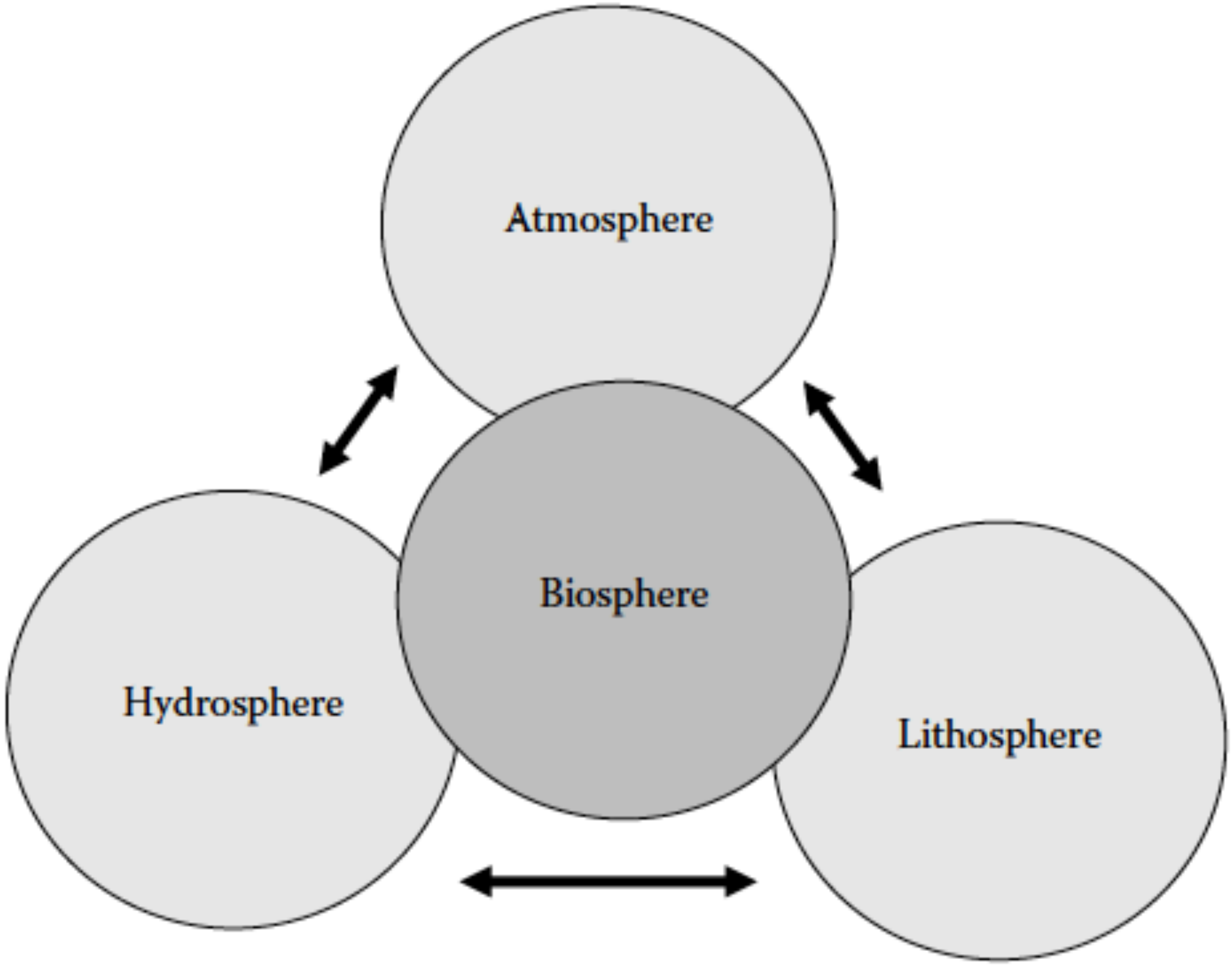

TOUR OF ACCOUNTING
OVER HERE WE HAVE OUR RANDOM NUMBER GENERATOR.
scottadams@aol.com
www.dilbert.com
NINE NINE NINE NINE NINE NINE
© 2001 United Feature Syndicate, Inc.
ARE YOU SURE THAT'S RANDOM?
THAT'S THE PROBLEM WITH RANDOMNESS: YOU CAN NEVER BE SURE.


0
X
s
P(X)
1
X(s)
Sample Space S
Measurement Space


$f_X(x)$
PDF
PMF
$\frac{1}{n}$
...
0
x
s_1 s_2 s_3 ... s_n
$F_X(x)$
CDF
CDF
1
$\frac{i}{n}$
$\frac{1}{n}$
0
x
s_1 s_2 s_3 s_n



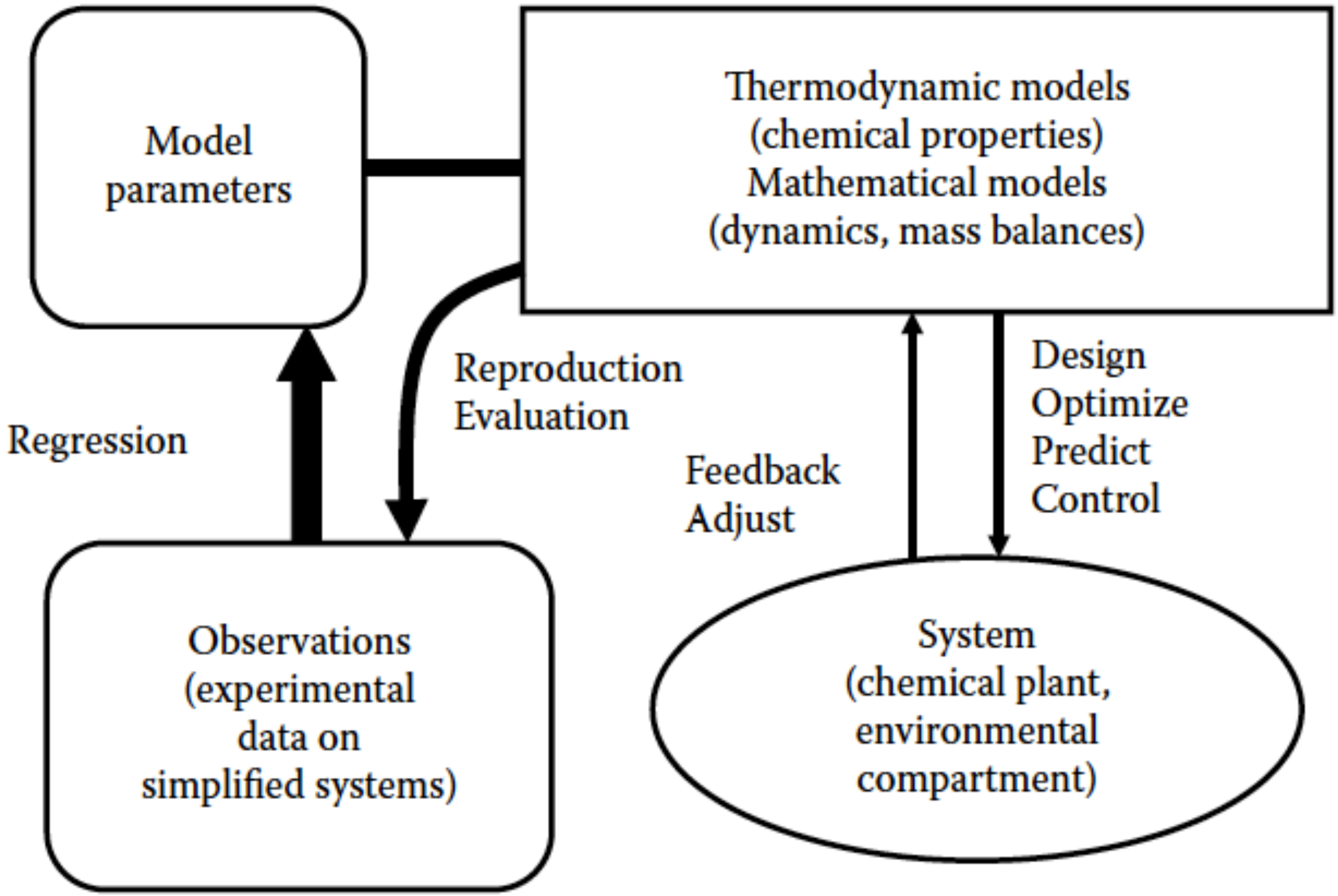
Random variable: sum of independent components? Product of independent component? What else? How environmental factor combines?



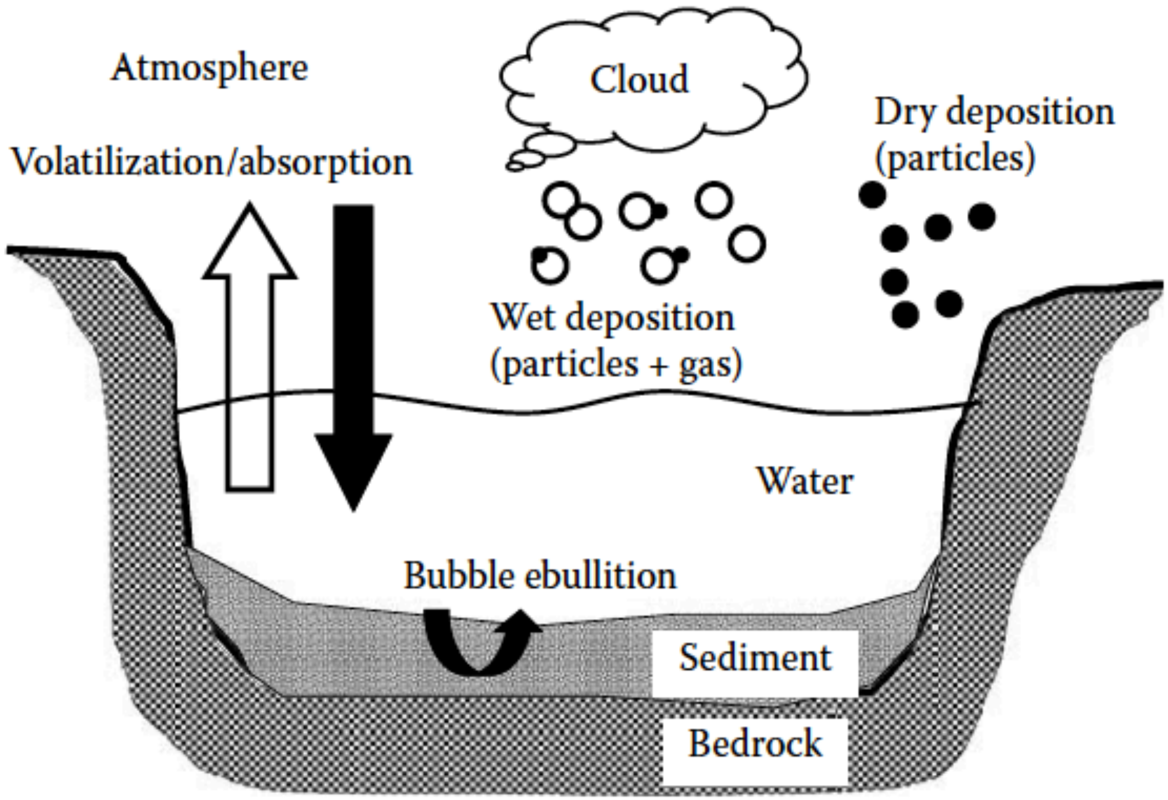
Equilibrium between various environmental compartments, reactions within compartments, and material exchange between the compartments.

Applications of Chemical Thermodynamics and Kinetics in Environmental Processes

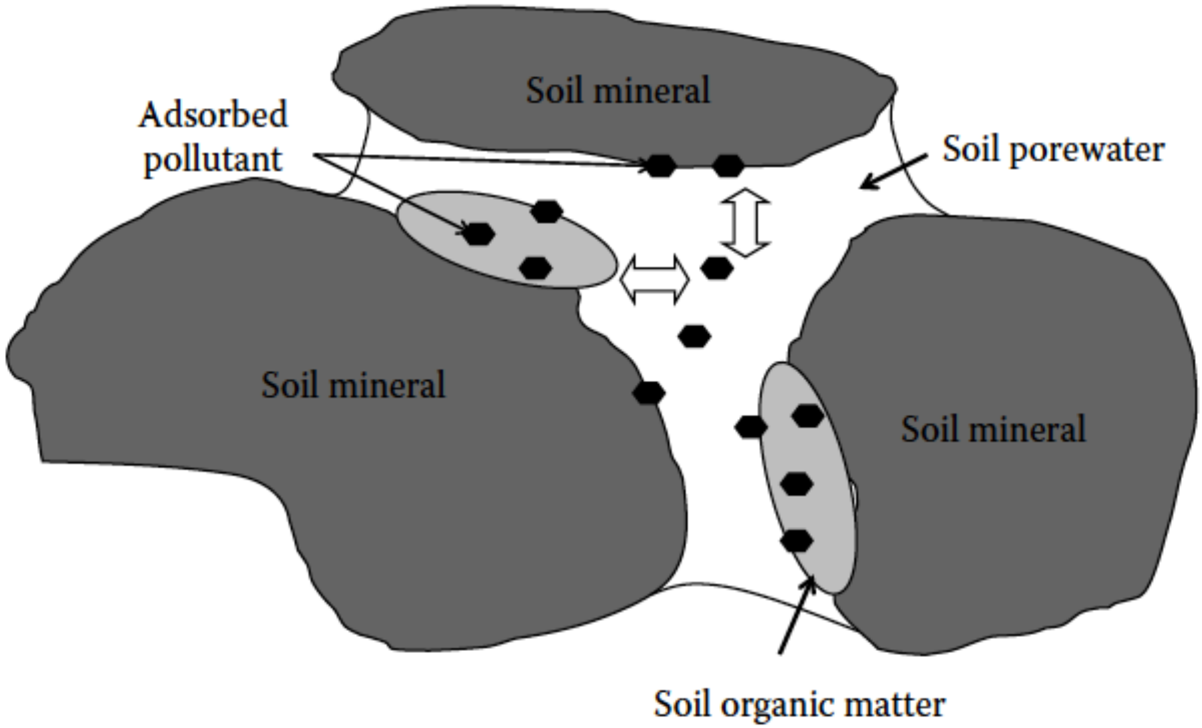
Process	Typical Equilibrium Representation	Thermodynamic Property	Kinetic Property
Solubility in water	A (pure) $\rightleftharpoons$ A (water)	Saturation solubility, C_i^*	Dissolution rate
Absorption in water	A (air) $\rightleftharpoons$ A (water)	Henry's law constant, K_H	Absorption rate
Precipitation from water	A (water) $\rightleftharpoons$ A (crystal)	Solubility product, K_{sp}	Precipitation rate
Volatilization from water	A (water) $\rightleftharpoons$ A (air)	Henry's law constant, K_H	Volatilization rate
Evaporation from pure liquid	A (pure) $\rightleftharpoons$ A (vapor)	Vapor pressure, P_i^*	Evaporation rate
Acid/base dissociation	$A \rightleftharpoons A^- + H^+$	Acidity or basicity constant, K_a or K_b	Acidification rate
Ion exchange	$A^+ + BX \rightleftharpoons BA + X^+$	Ion-exchange partition constant, K_{exc}	Ion-exchange rate
Oxidation/reduction	$A_{ox} + B_{red} \rightleftharpoons A_{red} + B_{ox}$	Equilibrium constant, K_i	Redox reaction rate
Adsorption from water	A (water) $\rightleftharpoons$ A (surface)	Soil/water partition constant, K_{SW}	Adsorption rate
Adsorption from air	A (air) $\rightleftharpoons$ A (surface)	Particle/air partition constant, K_{AP}	Adsorption rate
Uptake by biota	A (water) $\rightleftharpoons$ A (biota)	Bio-concentration factor, K_{BW}	Rate of uptake
Uptake by plants	A (air) $\rightleftharpoons$ A (plant)	Plant/air partition constant, K_{PA}	Rate of uptake
Chemical reaction	A + B $\rightleftharpoons$ Products	Equilibrium constant, K_{eq}	Rate of chemical reaction
Photochemical reaction	A + $(h\nu)$ $\rightleftharpoons$ Products	Equilibrium constant, K_{eq}	Rate of photolysis
Biodegradation reaction	A + (enzymes) $\rightleftharpoons$ Products	Equilibrium constant, K_{eq}	Michaelis–Menten and Monod kinetics constants



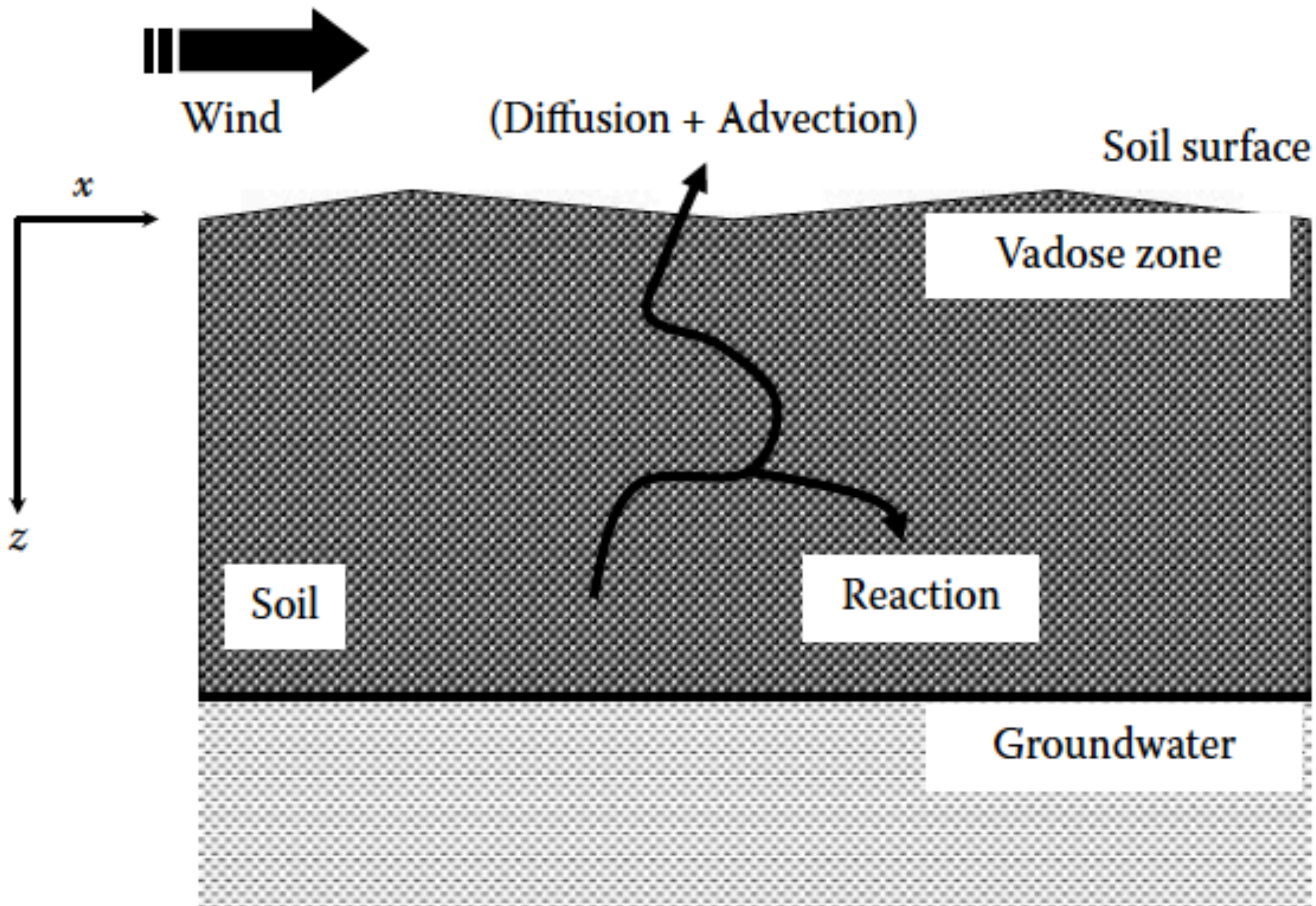
Role of chemical thermodynamics in understanding natural and engineered systems. (Modified from http://www.codata.org/codata02/04physci/rarey.pdf.)



Exchange of chemicals between the water and the atmosphere. Equilibrium is important in many cases such as volatilization–absorption, wet deposition, dry deposition, and gas bubble transport.



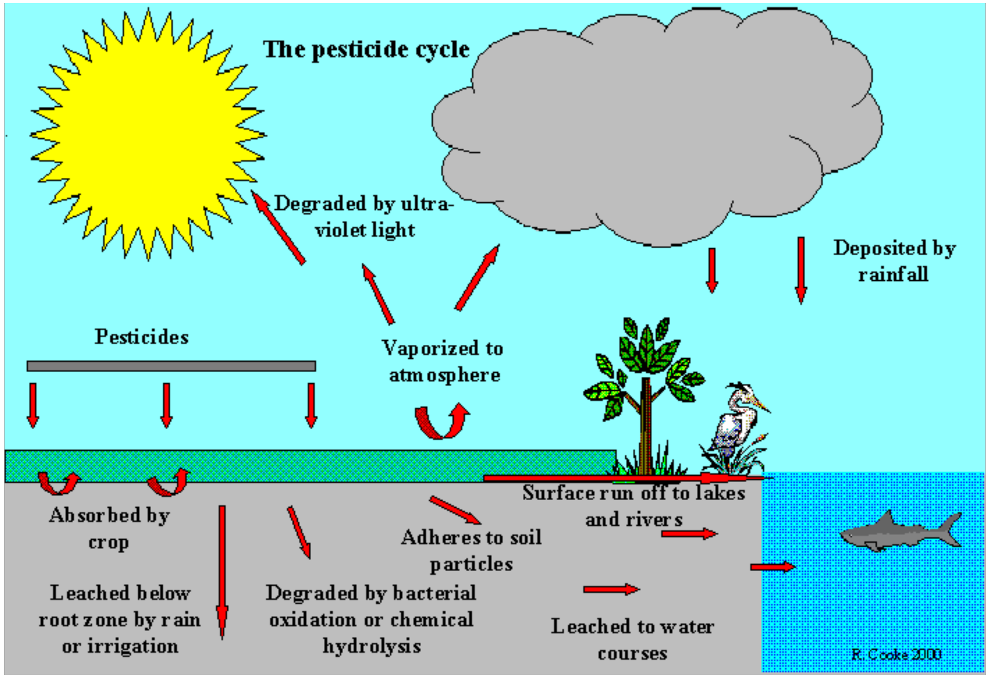
Partitioning of a solute between soil and porewater. Various niches are shown where pollutants reside in the soil pore.



Transport of a contaminant from the soil to the atmosphere.

The environmental variables actually observed are the **consequence** of thousands of events, some of which may be poorly defined or imperfectly understood.

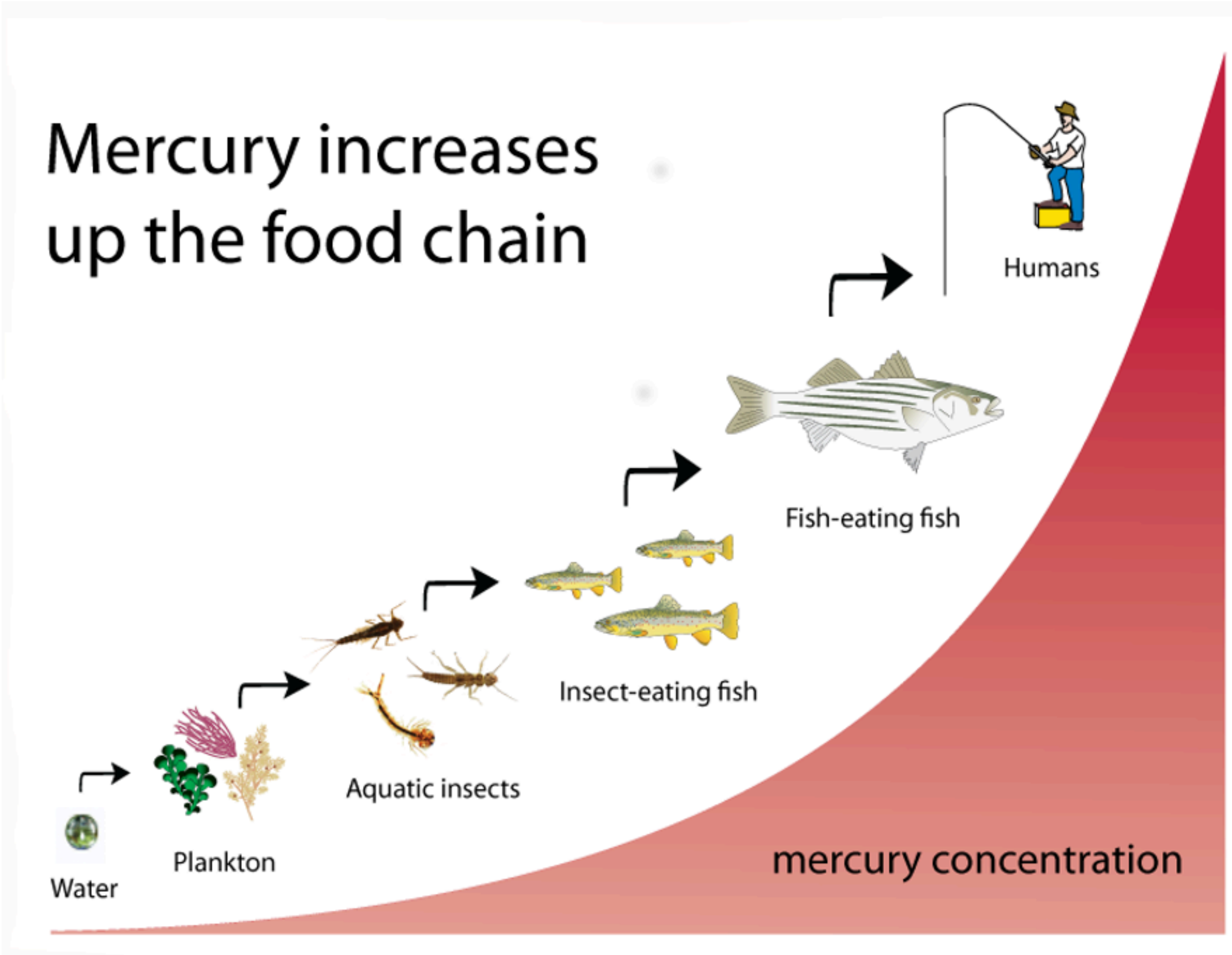
The concentration of a pesticide observed in a stream results from the combined influence of many complex factors, such as the amount of pesticide applied to crops in the area, the amount deposited on the soils, irrigation, rainfall, seepage into the soil, the nature of the surroundings terrain, porosity of the soil, mixing and dilution as the pesticide travels to the stream, flow rates of adjoining tributaries, chemical reactions, many other factors…



Factors will change with time and space.

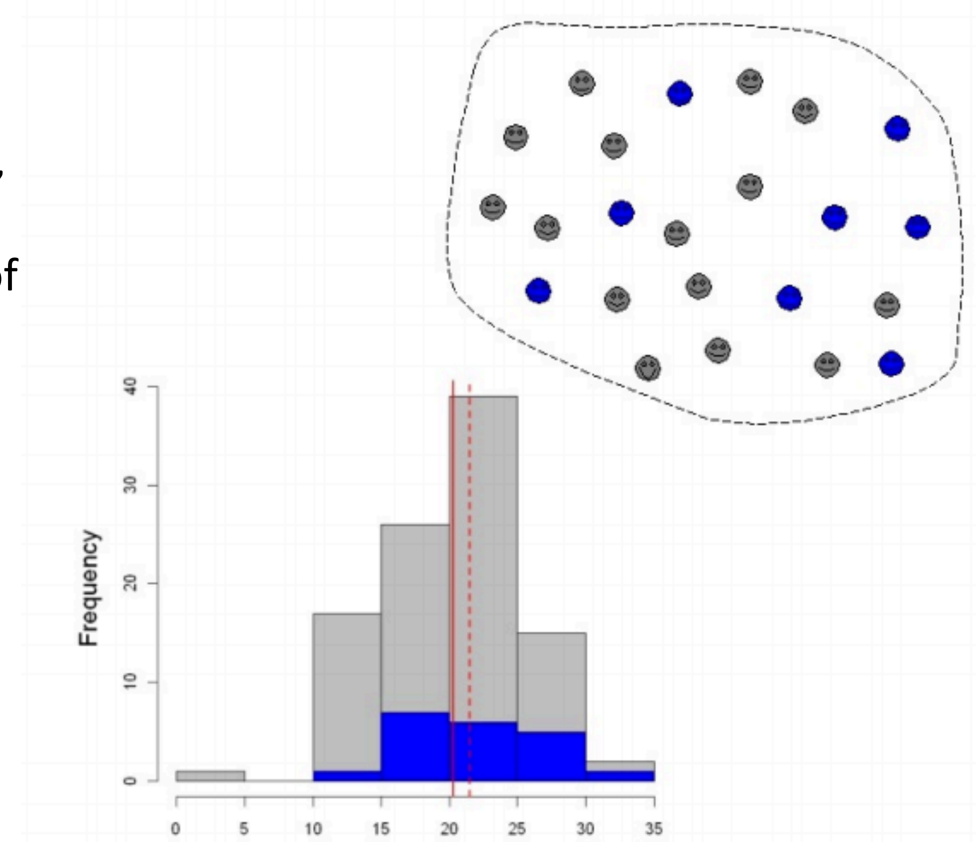
The concentrations of an **air pollutant** observed in a city often are influenced by hundreds or thousands of sources in the area, atmospheric variables (wind speed and direction, temperature, weather conditions, mechanical mixing and dilution, chemical reactions, interaction with physical surfaces or biological systems, other phenomena…)

Even more complex are the factors that affect pollutants as they move through the **food chain**, from source to soils, to plants, to animals, and to man, ultimately becoming deposited in human tissue or in body fluids.



Distribution analysis

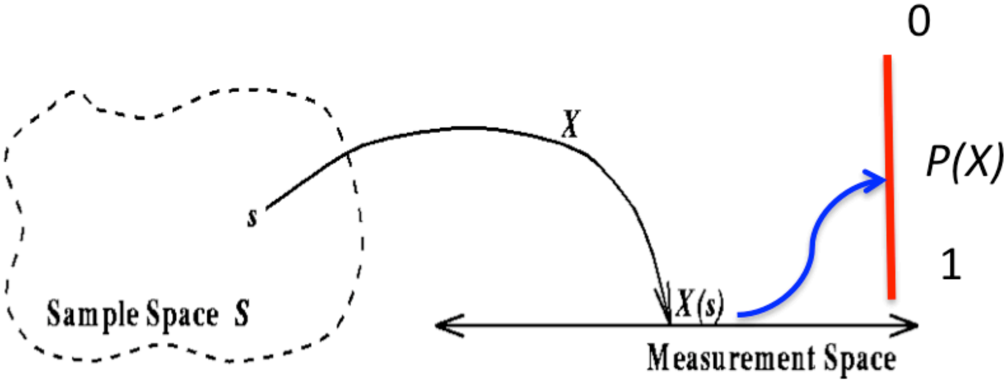
A **histogram** is a graphical display of tabulated frequencies (or probabilities), shown as bars; it shows what proportion of cases fall into each of several adjacent non-overlapping categories; and it is a way of binning the data.

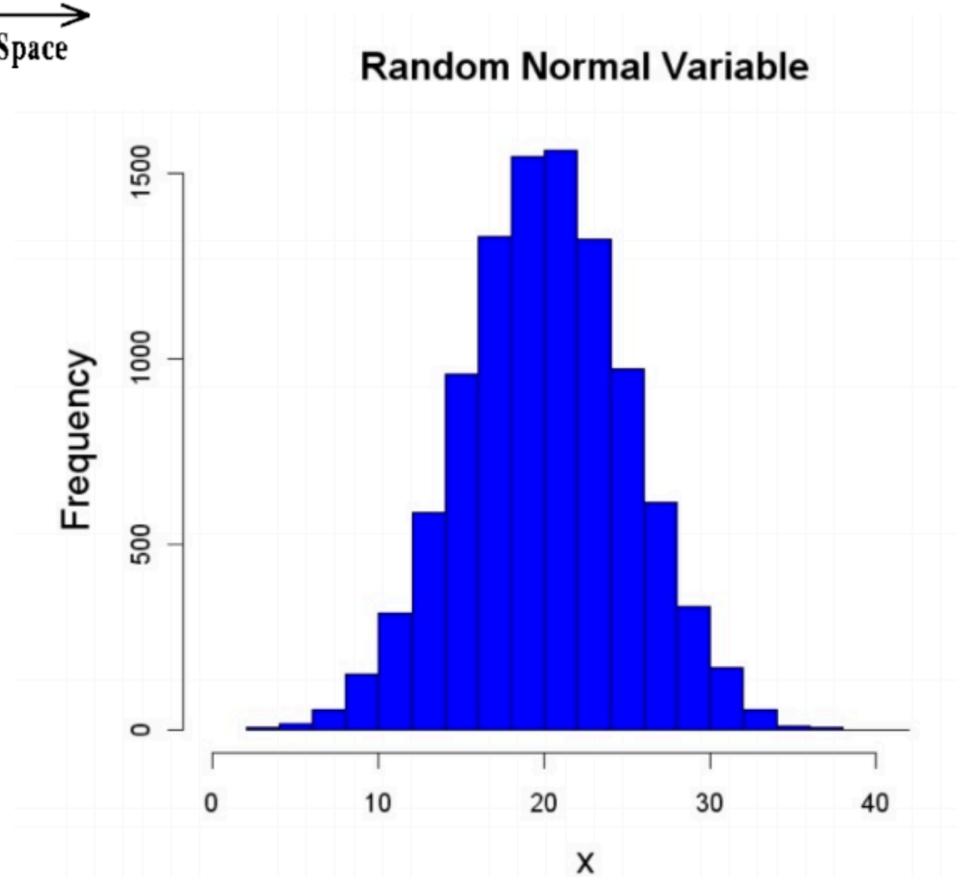


Histograms are extremely useful for quickly visualizing the distribution of values in a variable.

As environmental data are often **highly skewed**, a histogram will readily reveal that skew.

Moreover, extreme values in the distribution (i.e., potential **outliers**) often show up as isolated bars on the tail of the distribution.

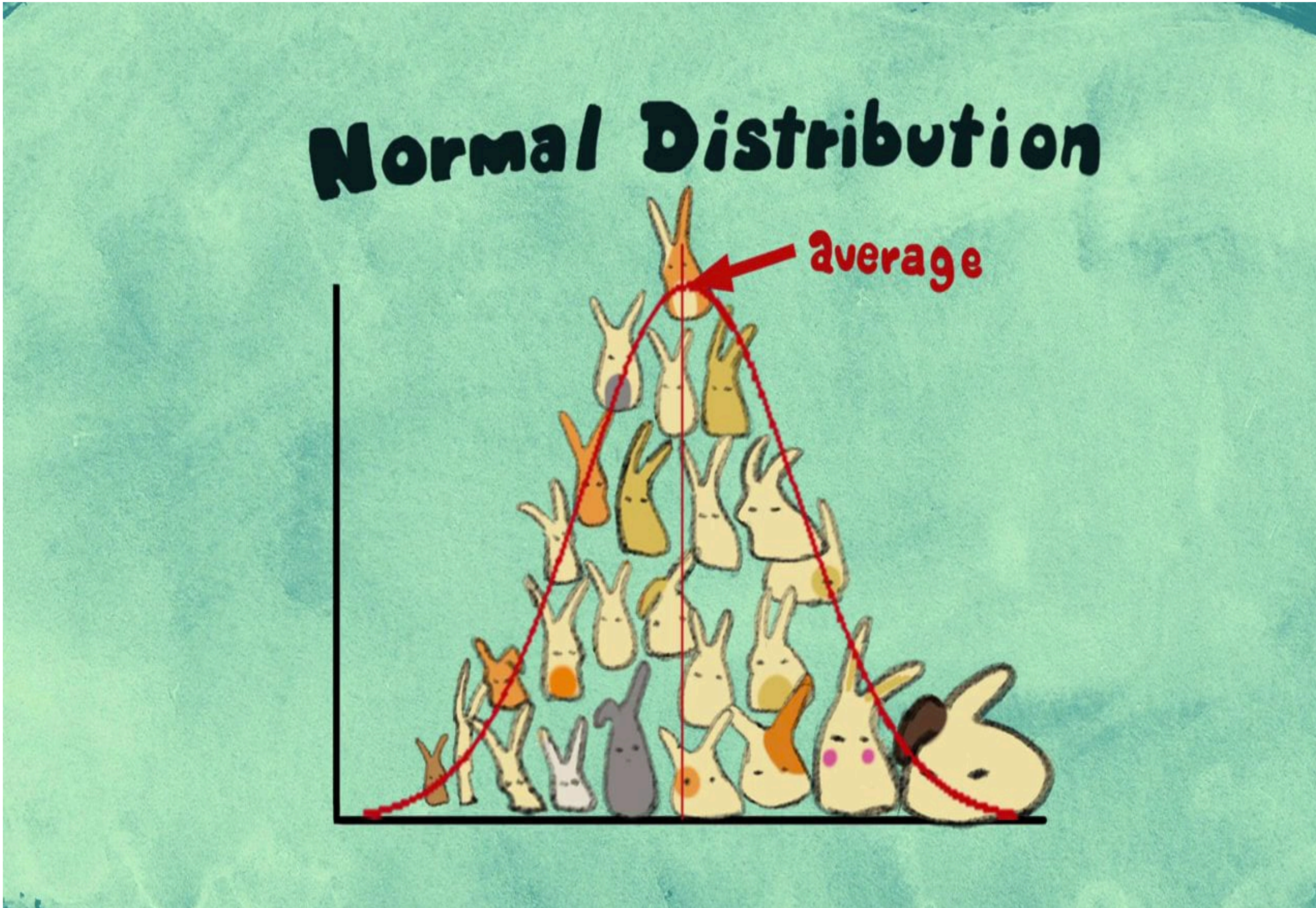


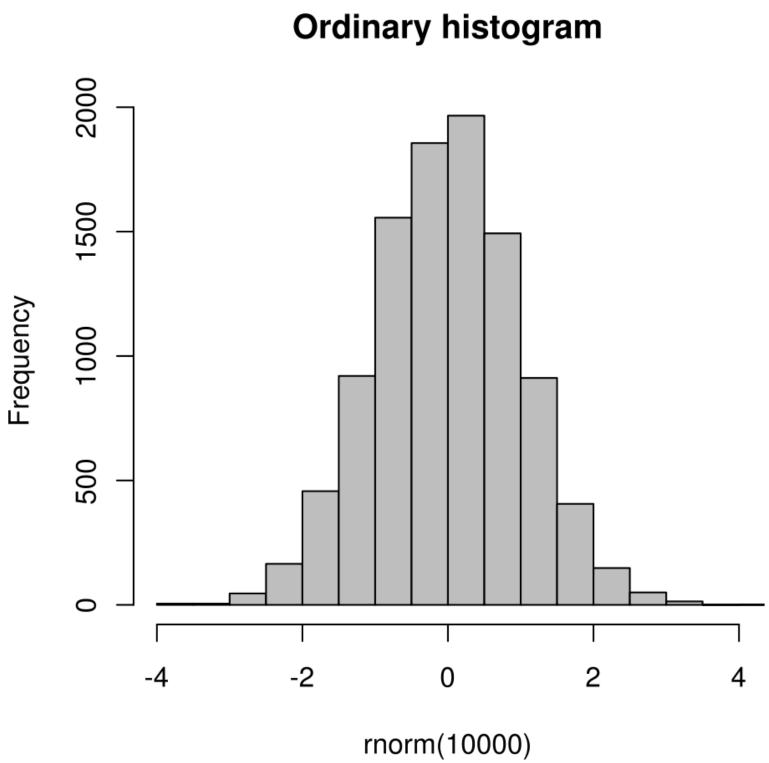


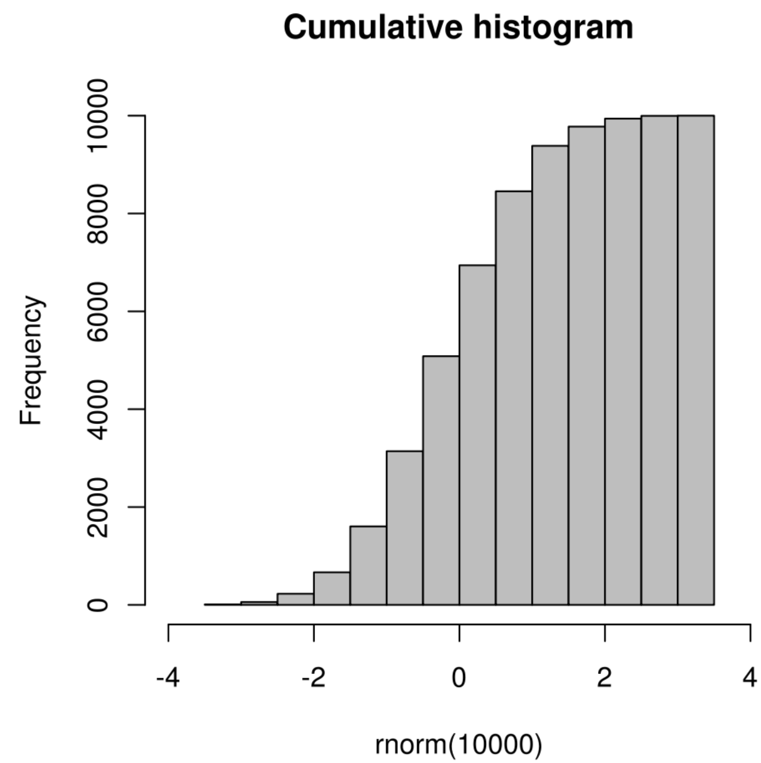
the "normal" distribution is frequently used as a reference framework for describing data characteristics and is the basis for most classical statistics.

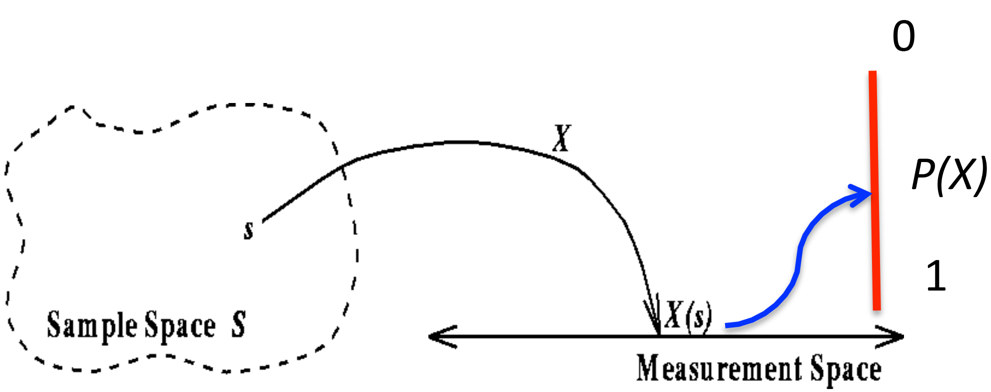
A normal distribution describes a data set that exhibits a symmetrical "bell- shaped" frequency distribution.

A collection of values that concentrate around a single central tendency (the average value) and trail off in both directions at the same rate.

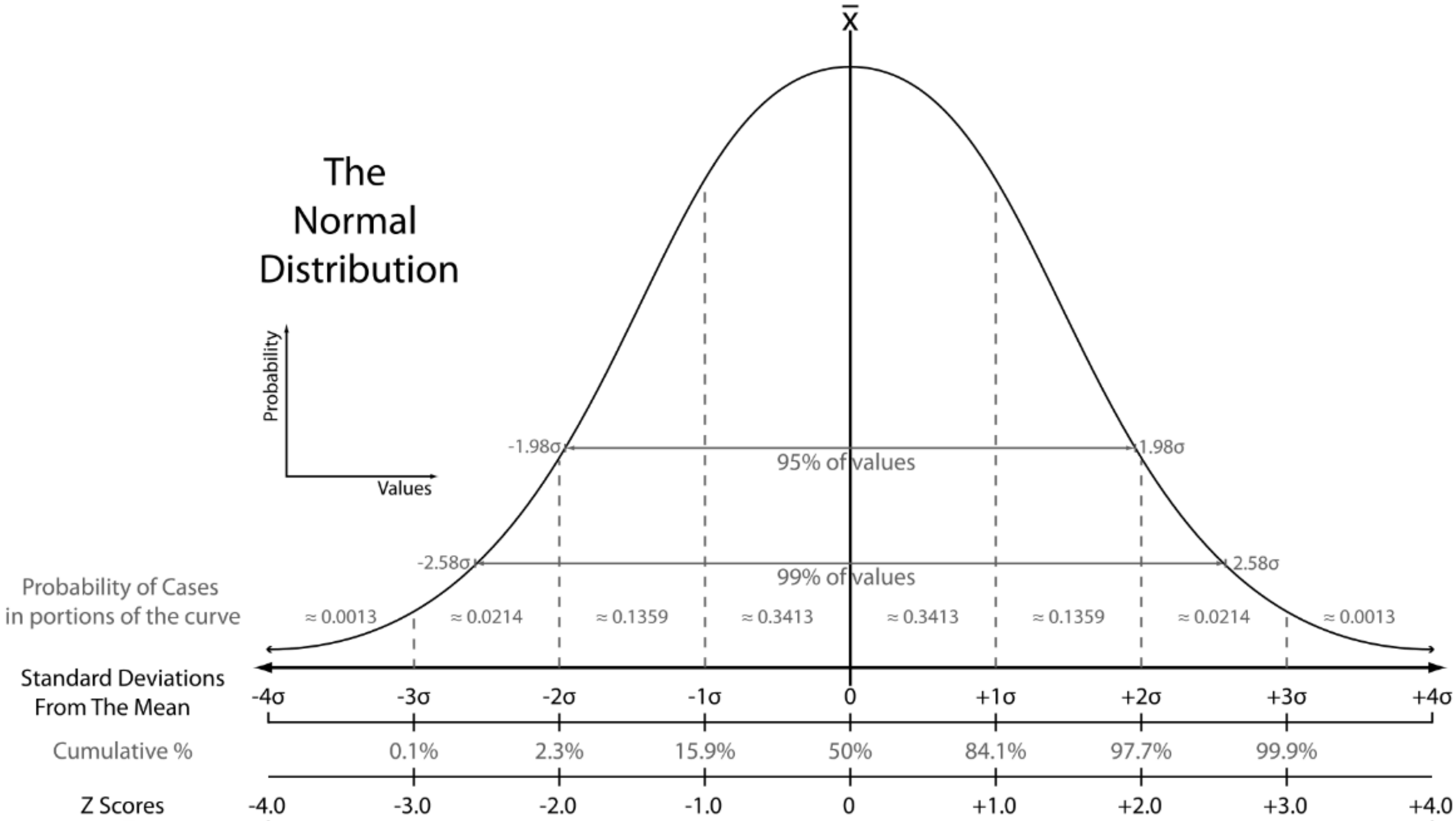

Normal Distribution
average





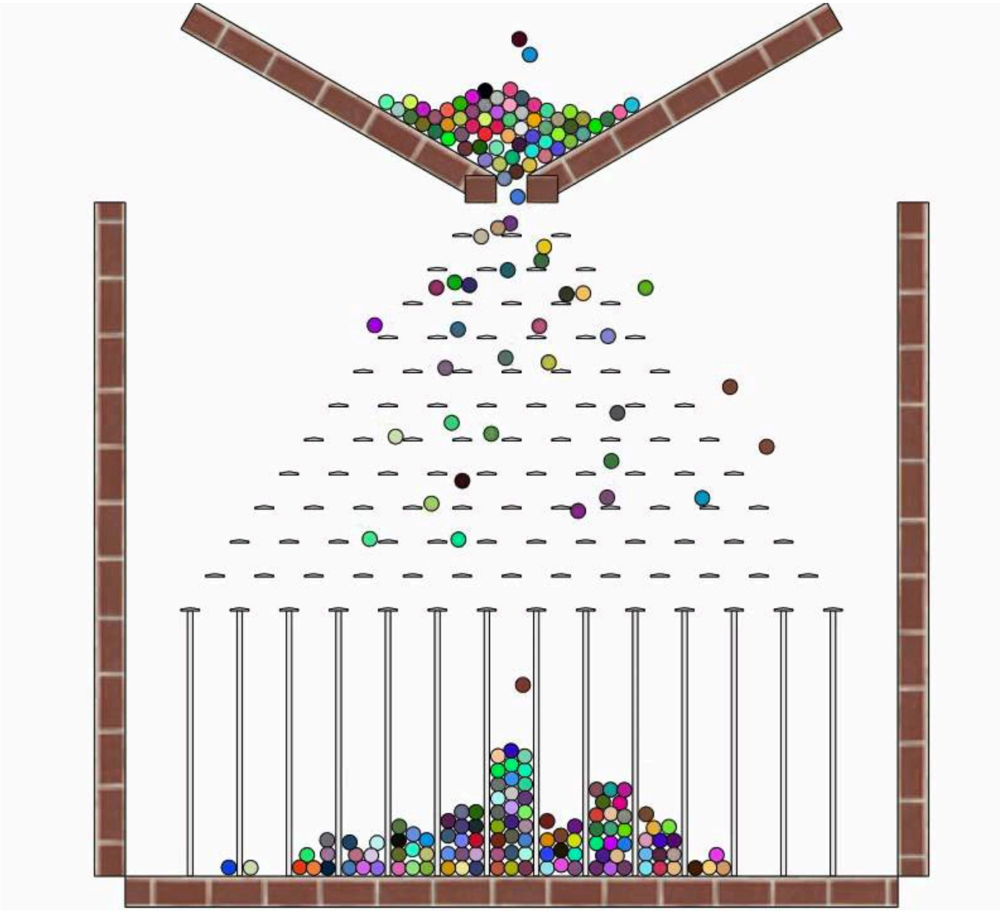


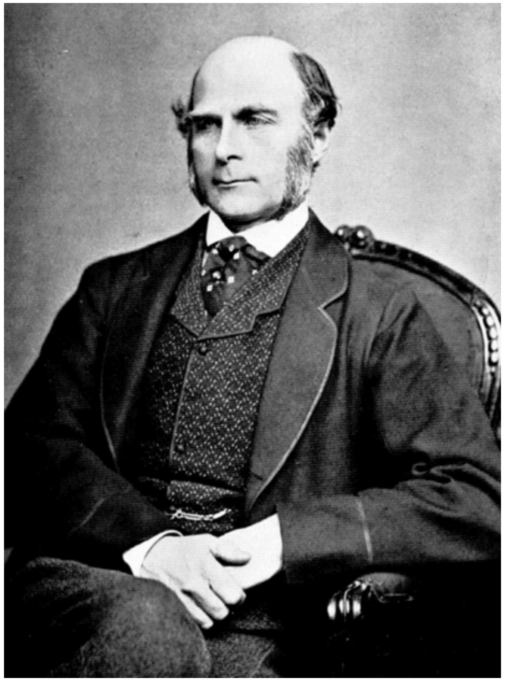
$$P(X) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$$



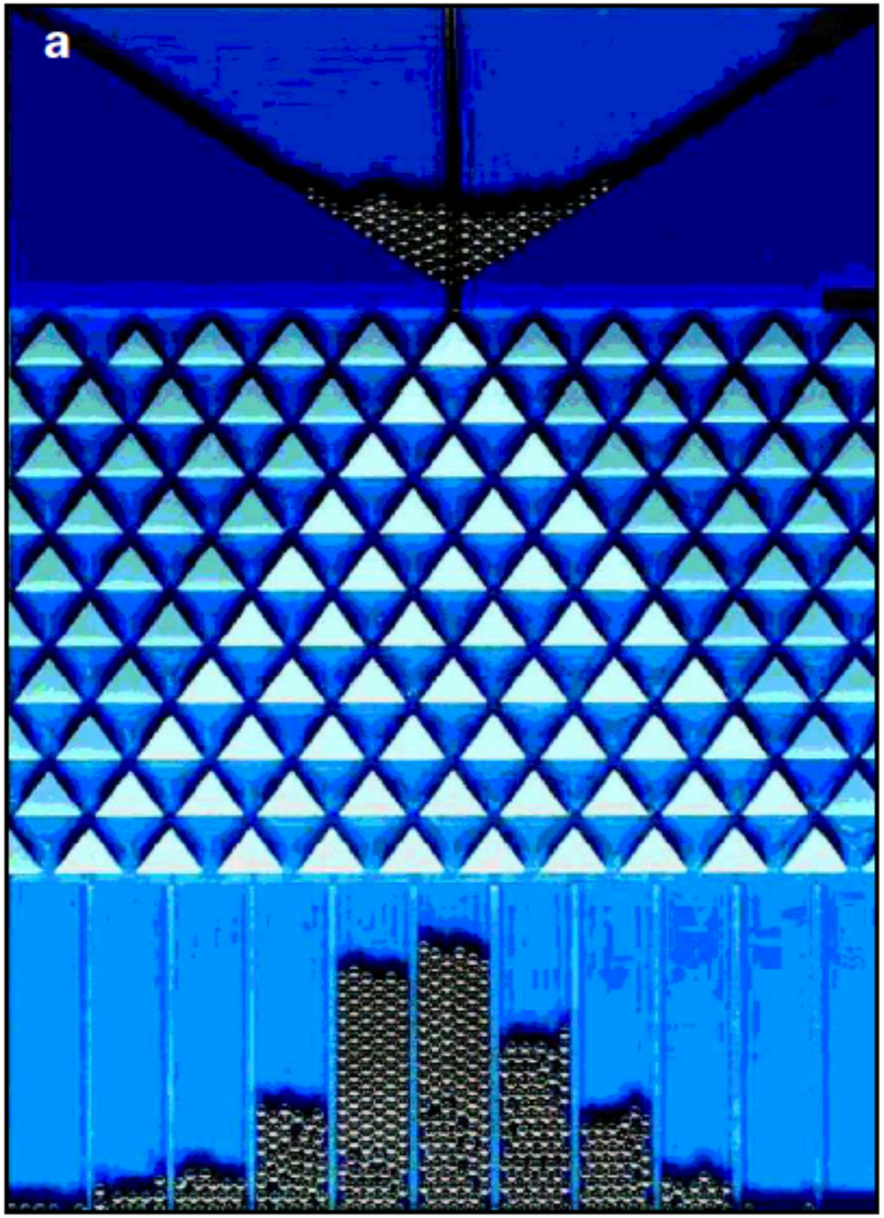
$$z_i = \frac{x_i - \bar{x}}{s}$$

The key premise is that a **distribution's shape** reveals information about the governing dynamics of the system that gave rise to the distribution.





Francis Galton (1822 – 1911)

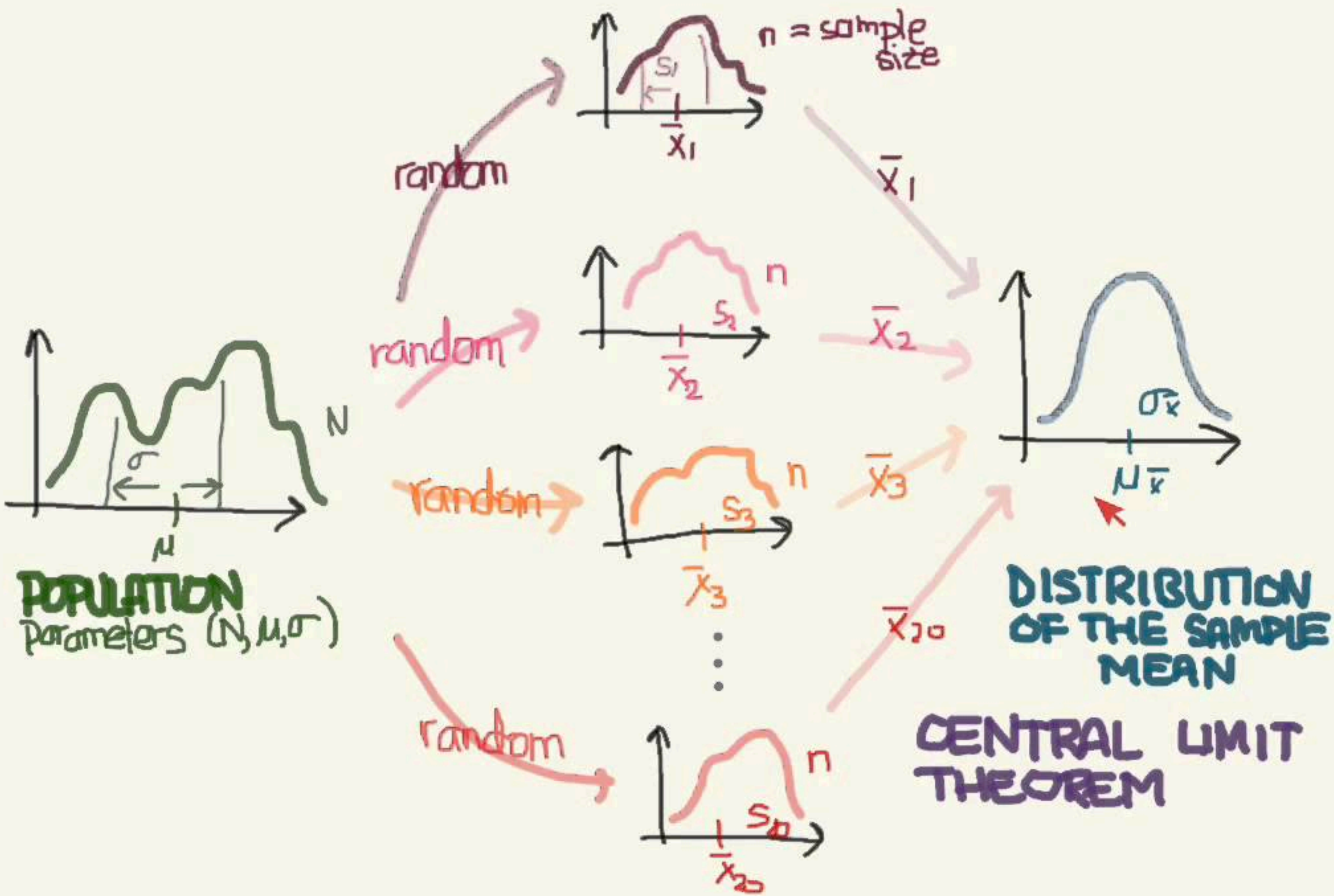


Particles fall from a funnel onto tips of triangles, where they are **deviated** to the left or to the right with equal probability (0.5) and finally fall into receptacles.

The tip of a triangle is at distance x from the left edge of the board, triangle tips to the right and to the left below it are placed at x + c and x – c.

The distribution is generated by many small random effects (according to the **central limit theorem**) that are Additive.

Roughly, the **central limit theorem** states that the distribution of the sum (or average) of a large number of independent, identically distributed variables will be approximately normal, regardless of the underlying distribution.



n = sample size
random
random
random
random
s_1
$\bar{x}_1$
s_2
n
$\bar{x}_2$
s_3
n
$\bar{x}_3$
s_{20}
n
$\bar{x}_{20}$
$\bar{x}_1$
$\bar{x}_2$
$\bar{x}_3$
$\bar{x}_{20}$
N
σ
μ
POPULATION
parameters (N, μ, σ)
$\sigma_{\bar{x}}$
$\mu_{\bar{x}}$
DISTRIBUTION OF THE SAMPLE MEAN
CENTRAL LIMIT THEOREM

Conditions for normal process

Many variables found in nature result from the summing of numerous unrelated components.

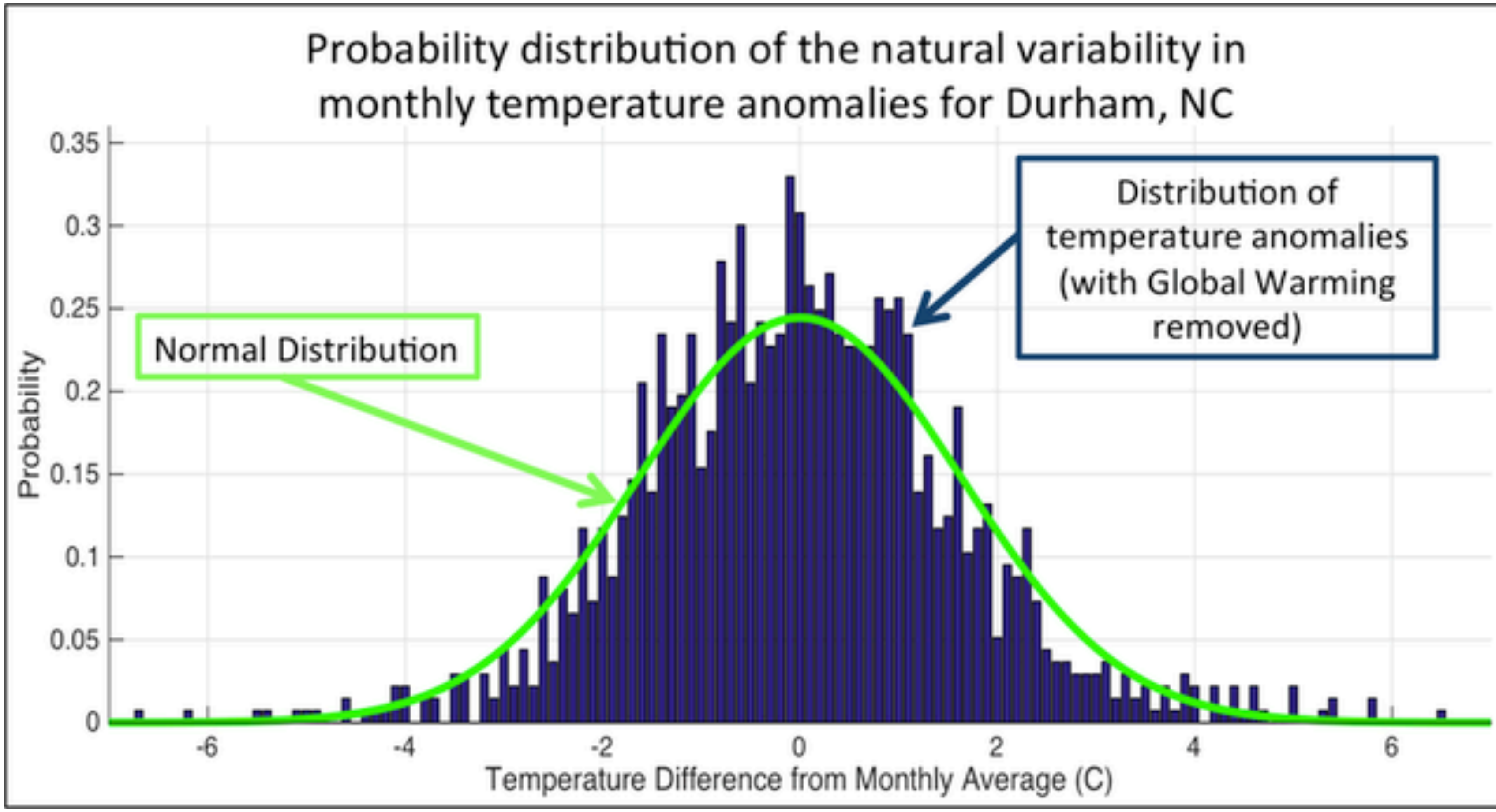
When the individual component are sufficiently unrelated and complex, then the resulting sum tend toward normality as the number of component becomes increasingly large.

Two important conditions for normal process

1) Summation of many continuous random variables

2) Independence of these random variables

Dynamic governing the "additive" natural system → Component-dominant dynamics



Probability distribution of the natural variability in monthly temperature anomalies for Durham, NC
Distribution of temperature anomalies (with Global Warming removed)
Normal Distribution
Probability
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-6
-4
-2
0
2
4
6
Temperature Difference from Monthly Average (C)

Scaling laws and geochemical distributions

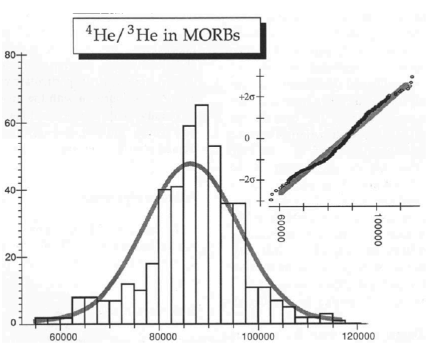
C.J. Allègre & E. Lewin, EPSL, 132, 1995

conditions of good mixing (i.e., complete stirring and homogenisation) require time and vigorous dynamics in the system.

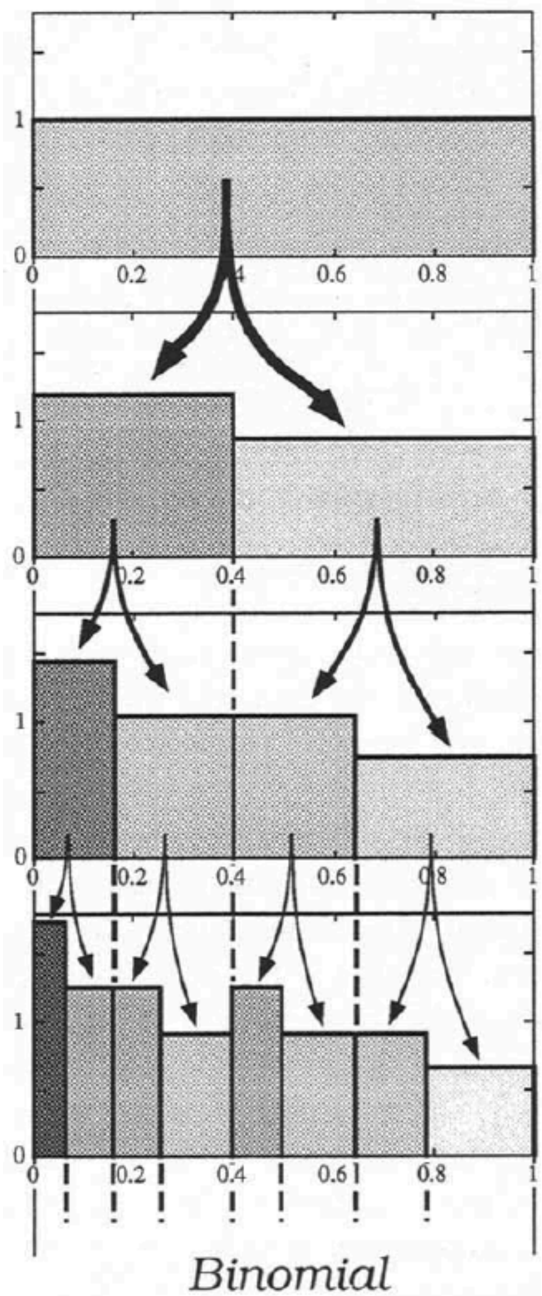
Under such conditions, one statistical theorem dominates all the reasoning about the mixing of the different distributions. This is the central limit theorem.

This theorem says that, with a rather unconstrained hypothesis, the average of an increasing number of random variables tends to follow the normal Laplace-Gauss probability distribution, mainly provided that these random variables are altogether independent.

The practical consequence of this is that the mixing of several distributions will rapidly converge after the addition of less than a tenth to the Laplace-Gauss normal distribution. This is why in Nature we observe several Gaussian distributions in multiple stage mixing situations.

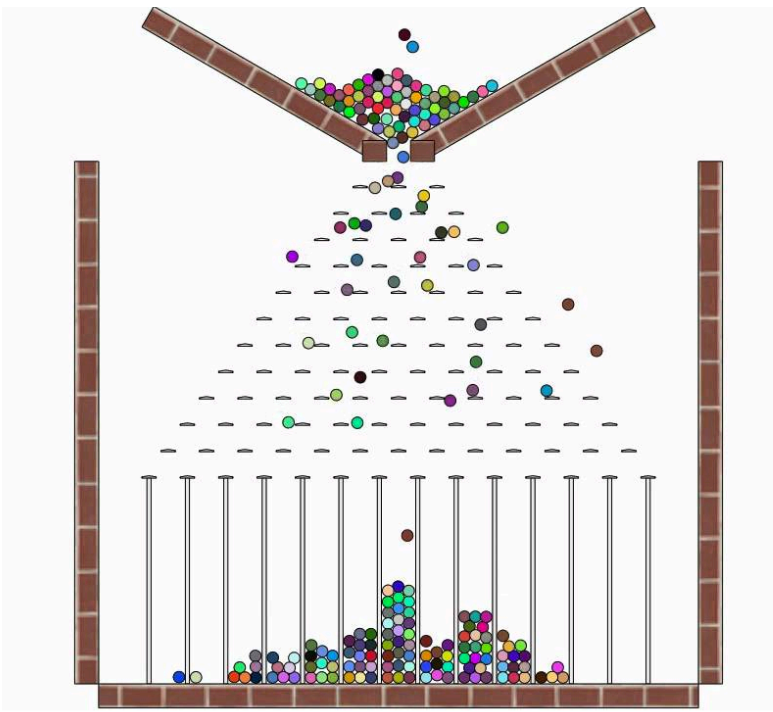


$^4He/^3He$ histogram in MORB's (44 samples from all ocean ridges, away from hotspot influence)



we will start with a crustal segment of mass M_0, being uniformly concentrated in some considered trace element with concentration C_0.

Repeated mixing operations produce normal distributions, while uncompleted mixing of a small number of components produces bimodal, or even multimodal, distributions.



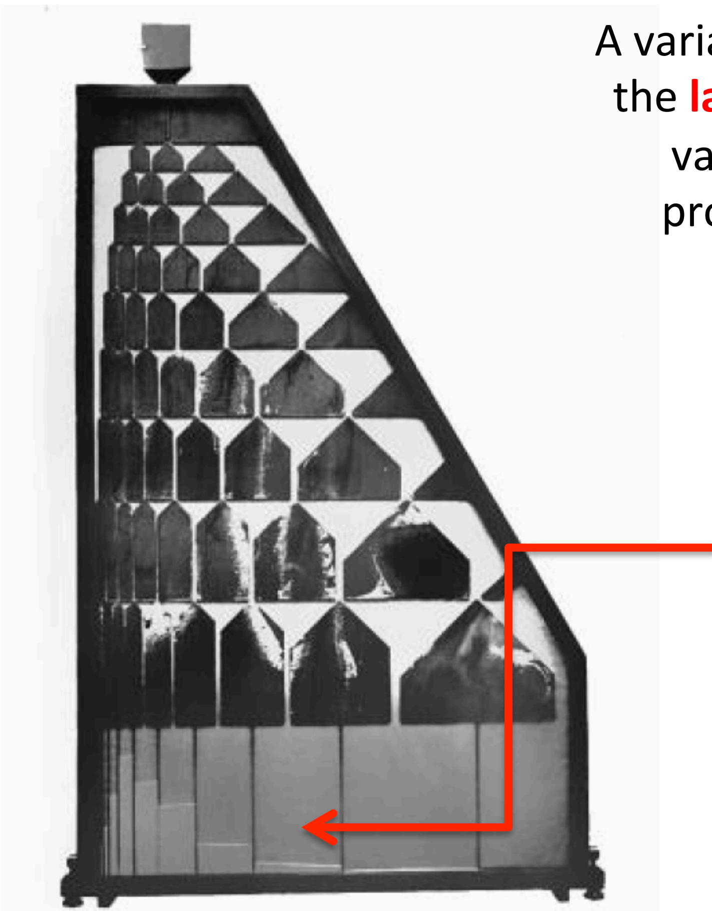
Use of the normal distribution to approximate binomial probabilities.

The Central Limit Theorem is the tool that allows us to do so…

What's the Kapteyn process?

'[The log-Normal distribution] is one of the most important classes occurring in Nature.'

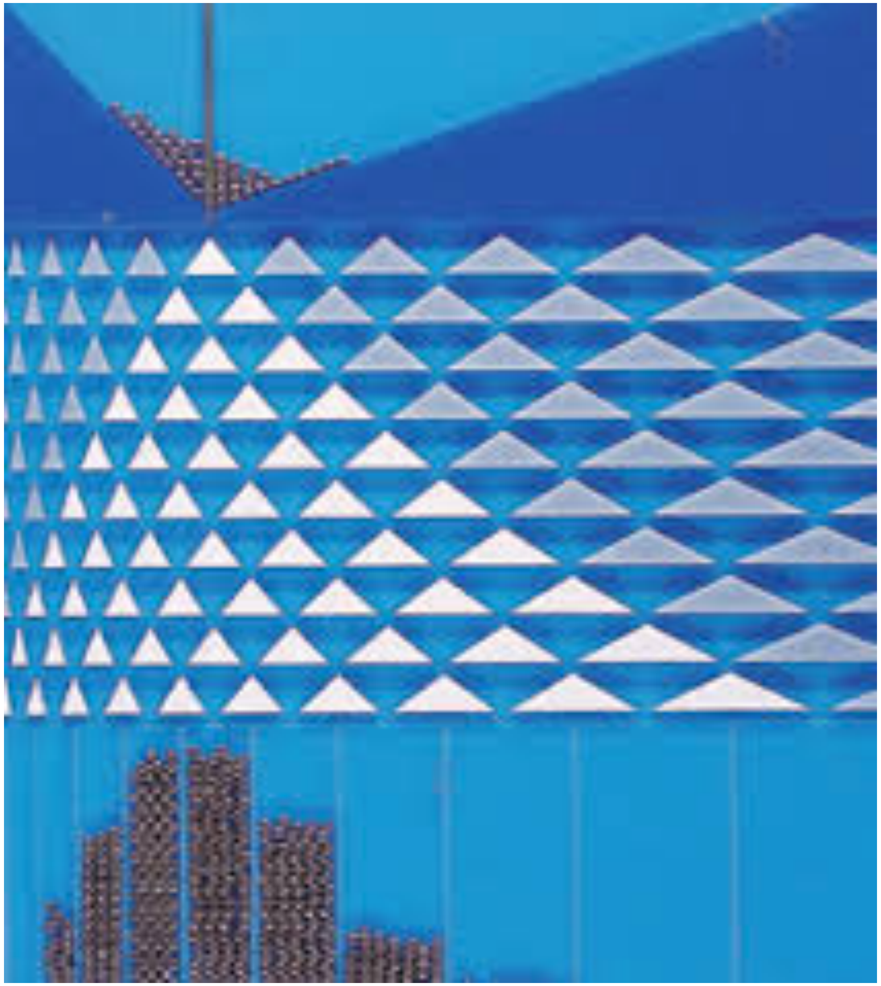
JC Kapteyn in "Skew Frequency Curves in Biology and Statistics" (1903)



A variate subject to process of change is said to obey the **law of proportionate effect** if the change in the variate at any step of the process is a random proportion of the previous value of the variate

$$X_j - X_{j-1} = \varepsilon_j X_{j-1}$$

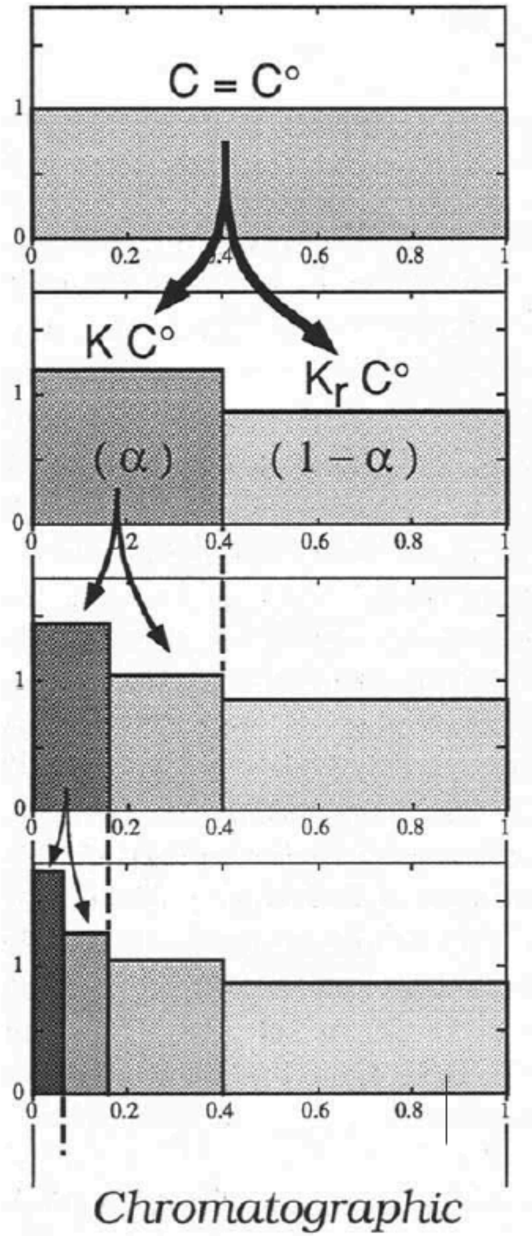
The sand finally arriving in the receptacles placed at the bottom of the machine forms a **skew histogram** approximating to that given by a two-parameter **lognormal distribution**.



If the tip of a triangle is at distance x from the left edge of the board, triangle tips to the right and to the left below it are placed at $(x \cdot c)$ and (x/c) for the log-normal, with c constant.

The distribution is generated by many small random effects (according to the central limit theorem) that are multiplicative.

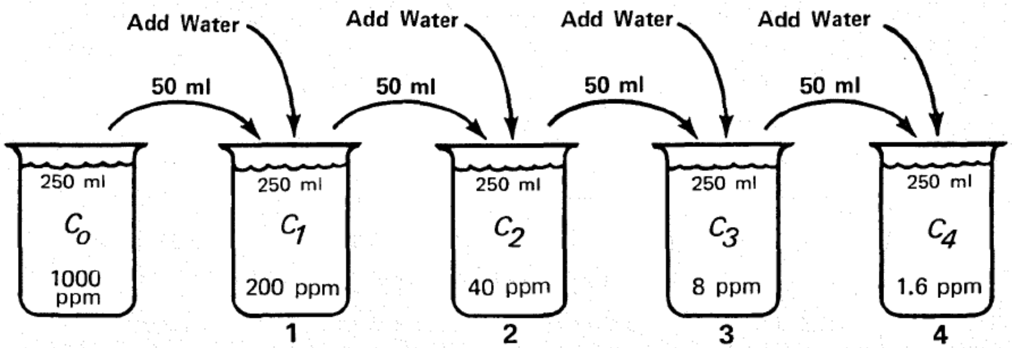
This follows from the multiplicative version of the central limit theorem, which proves that the product of many independent, identically distributed, positive random variables has approximately a log-normal distribution.



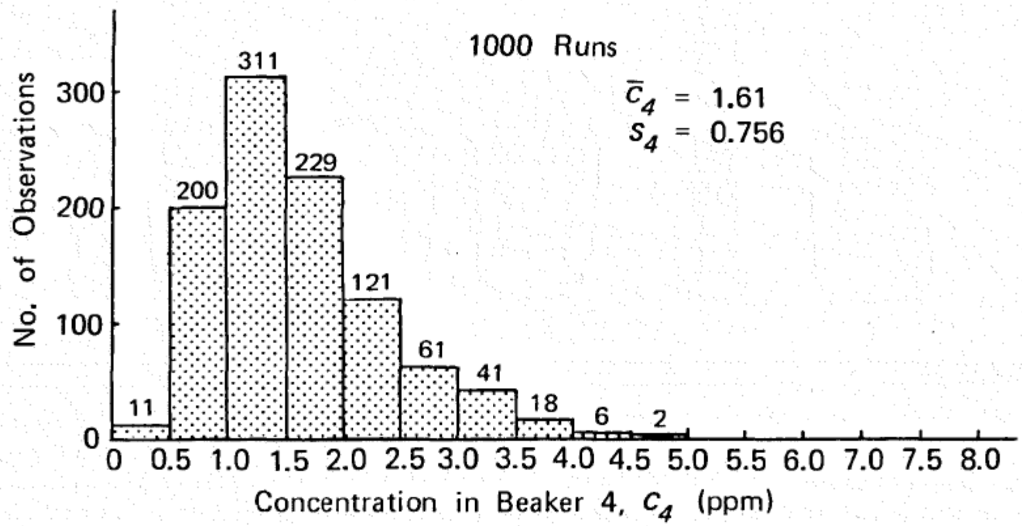
The **chromatographic enrichment**: at each level, only the most enriched segment is re-fractionated.

If we consider a portion of continental crust that has had a long geological history (e.g., having suffered successive orogenic, anatexis and high-grade metamorphism episodes) the final element concentration distributions are **log-normal**.

Indeed, seeing as continental crust is subject to episodic circulation of fluids, with dissolution, redeposition and migration, both in supergene and in deep-seated conditions, a long geological history often means a superimposition of several enrichment/ depletion episodes.



beaker experiment in which a beaker containing a pollutant at concentration c_0 undergoes successive stages of "deterministic" dilution.



Histogram of the final concentration in Beaker 4 generated by computer simulation of 1,000 successive random dilution experiments.

In general, if *m* denotes the number of successive dilutions in a successive random dilution process, then the final concentration will be the product of the initial concentration c_0 and *m* dilution factors:

$$C_m = c_o D_1 D_2 \ldots D_m = c_o \prod_{i=1}^{m} D_i$$

$$C_m = c_o D_1 D_2 \dots D_m = c_o \prod_{i=1}^{m} D_i$$

Consider the general case in which m is a large number. If we take logarithms of both sides of Equation:

$$\log C_m = \log c_o + \log D_1 + \log D_2 + \dots + \log D_m = \log c_o + \sum_{i=1}^{m} \log D_i$$

logarithm of the final concentration

sum of log c_0 and the logarithms of the m dilution factors. If the dilution factors $D_1, D_2, \dots, D_m$ are independent random variables, then the logarithms of the dilution factors also will be independent random variables.

By the additive form of the Central Limit Theorem, the right-hand side of Equation is approximately normally distributed, since it is the sum of *m* independent random variables.

This conclusion **does not depend** on the nature of the individual distributions of the dilution factors D_i; they may have the same distribution or different distributions. Since log C_m is approximately normally distributed, then C_m will be approximately lognormally distributed.

This analysis parallels development of the **Law of Proportionate Effect**, which initially was proposed by Kapteyn to explain the lognormal distribution for observed biological variables.

A **lognormal process** is one in which the random variable of interest results from the product of many independent random variables multiplied together.

Interaction-dominant dynamics are associate with systems that entail tightly coupled processes spanning a wide range of temporal or spatial scales, including fractal systems.

They refer to systems that entail **multiplicative** and/or **interdependent feedback transactions** among the processes that govern the system dynamic's.

Geochimica et Cosmochimica Acta, 1954, Vol. 5, pp. 49 to 73. Pergamon Press Ltd., London

The lognormal distribution of the elements[1]

(A fundamental law of geochemistry and its subsidiary)

L. H. AHRENS

Department of Geology and Geophysics, Massachusetts Institute of Technology[2]

(*Received* 3 *November* 1953)

ABSTRACT

Frequency distribution plots of K, Rb, Sc, V, Co, Ga, Cr, and Zr in Ontario diabase, Sc, V, Ga, Cr, La, and Zr in Canadian granite, K, Rb, and Cs in New England granite and F and Mo in granite from various localities are regular, but assume decided positive skewness when dispersion is large; hence, distribution of concentration is not normal. All distributions become normal, or nearly so, provided the variate (concentration of an element) is transformed to log concentration: this leads to a statement of a fundamental (lognormal) law concerning the nature of the distribution of the concentration of an element in specific igneous rocks.

A subsidiary law concerning the relationship between averages and most prevalent concentrations follows as a direct consequence of the fundamental law.

Dispersions of different elements ca
lognormal law. A comparison of the d
the strikingly high uniformity of abund
show a totally different magnitude of
scandium is small in diabase and extre

INTERNATIONAL ASSOCIATION OF GEOCHEMISTRY AND COSMOCHEMISTRY

IAGC NEWSLETTER - NO. 3, OCTOBER 1973

Andrei Borisovich Vistelius (1915–1995), along with William Christian Krumbein (1902–1979) and John Cedric Griffiths (1912–1992), were dominant figures in the formative and development years of mathematical (or quantitative) geology as a subdiscipline of geology.

THE JOURNAL OF GEOLOGY

January 1960

THE SKEW FREQUENCY DISTRIBUTIONS AND THE FUNDAMENTAL LAW OF THE GEOCHEMICAL PROCESSES[1]

ANDREW B. VISTELIUS
Laboratory of Aeromethods, Academy of Sciences U.S.S.R., Leningrad

ABSTRACT

The frequency distribution of concentrations of chemical elements has been investigated. Modern statistical methods for the analysis of large and small samples have been used. As a result of the work, the formulation of the fundamental law of the geochemical processes is proposed. This law can be applied in many cases to the investigation of deposits. The author points out the difference between "the distribution of the concentrations" in V. M. Goldschmidt's sense and "the probability distribution of the concentrations" of the present paper.

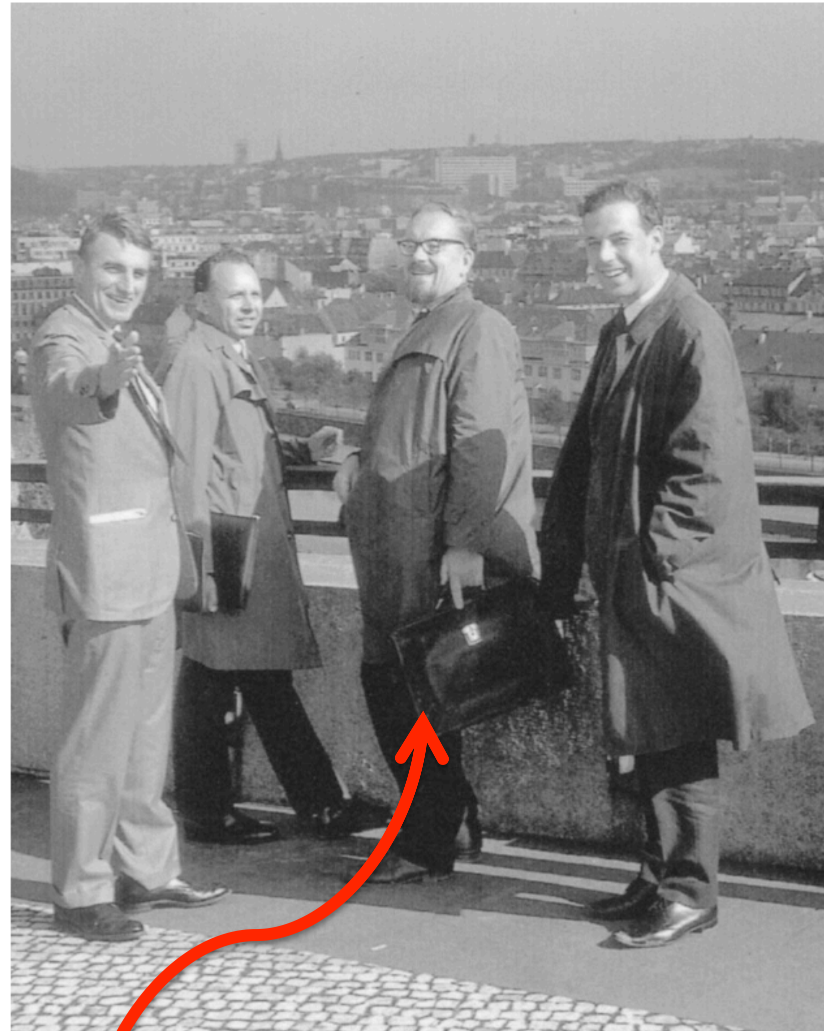
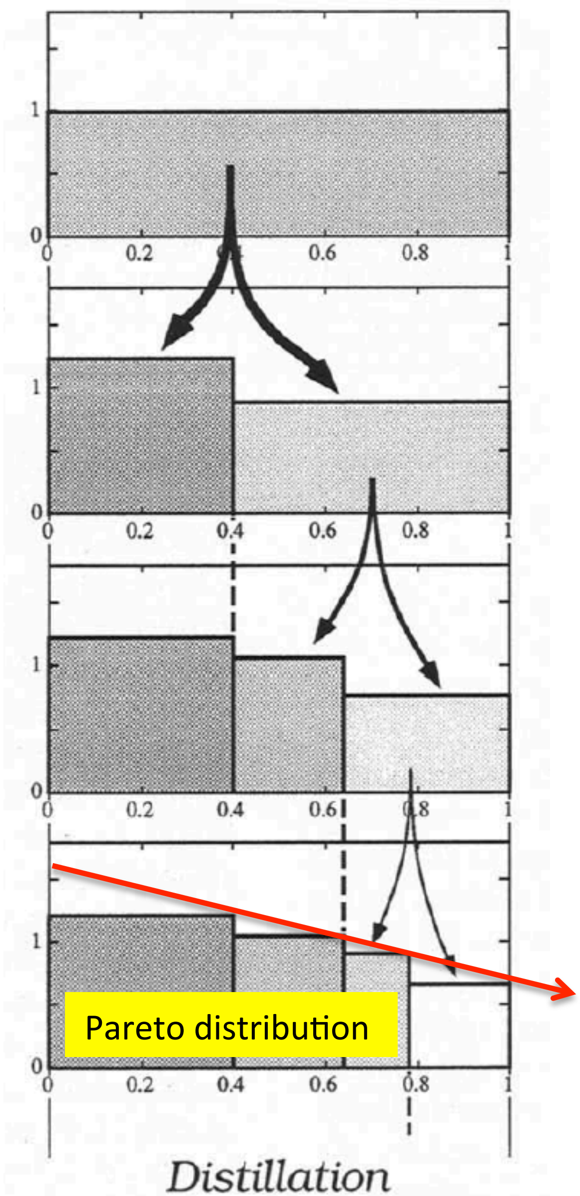


Figure 2. Prague, Czechoslovkia prior to start of IGC in 1968. From left to right: John Harbaugh (USA), B. Soukup (Czechoslovkia), Vistelius (USSR), and Frits Agterberg (Canada).



The distillation process: at each level, only the 'residual' segment is refractionated.

classical model used in **geochemistry** in order to explain the extreme enrichments observed for some trace elements, especially in crystalline rocks.

The background of the model is a **closed system** in which a material crystallises from a melt. Trace elements are partitioned between the melt and the crystallising solid.

Considering one of these trace elements known as incompatible (i.e., they are enriched in the melt relative to the crystallising solid), the successive melts become progressively more enriched.

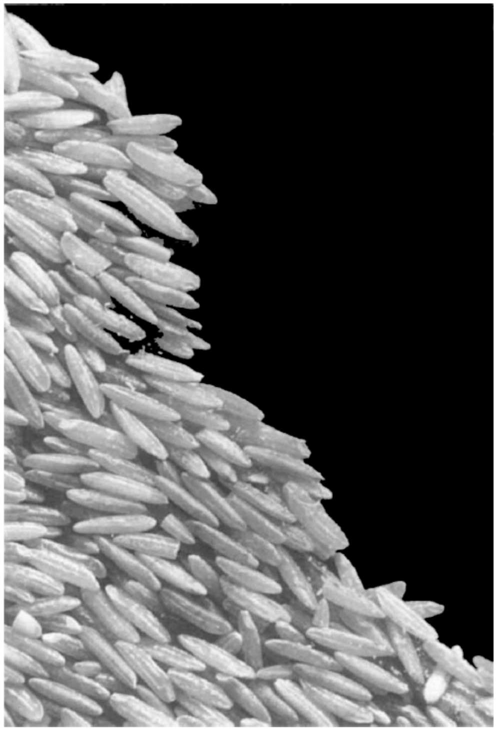
An **inverse power-law distribution** is a so-called ***heavy-tailed distribution***; the heavy tail represents large magnitude, but rare events.

Thus, it expresses a salient positive skew. If the extreme right tail of an event distribution decays as a power function, then the probability of observing a particular event magnitude, *p(x)*, is the inverse of the *x* value itself, raised to the scaling exponent α (alpha) that is $p(x) \approx x^{-\alpha}$.

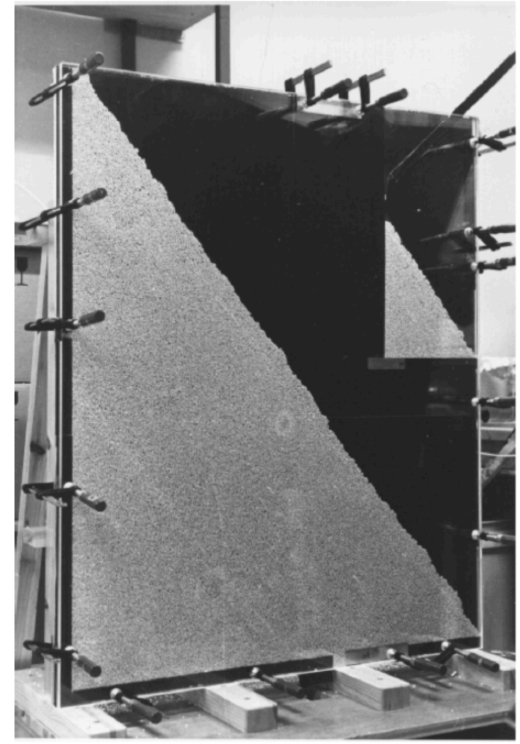
The formal mathematical equation of the inverse power-law probability density function is $p(x) = b \cdot x^{-\alpha}$, where *b* is a positive constant. The scaling exponent **α** quantifies the **rate of decay** of the distribution's tail.

Circular, interdependent feedback transactions likely govern systems that express inverse power-law scaling. Moreover, power law behavior is symptomatic of self-organizing physical systems poised near a critical point.

One of several model systems for studying the behavior of self-organized and critical systems is a simple rice pile…

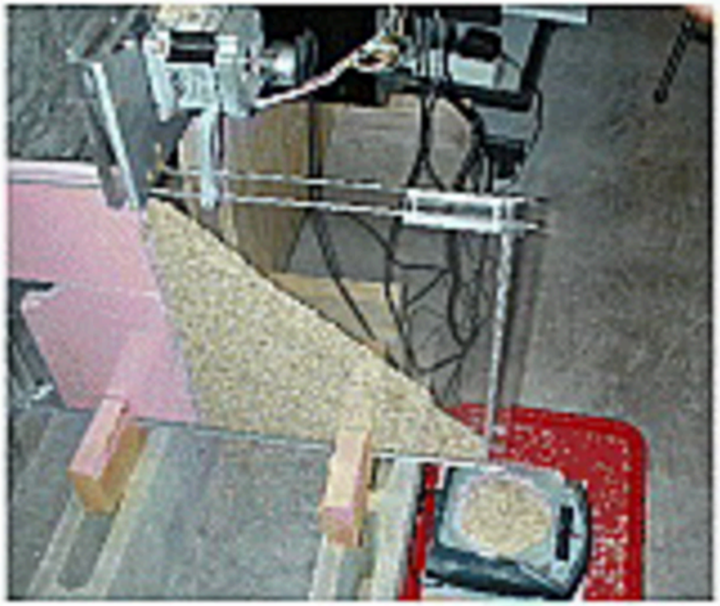


Actual rice pile experiments use an apparatus that makes detailed measurements of rice grain activity, as kernels are continuously added to and exit the pile.



Initially, small, localized piles emerge within the larger pile. As the local piles grow, avalanches unfold. At a critical point, a holistic coordinative balance emerges throughout the system. The balance is governed by two competing sources of constraint: friction and inertia

From that point on, the rice pile maintains a time-invariant organization, even in the face of the constant perturbation induced by the intermittent clusters of inflowing and avalanching rice.



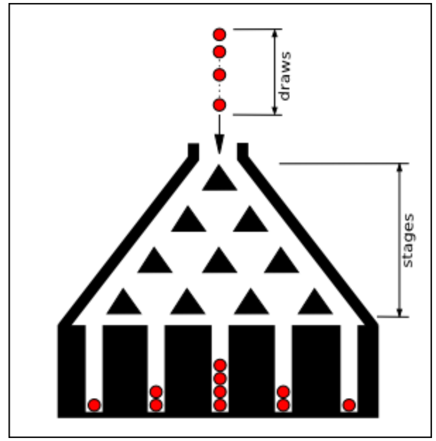
When a rice pile is in a critical regime the effects of perturbation are no longer proportional to the size of the perturbation—adding one new grain might result in no change, a tiny avalanche, or a large avalanche, affecting the entire pile.

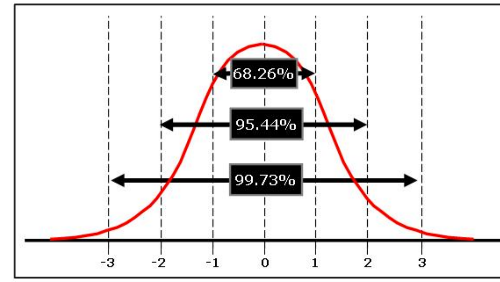
In the long run, small avalanches occur frequently and occasional very large avalanches unfold, all the while the pile maintains a time-invariant average height and slope. An inverse power-law distribution neatly summarizes the relationship between the avalanche magnitudes (indexed by grain counts) and their frequency of occurrence.

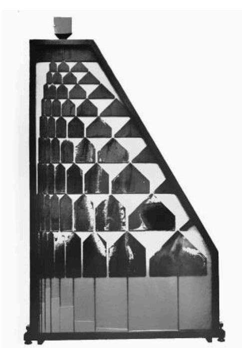
By way of summary, the model rice pile system only reaches a **critical state** when certain grain size and smoothness requirements are met. For instance, if one adds a constraint that **changes the balance** between inertia and friction so that one or the other term dominates the interactions, the empirical consequences of feedback are minimized, and the rice pile converges on a characteristic relaxation time. Systems in which the **effects of feedback** are negligible but that are still governed by multiplicative interactions exhibit **lognormal** instead of **power law** behaviour.

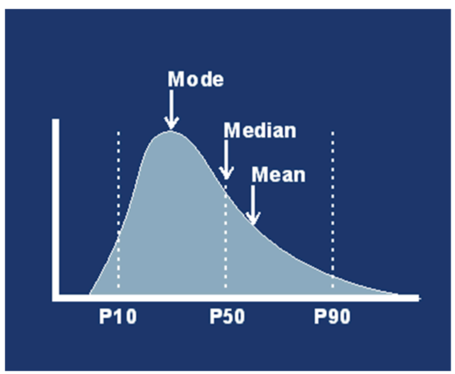
One may envision a loose continuum of ideal distributions spanning the general taxonomy of **component-dominant** and **interaction-dominant dynamics**.

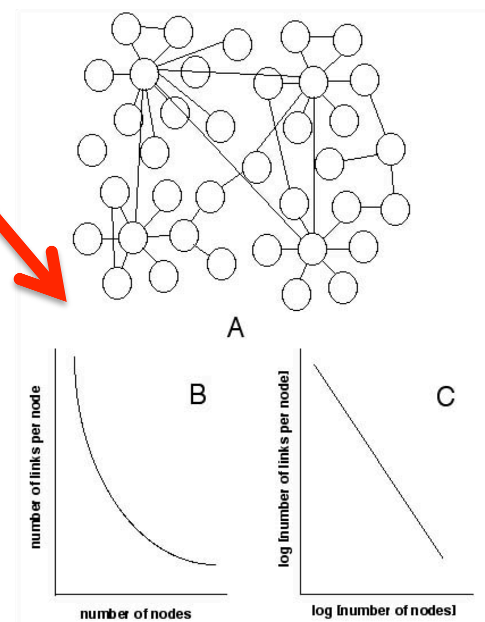
At one extreme, there is the **Gaussian distribution**, signature of weak unsystematic additive interactions among independent random variables.











At the other extreme, there is the **heavy-tailed inverse power-law**, the signature distribution of interdependent feedback dynamics. The moderately skewed lognormal stands between these two extremes; it arises from multiplicative interactions among independent variables.