Creating Mathematics That Describes Brain and Information

(1) the structure of information: surprise, probability, and geometry

RIKEN
Brain Science Institute

Shunichi Amari

Mathematics is a culture of humankind

Babylonian Mathematics

Arabic, Chinese, Greek, Japanese Mathematics

Euclidean geometry

the brain structure

the concept of number- animal, chimpanzee

geography- mouse, geometry

-human: abstract thinking

The Brain Structure

-the pleasure of thinking things out

abstract thinking

games, puzzles

mathematics, culture

The science and technology in the 21st century

What is human?

What is society?

- Physics
- Mathematical science
- Biology
- Information science
- Human science

fusion

Mathematical thinking

Physics- a huge empire

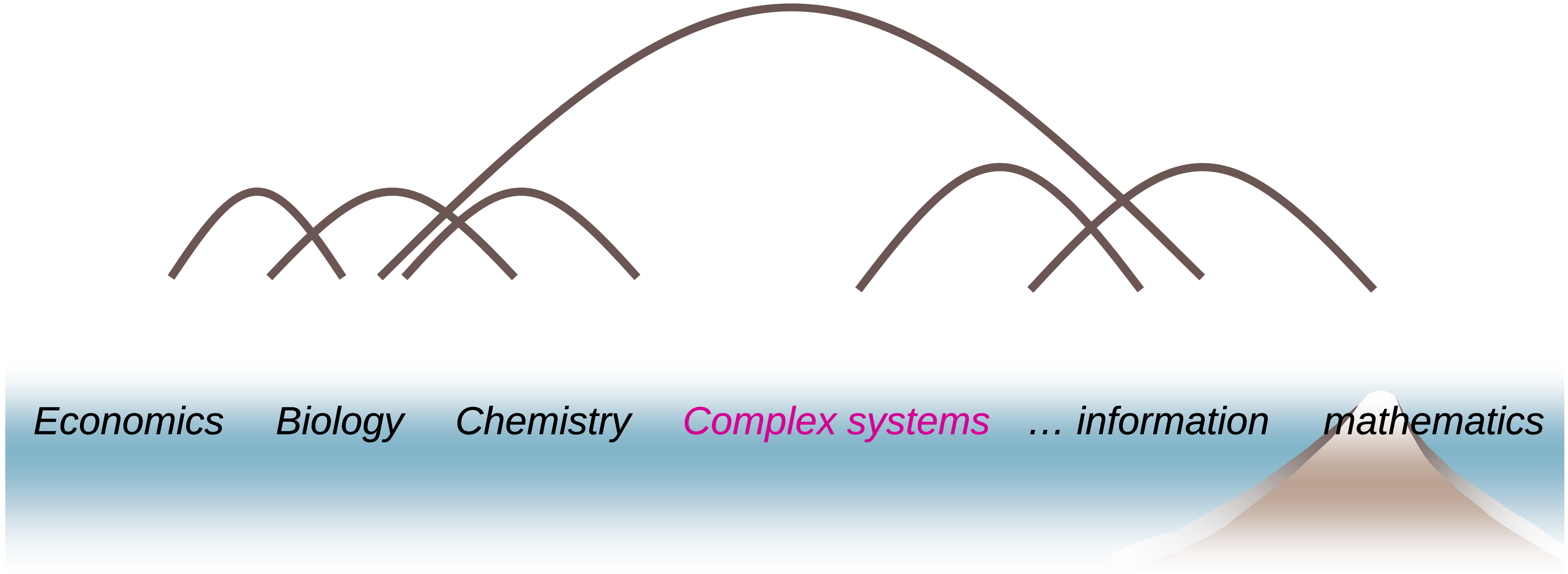
Theory of relativity

Quantum mechanics

Condensed matter physics

nanotechnology

electronics



Biology —— Reductionism

Natural history *Molecular biology* *Genes*
Biotechnology

Diversity of species *Structure and function*

Fundamental principles of life

Brain / System biology

Information science/engineering
— Turing and Gögel's Curse

Telephone
radio

Information theory
cybernetics

Internet

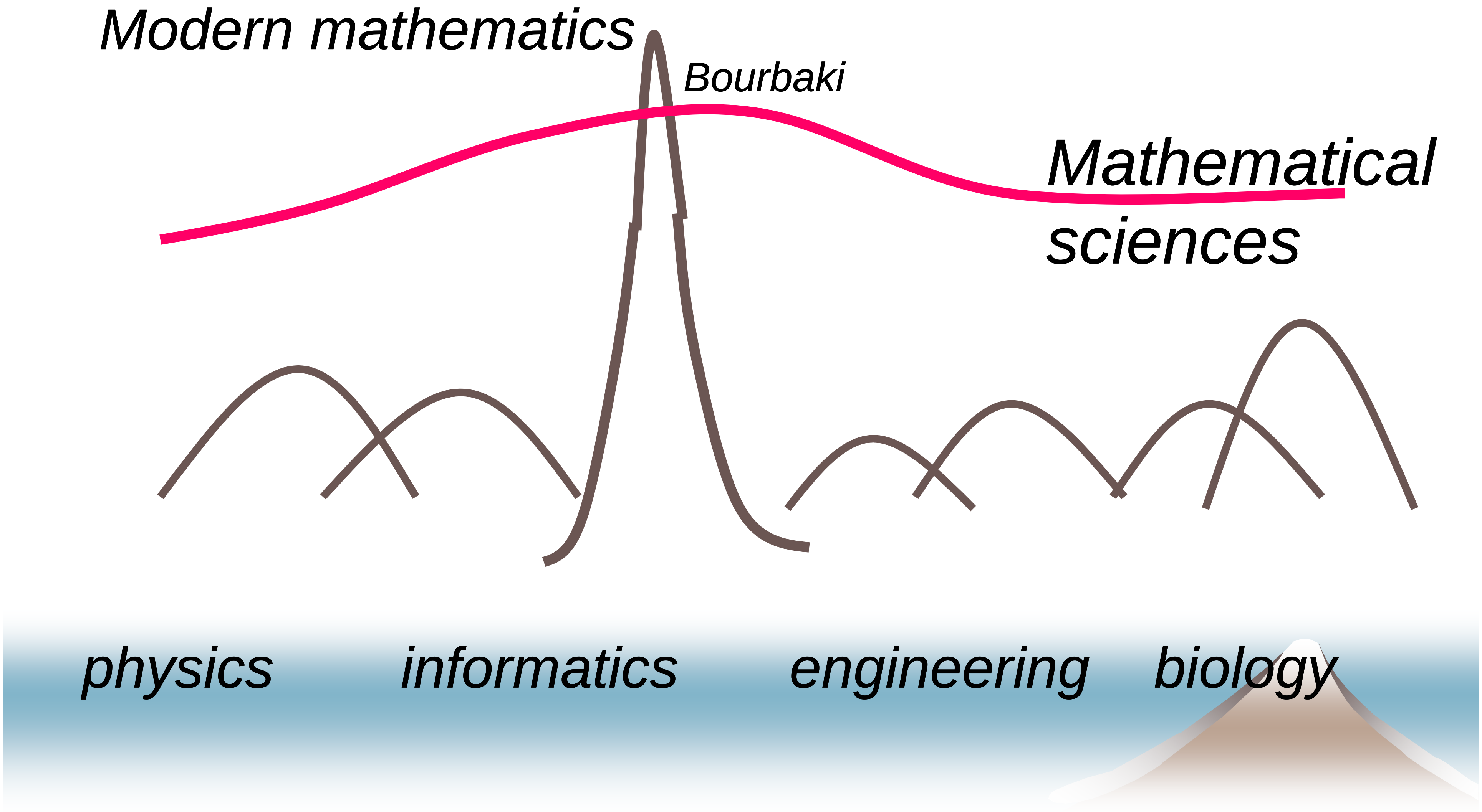
Information society

Vacuum tube

Computer
Transistor

Artificial Intelligence ,Robots,
Brain, Bioinformatics

Mathematics- an ivory tower



The mathematical principle of information: the amount of information

A or B? -1 bit

A,B,C… : k characters

f(k) bits

The additivity of information

m

xxxxxxx

k xxxxxxx

xxxxxxx

$$yf'(xy) = f'(x)$$

$$yf'(y) = c$$

$$f'(y) = c / y$$

f(km)= f(k) + f(m)

f(xy) = f(x) + f(y)

f(k) = log k

Information and probability

p = 1/k

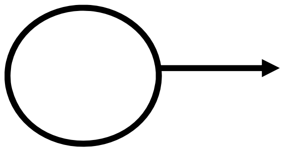
f(p) = -log p = log k

x x x x x x x x x x x x   n

m

$$I(p) = I\left(\frac{m}{n}\right) = \log n - \log m = -\log p$$

information, probability

Information source  $x_1 x_2 x_3 \mathrm{L}$

Information entropy $p_i = \mathrm{Prob}\{x = i\}$

$$H = -\sum p_i \log p_i$$

Information contraction

$$p = \mathrm{Prob}\{x = 1\}$$

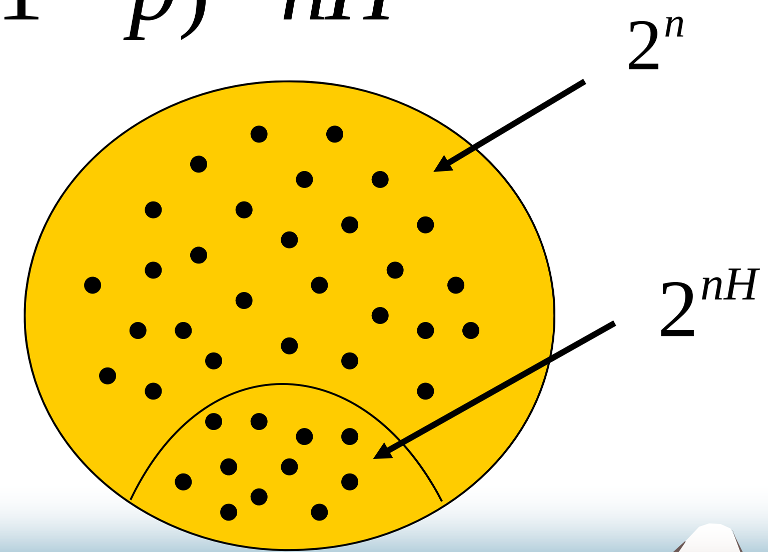
$$\underbrace{011001110101}_{n}$$

entropy

$$H = -p\log p - (1-p)\log(1-p) \quad nH$$

2^n $\quad$ 2^{nH} Frequently appear

$$\log 2^{nH} = nH \text{ bits}$$



Law of large numbers

$$0101100110\mathrm{L} \rightarrow \#1 = np; \quad n = 10,000; p = 0.3$$

$$p = \mathrm{Prob}\{x = 1\} \quad H = -p\log p - (1-p)\log(1-p)$$

$$\sum x_i = np \pm A_n$$

$$A_n \rightarrow 0, c, \infty ?$$

$$\sum x_i = np \pm \frac{\sqrt{n}}{2}$$

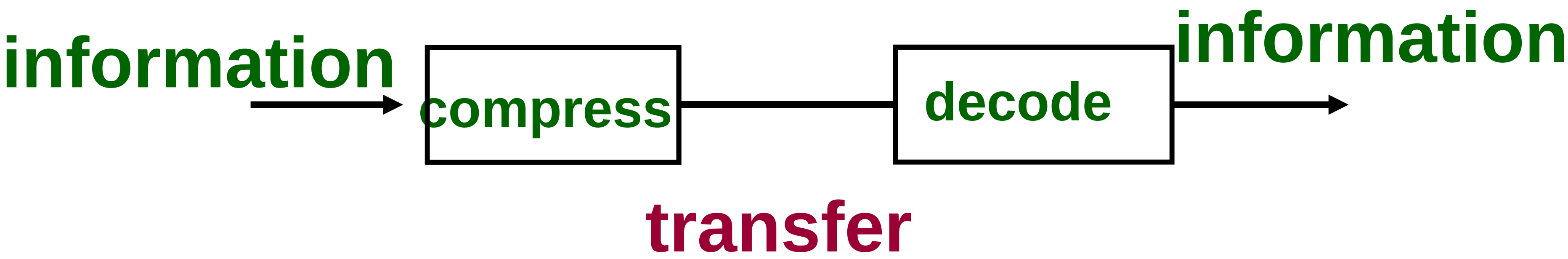
$$ {}_{n}C_{np} = \frac{(np)!(n-np)!}{n!} $$

$$ n! = \left(\frac{n}{e}\right)^{n} \sqrt{2\pi n} \approx n^{n} $$

$$ {}_{n}C_{np} = 2^{nH} $$

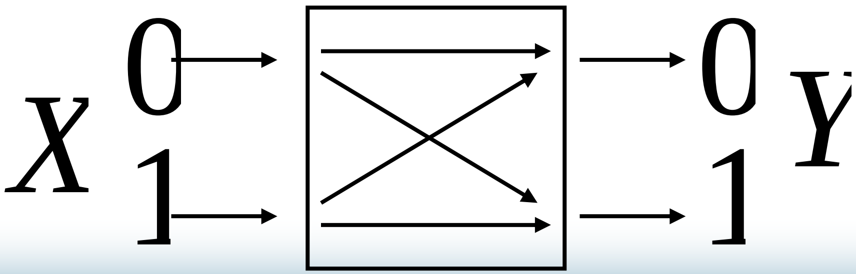
Information transfer

facsimile, image, audio,…



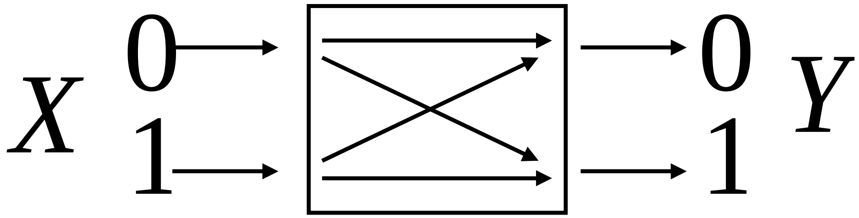
Information transfer: speech path

noise: error



$$\begin{cases} 0 \longrightarrow 1 & q \\ 1 \longrightarrow 0 & q \end{cases}$$

Mutual information



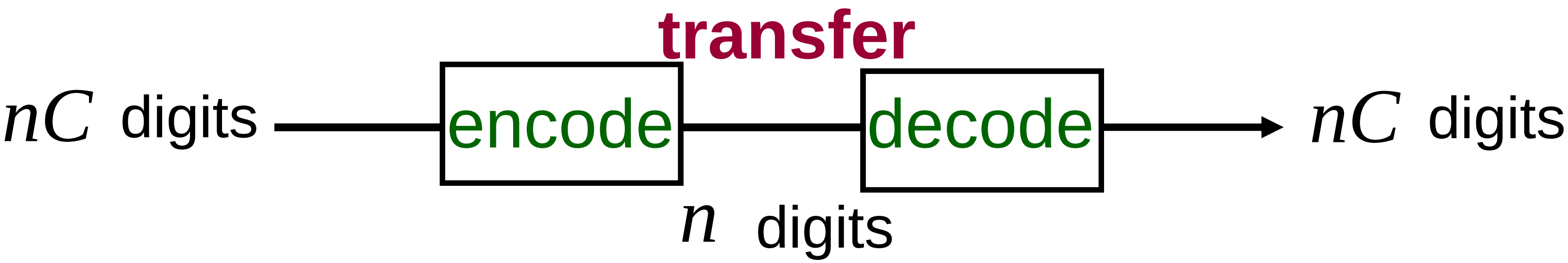
$$I(X,Y) = H(X) - H(X|Y)$$
$$= H(X) + H(Y) - H(X,Y)$$

Information capacity: $C = 1 - H(q)$

Shannon's great discovery: transfer without errors

$010\cdots0$ n digits

$010\cdots0$ nC digits

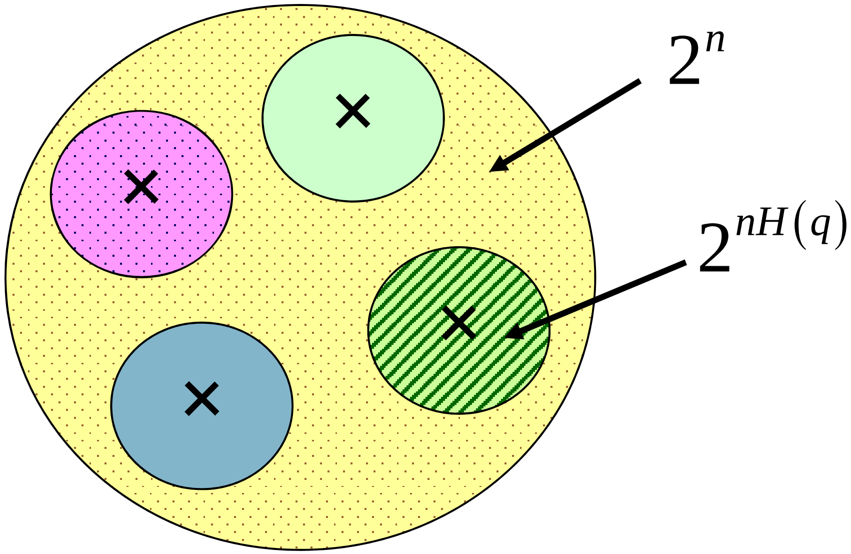


The Probability of error is "0" !

Signal and code

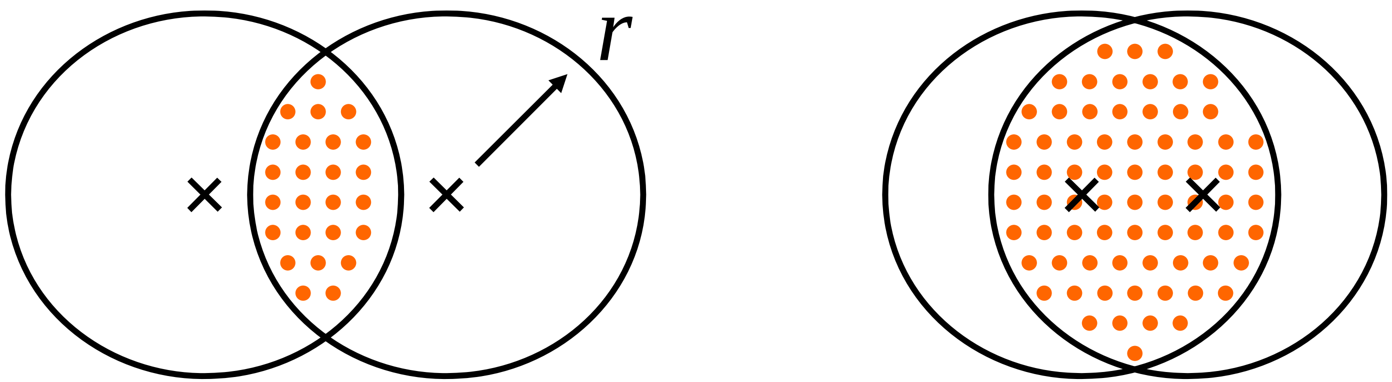
$$\frac{2^n}{2^{nH}} = 2^{n(1-H)}$$

$$= 2^{nC}$$



N-dimensional ball and intersection

$$\frac{(r+\varepsilon)^n}{r^n} = \frac{r^n + n\varepsilon r^{n-1}}{r^n} = 1 + \frac{n\varepsilon}{r} \qquad V = c_n r^n$$



Error correcting code：Hamming code

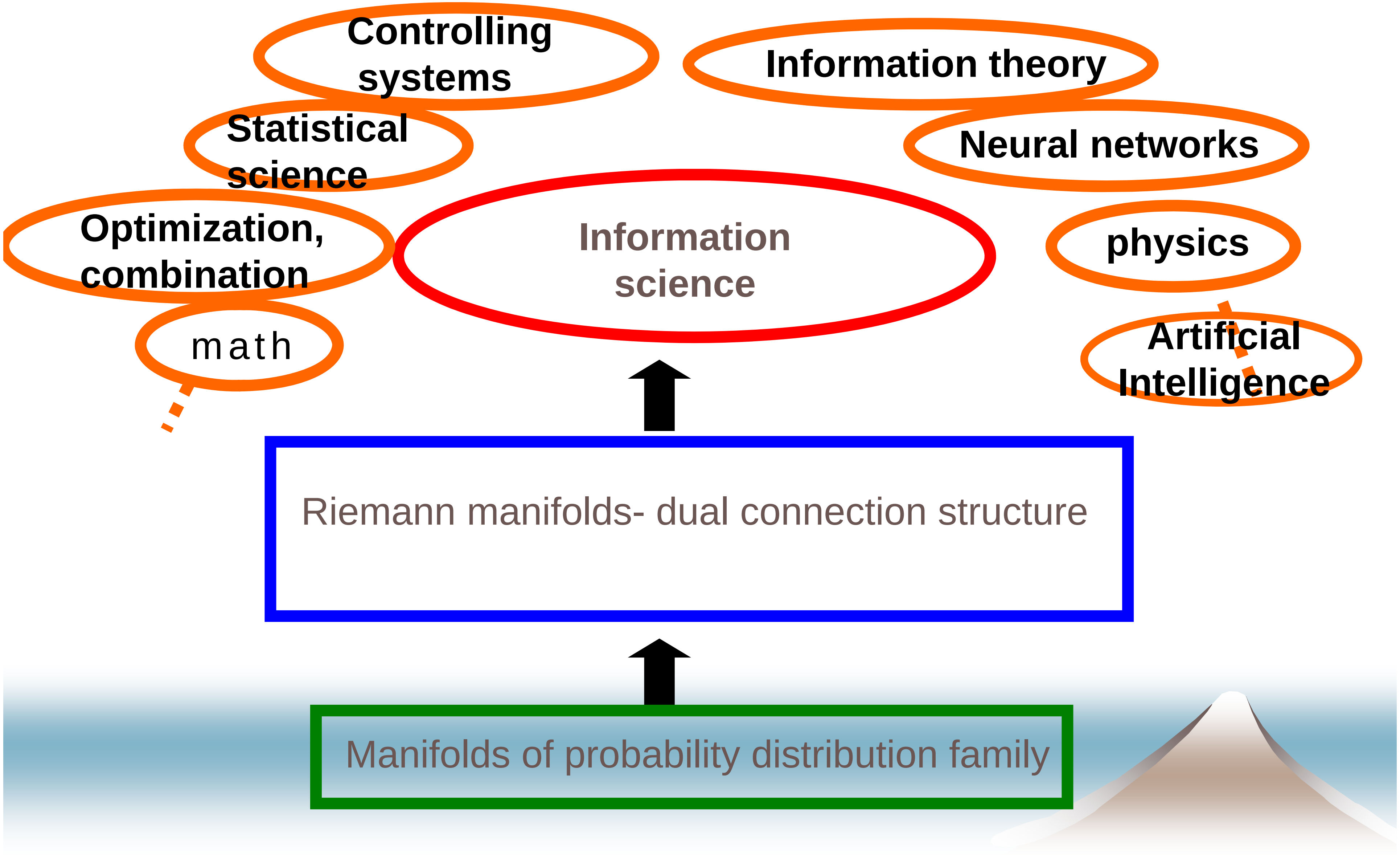
Computers, CDs, …

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_7 \end{bmatrix} = 0$$

$$2^4 \to 2^7$$

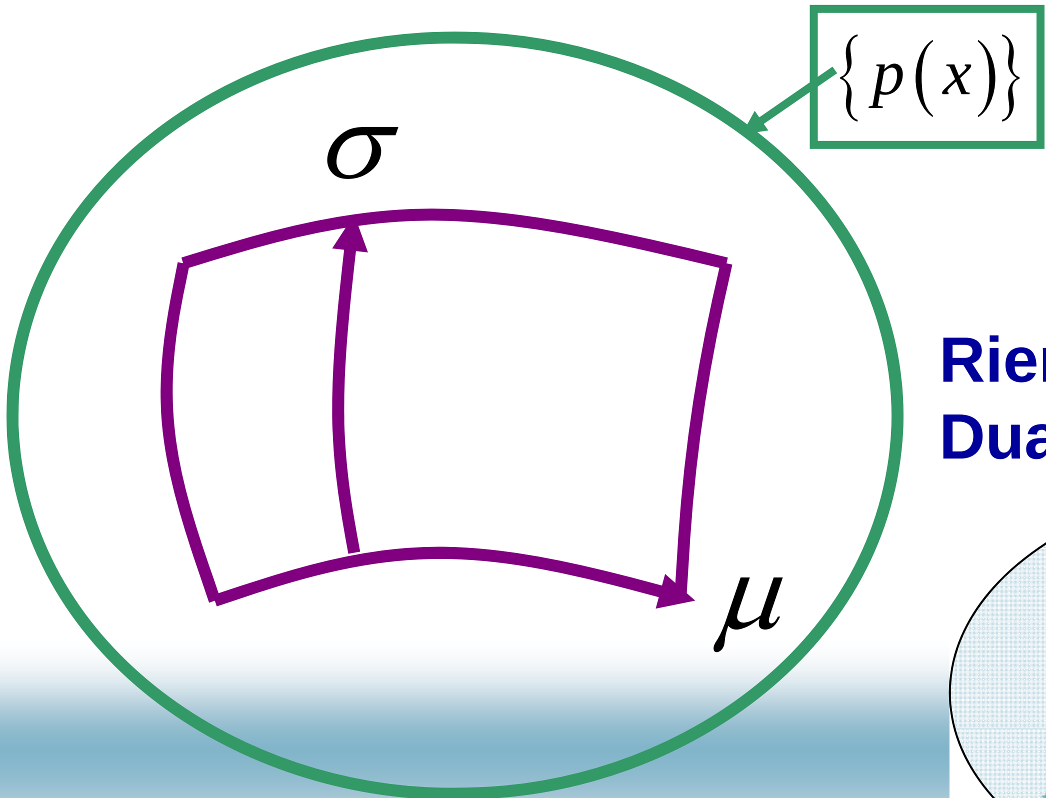
Galois field, algebraic geometry codes

Information geometry



What is Information geometry?

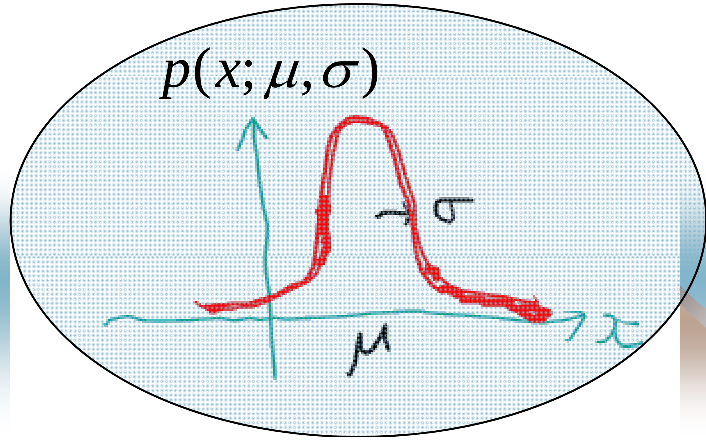
$$S = \{ p(x;\mu,\sigma) \} \qquad p(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$$



$$S = \{ p(x;\theta) \}$$

Riemann geometry

Dual affine connection



The invariance principle: $S=\left\{p\left(x,\theta\right)\right\}$

1. ***Independent of the parameterization***

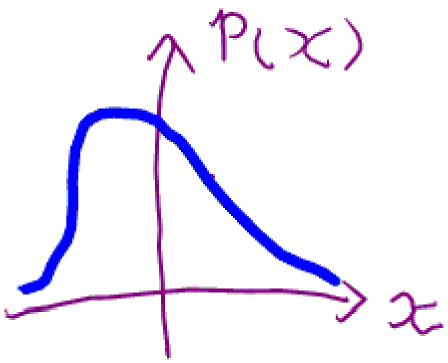
$$\xi=\xi\left(\theta\right),\quad \overline{p}\left(x,\xi\right)\qquad D=\sum\theta_i^2\ \neq\ \sum\xi_i^2$$

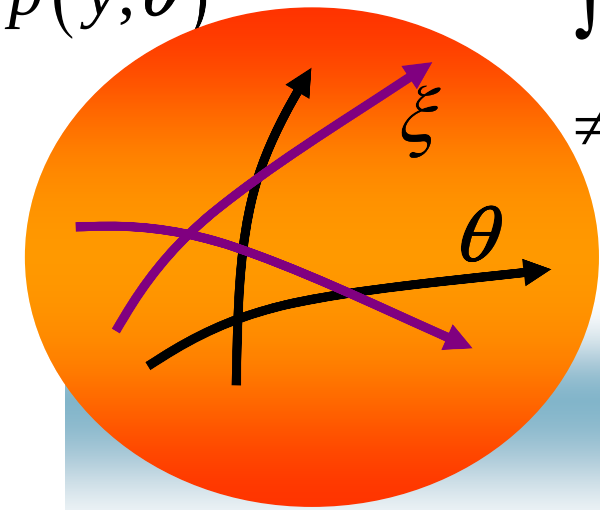
2. ***Independent of the apparent scale of the probability variable***

$$y=y\left(x\right),\qquad \overline{p}\left(y,\theta\right)$$

$$\int\left|p\left(x,\theta_1\right)-p\left(x,\theta_2\right)\right|^2dx$$

$$\neq\int\left|\,\overline{p}(y,\theta_1)-\overline{p}(y,\theta_2)\,\right|^2dy$$



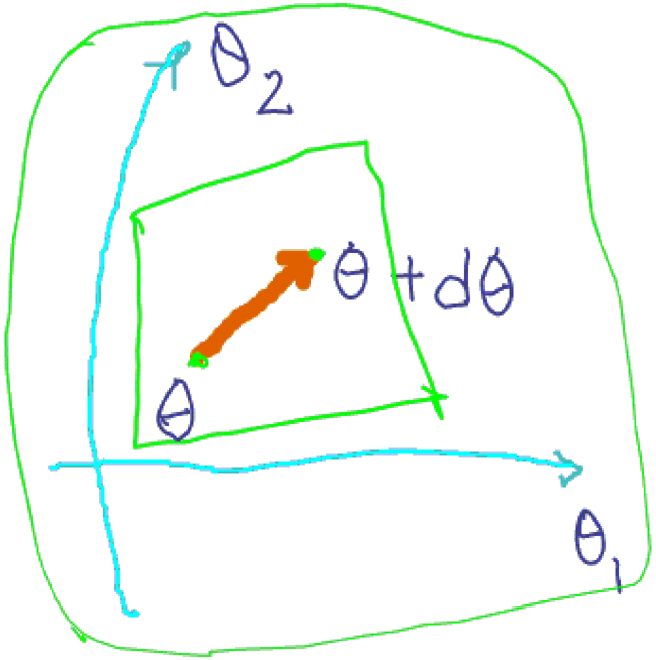


The two geometrical structure

Riemannian and affine connection

$$ds^2 = \sum g_{ij}(\boldsymbol{\theta})\, d\theta_i d\theta_j$$

$$ds^2 = \frac{1}{2} D\left[p(x, \boldsymbol{\theta}) : p(x, \boldsymbol{\theta} + d\boldsymbol{\theta}) \right]$$



Fisher information

$$g_{ij} = E\left[\frac{\partial}{\partial \theta_i} \log p \, \frac{\partial}{\partial \theta_j} \log p \right]$$

Affine connection

affine connection

covariant derivative

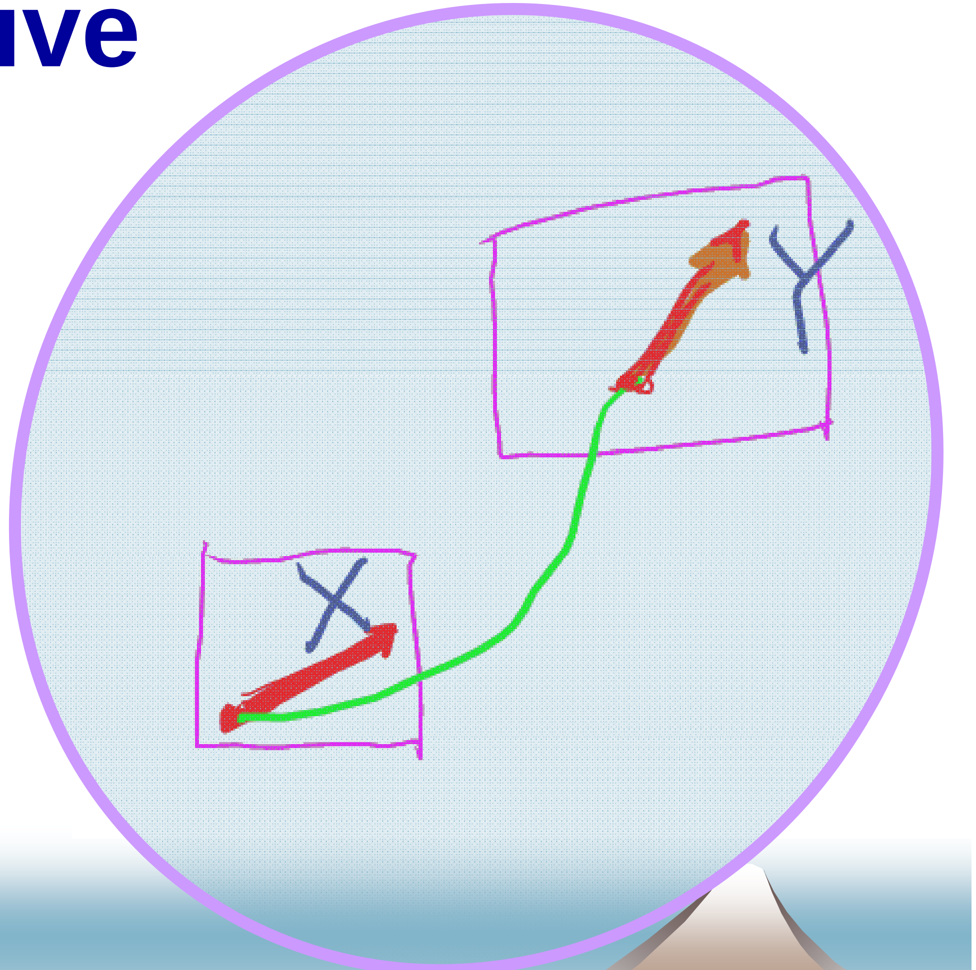
$$\Pi_c X = Y$$

geodesics $\Pi \dot{X} = \dot{X} \quad X = X(t)$

$$s = \int \sqrt{\sum g_{ij}(\theta) d\theta^i} \, d\theta^j$$

minimal distance

straight line



Geometry

Euclid

Bolyai, Lobachevsky

Gauss

Riemann

Einstein

String theory

Two affine connections (∇, ∇^*)

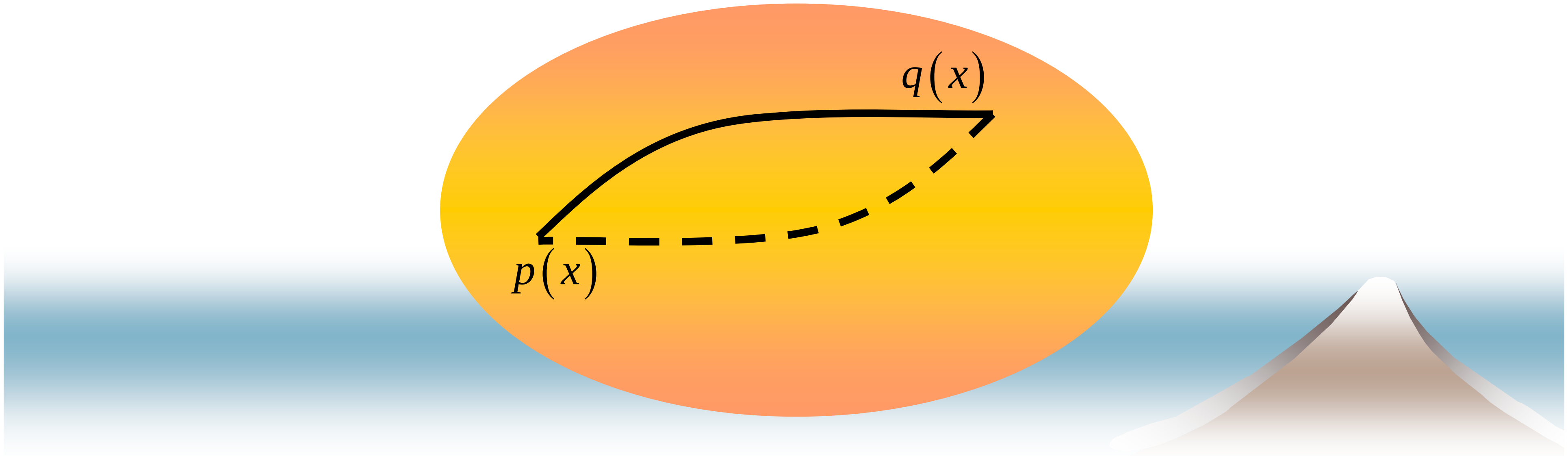
$$(\Pi, \Pi^*)$$

e-geodesic

$$\log r(x,t) = t \log p(x) + (1-t) \log q(x) + c(t)$$

m-geodesic

$$r(x,t) = tp(x) + (1-t)q(t)$$



Neuronal firing

$$x_1 \quad x_2 \quad x_3 \quad \cdots\cdots \quad x_n$$

$$p(\mathbf{x}) = p(x_1, x_2, \ldots, x_n), \quad x_\mathrm{i} = 0,1$$

$\eta_i = E[x_i]$ ---- firing rate

$v_{ij} = Cov[x_i, x_j]$ ----correlation

What is higher-order correlations?
Are they orthogonally decomposable?

Probabilistic inference

$$p(x, y, z, r, s)$$

$$p(x, y, z \mid r, s)$$

$$x, y, z, \ldots = 1, -1$$