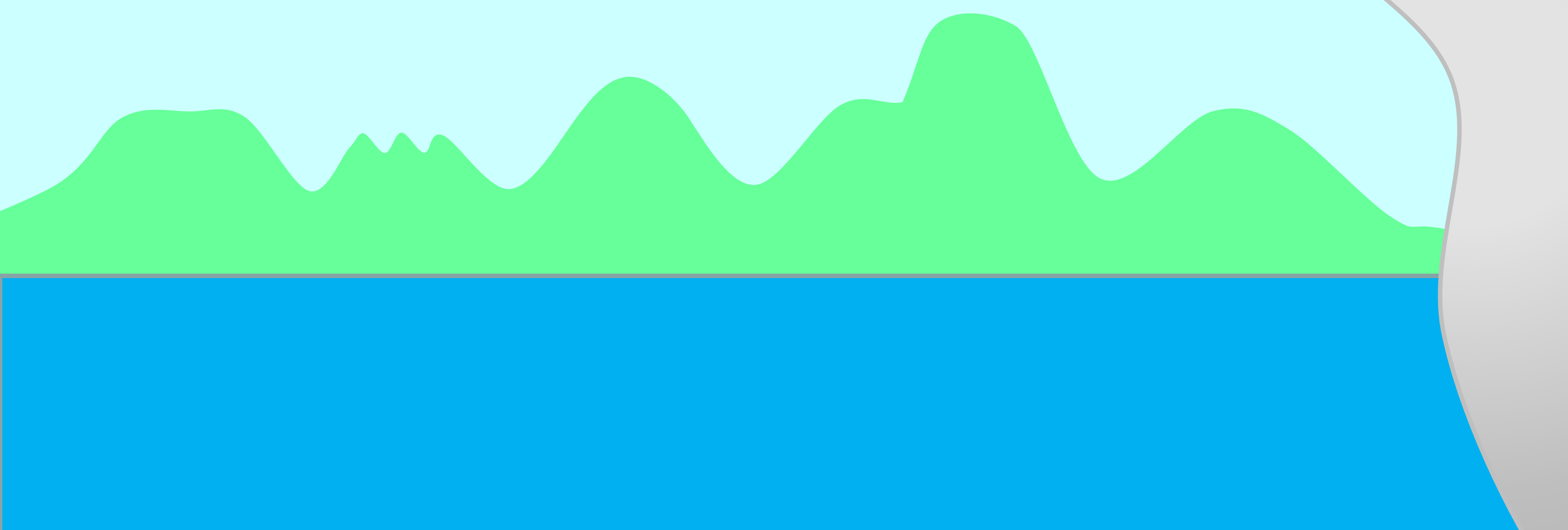
学習理論よ 何処へ

情報論的学習理論と機械学習研究会

東京大学安田講堂 2017年11月10日

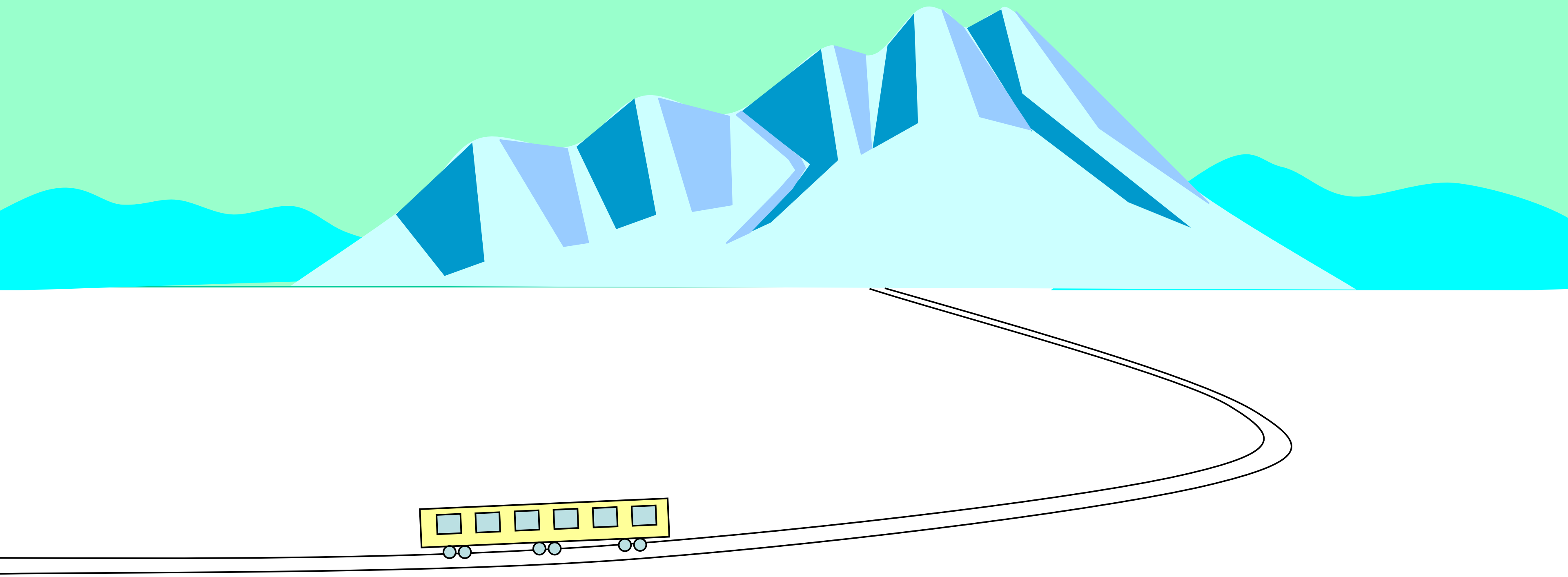
この講演では東京大学・佐藤一誠先生に
お世話いただきました．御礼申し上げます．

東京工業大学　渡辺澄夫

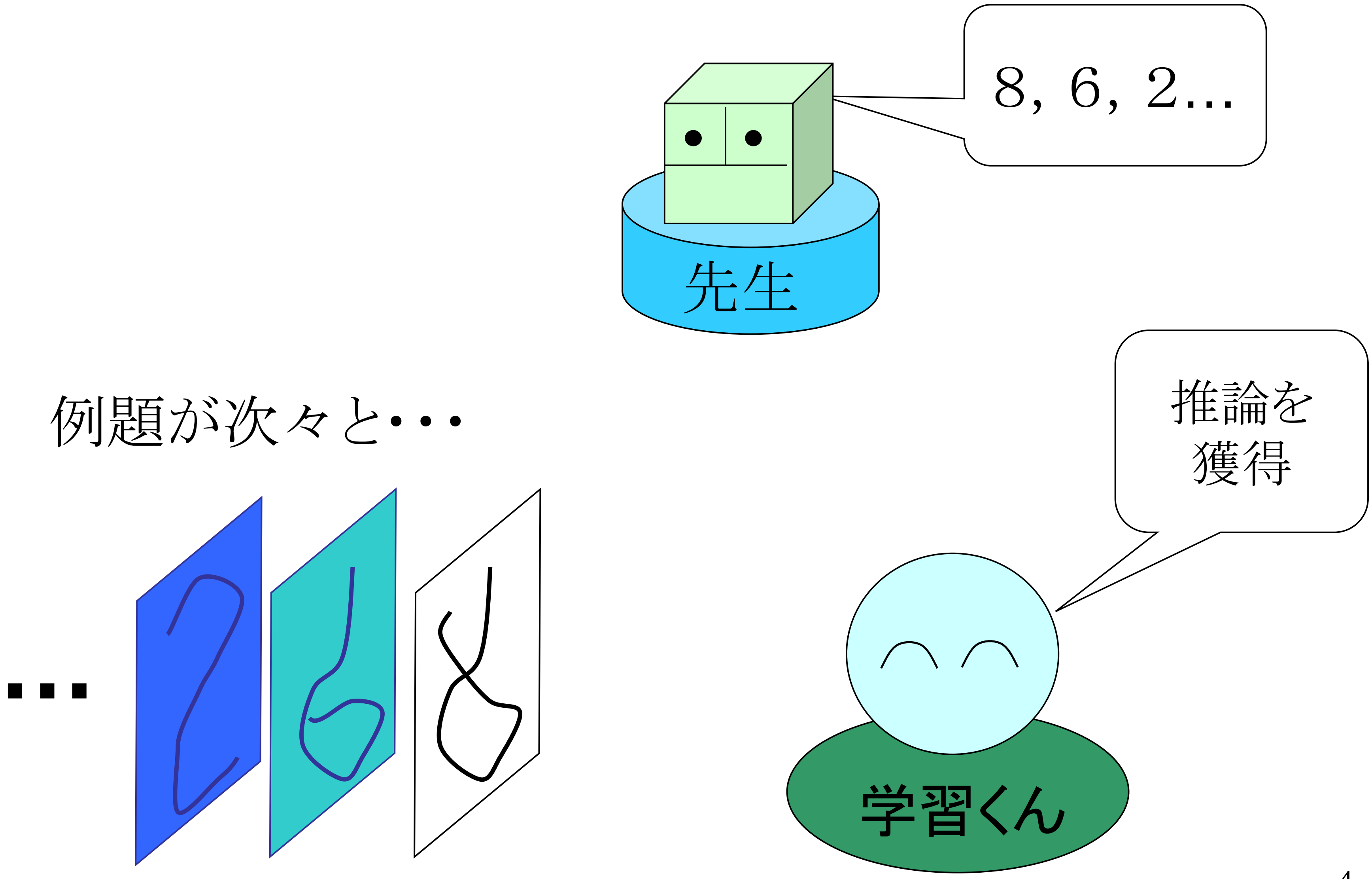


目次

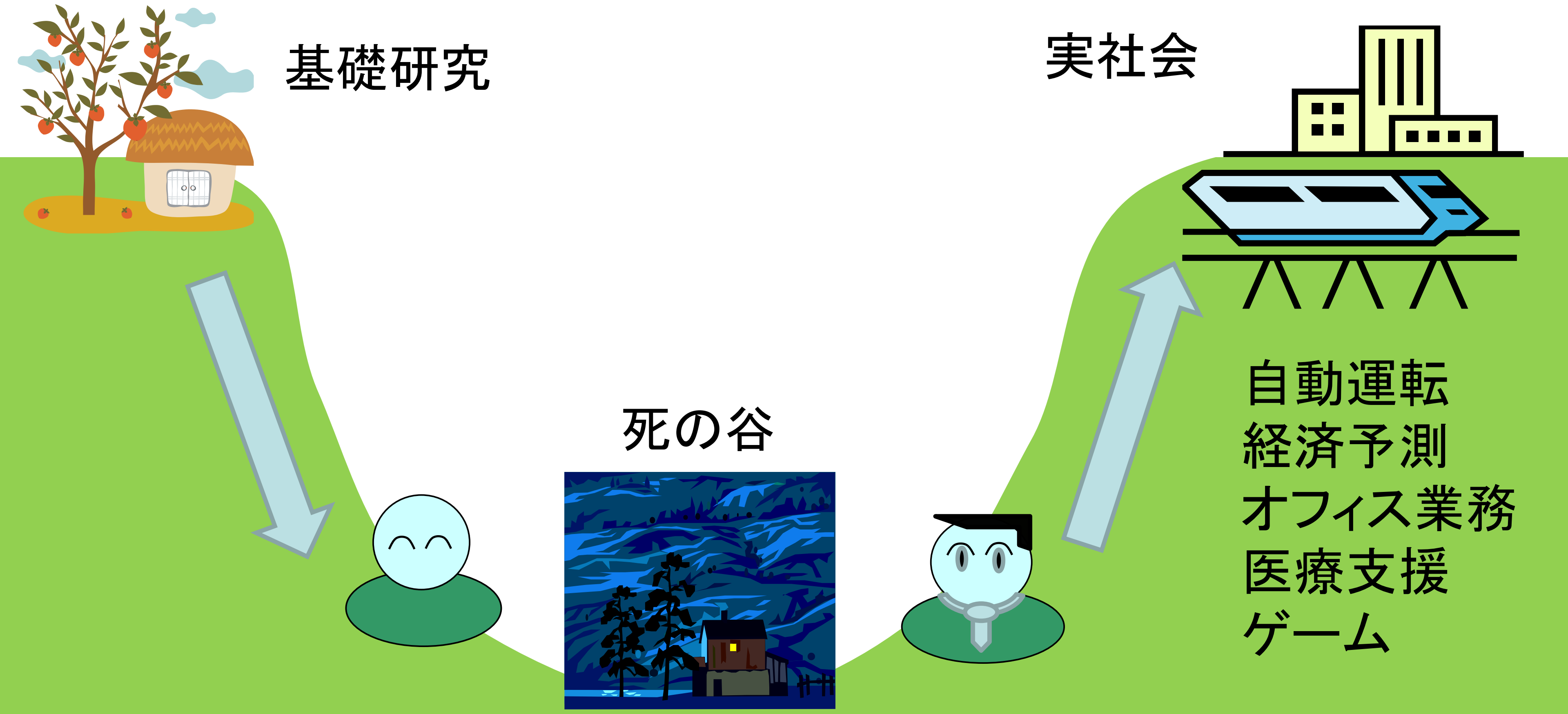
第1話　「人間力」の成功と挫折



基礎研究のころ



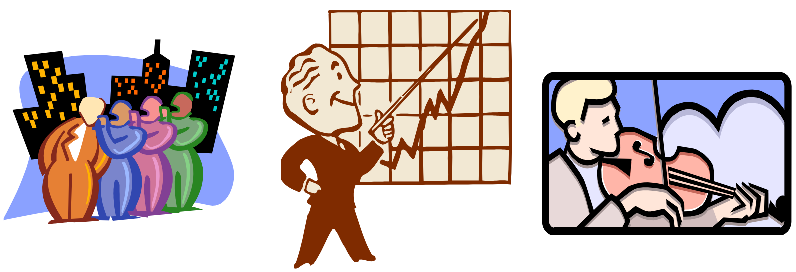
死の谷を越えて



基礎研究が社会で役立つには、死の谷を通らなくてはならないといわれている・・・が人間力で突破した・・・？

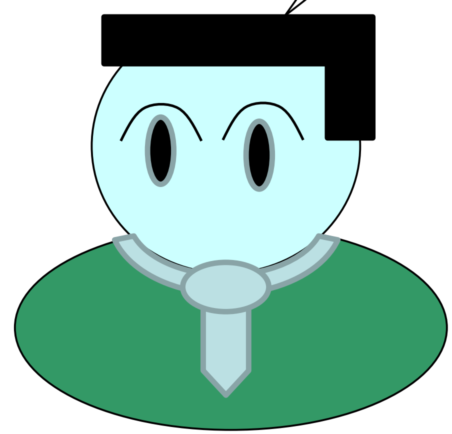
社会に出た

人間力・構想力・
コミュニケーション力
で死の谷を突破せよ。



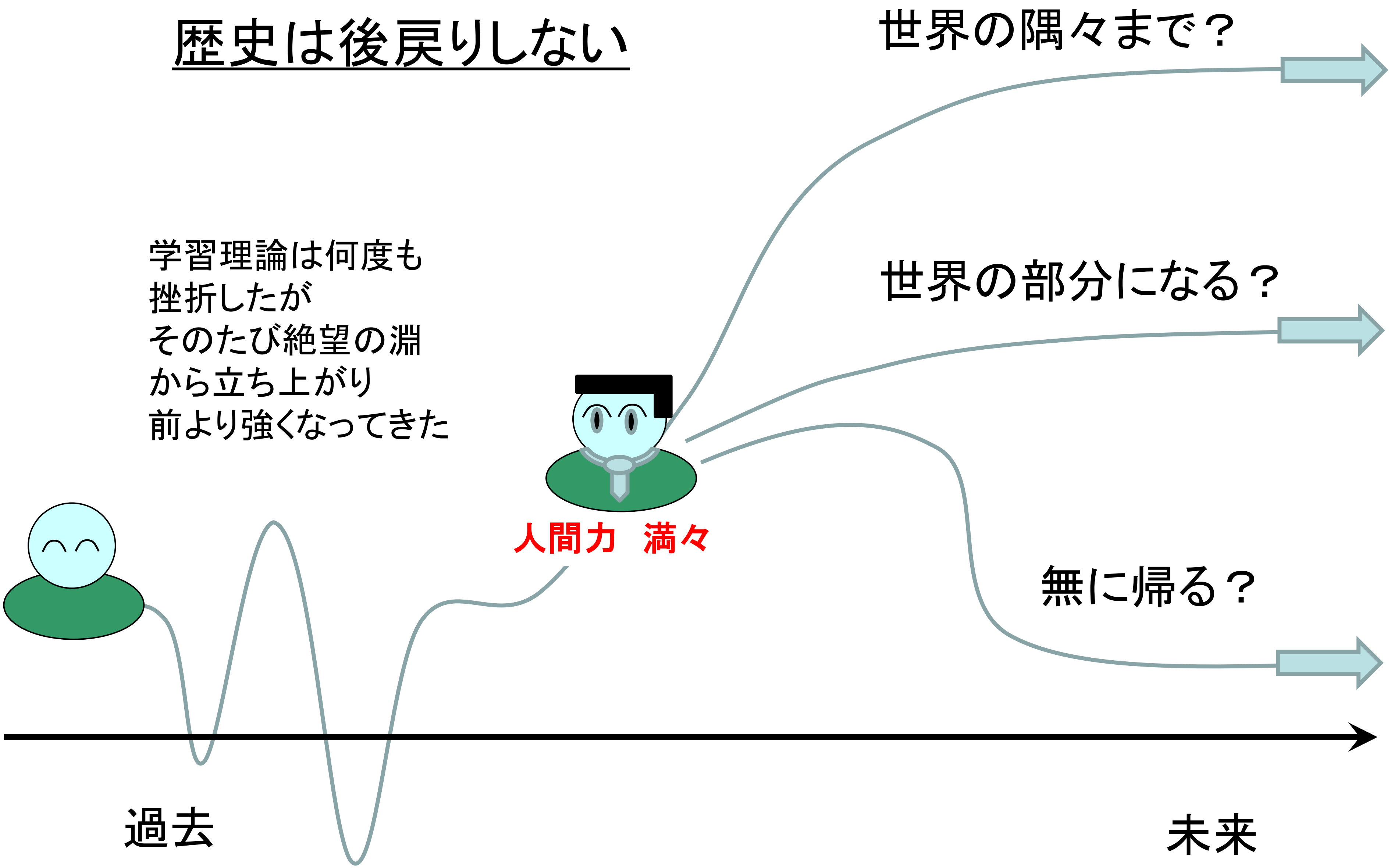
SNS解析、株価予測、芸術生成、
自動運転、購買解析・・・・・

アウトプットに
学問は不要

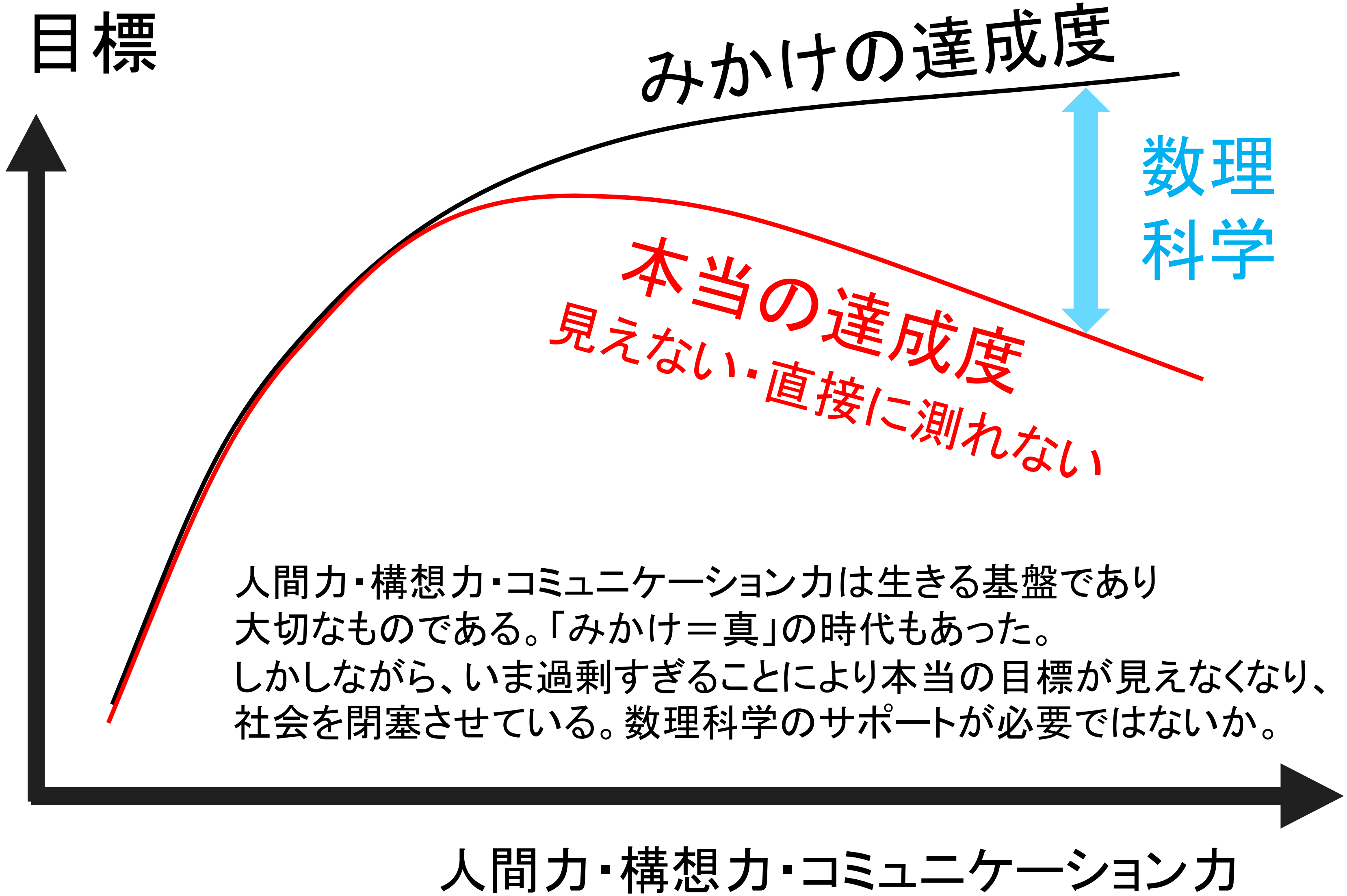


歴史は後戻りしない

学習理論は何度も
挫折したが
そのたび絶望の淵
から立ち上がり
前より強くなってきた

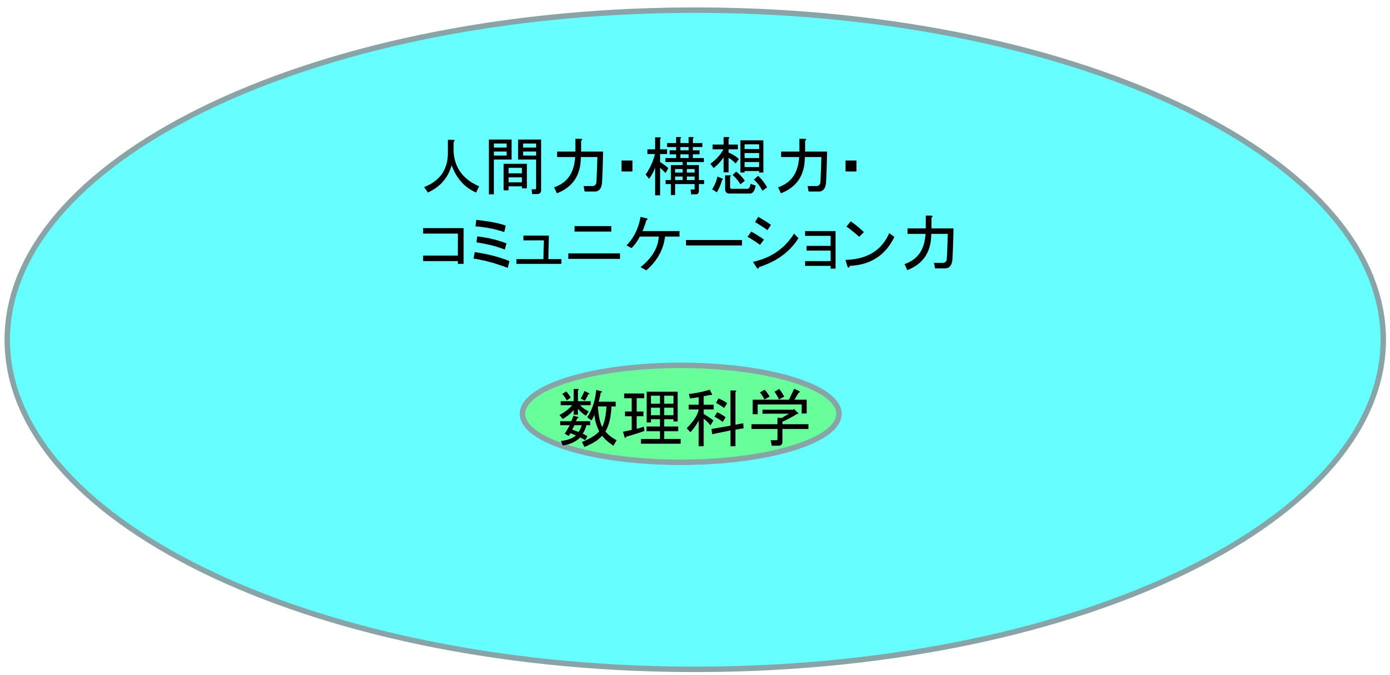


本当の達成度は見えない



過剰な人間力が未来を閉塞させている

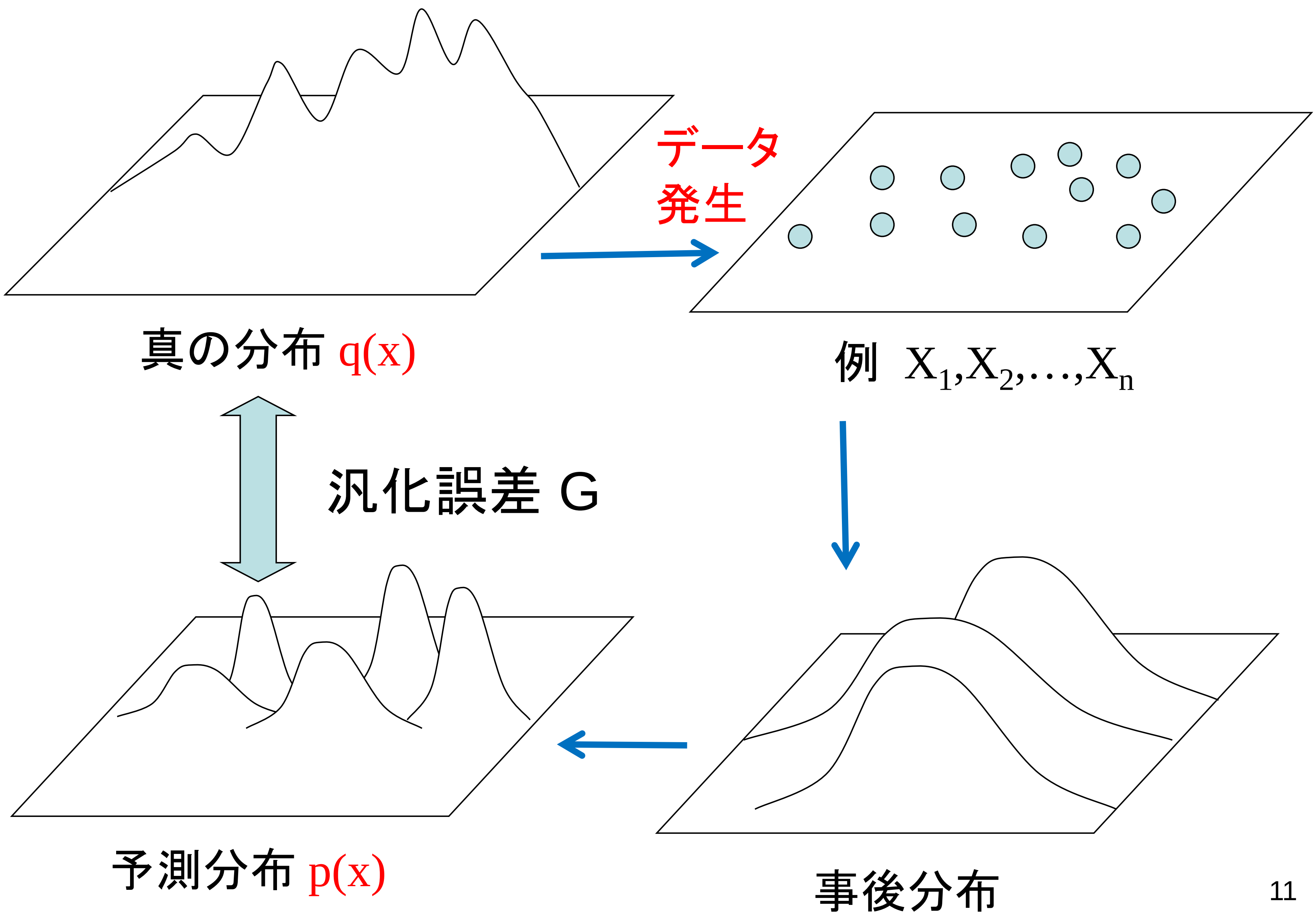
諸外国のAI研究は「早い遅い」の違いはあっても、必然的に社会や産業を変えていくだろうと思われているが、日本のAI研究は「無に帰すのでは」と危ぶまれている。なぜなのだろうか。



日本製「人工知能」システム

第2話　学習理論

学習の枠組み



汎化誤差の法則

定理　真の分布がモデルで実現可能であり、尤度関数
あるいは事後分布がガウス関数で近似できるならば

$$E[\,G\,] = \frac{d}{2n} + o\left(\frac{1}{n}\right)$$

d : パラメータの数, n: データの数

モデルと尤度関数

$N(x)$: 平均m, 分散1の正規分布

(1) モデル1

$$p(x|a,s) = (1/s)\, N(\, (x-a)/s\,)$$

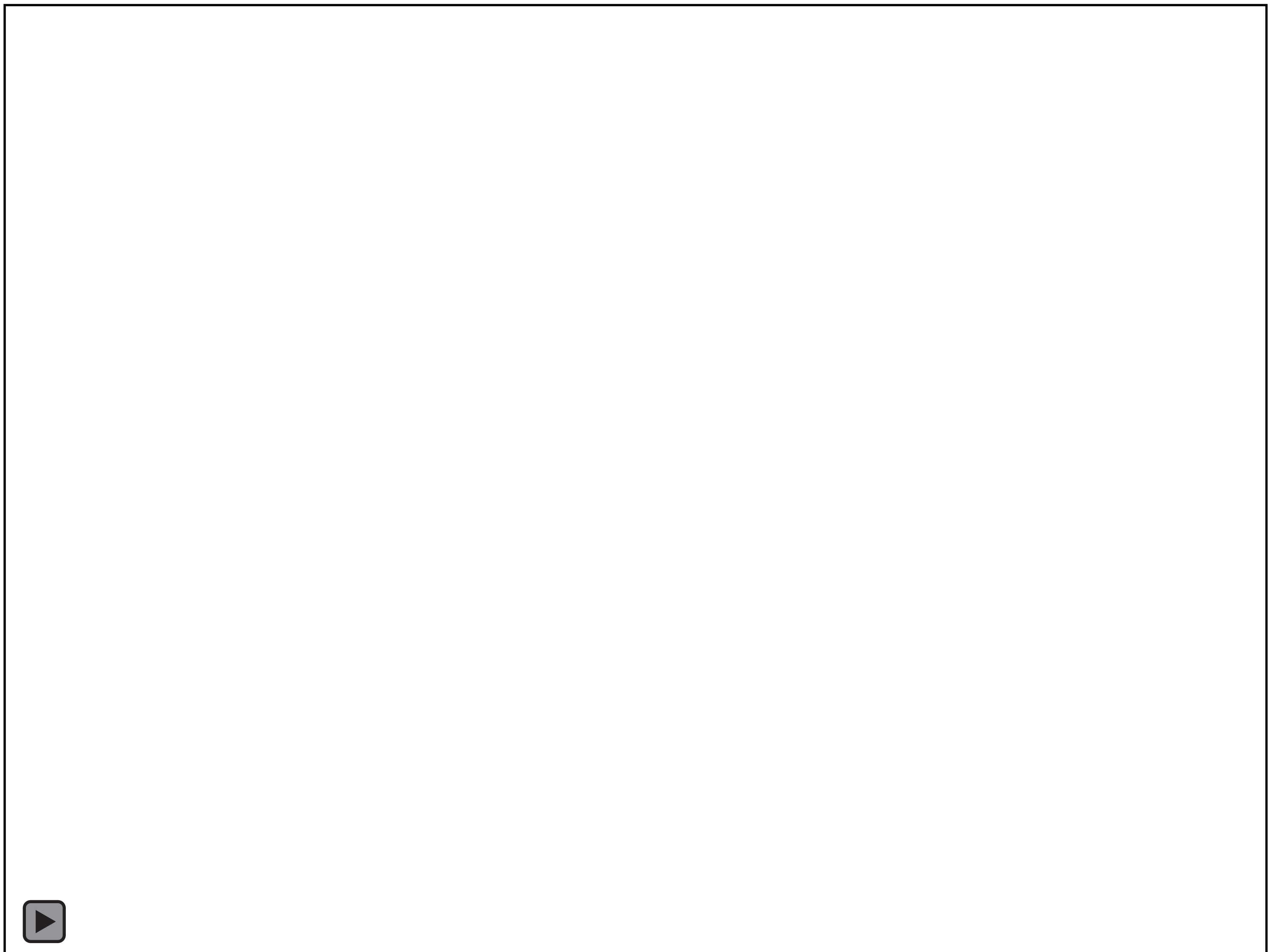
(2) モデル2

$$p(x|a,b) = (1-a)\, N(x) + a\, N(x-b)$$

尤度関数　$p(X_1|w)\, p(X_2|w) \cdots p(X_n|w)$

正則モデルの事後分布, n=30

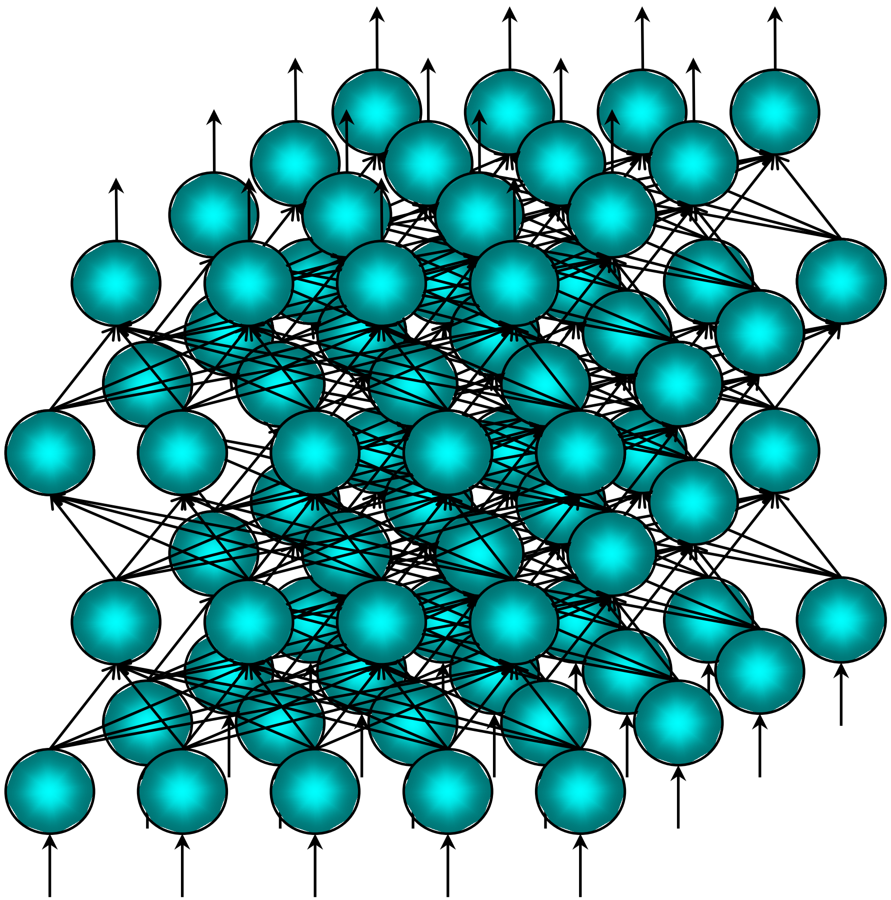
モデル $(1/s)\,N((x-a)/s)$,　　真 $a=0$, $s=1$.



非正則モデルの事後分布， n=3000
モデル (1−a)N(x) + aN(x−b)，真 a=0.5, b=0.3

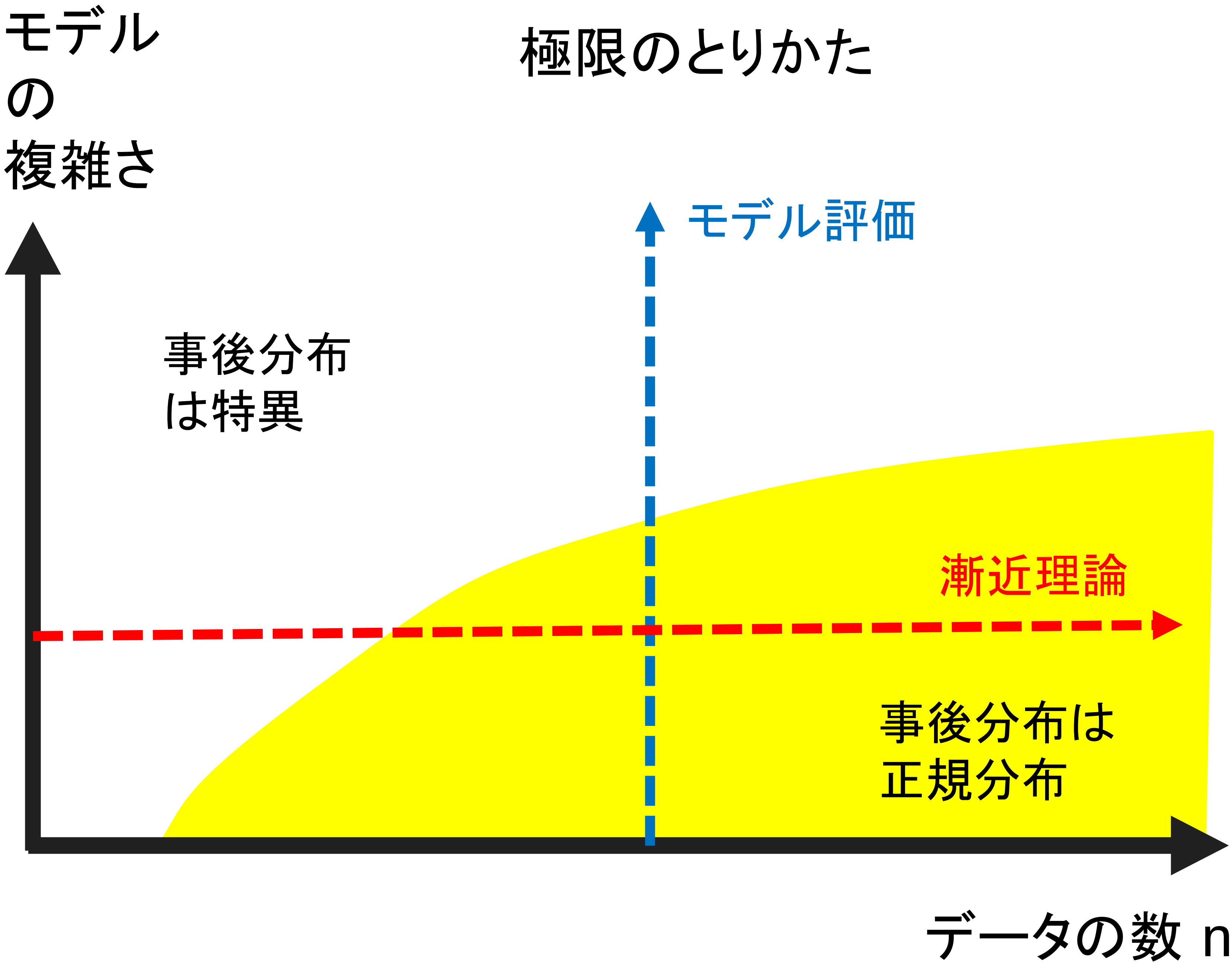
事実をありのままに見ることは難しい

人間力 「データが多ければ尤度関数はガウス関数で近似できる」



「機械学習のモデル
では成り立たない」という
事実は30年近く前から
知られていた

過剰すぎる人間力が
自然なままの物事を
みることを阻害する


極限のとりかた
モデルの複雑さ
モデル評価
事後分布は特異
漸近理論
事後分布は正規分布
データの数 n

汎化誤差の法則

定理　真の分布がモデルで実現可能であれば　いつでも

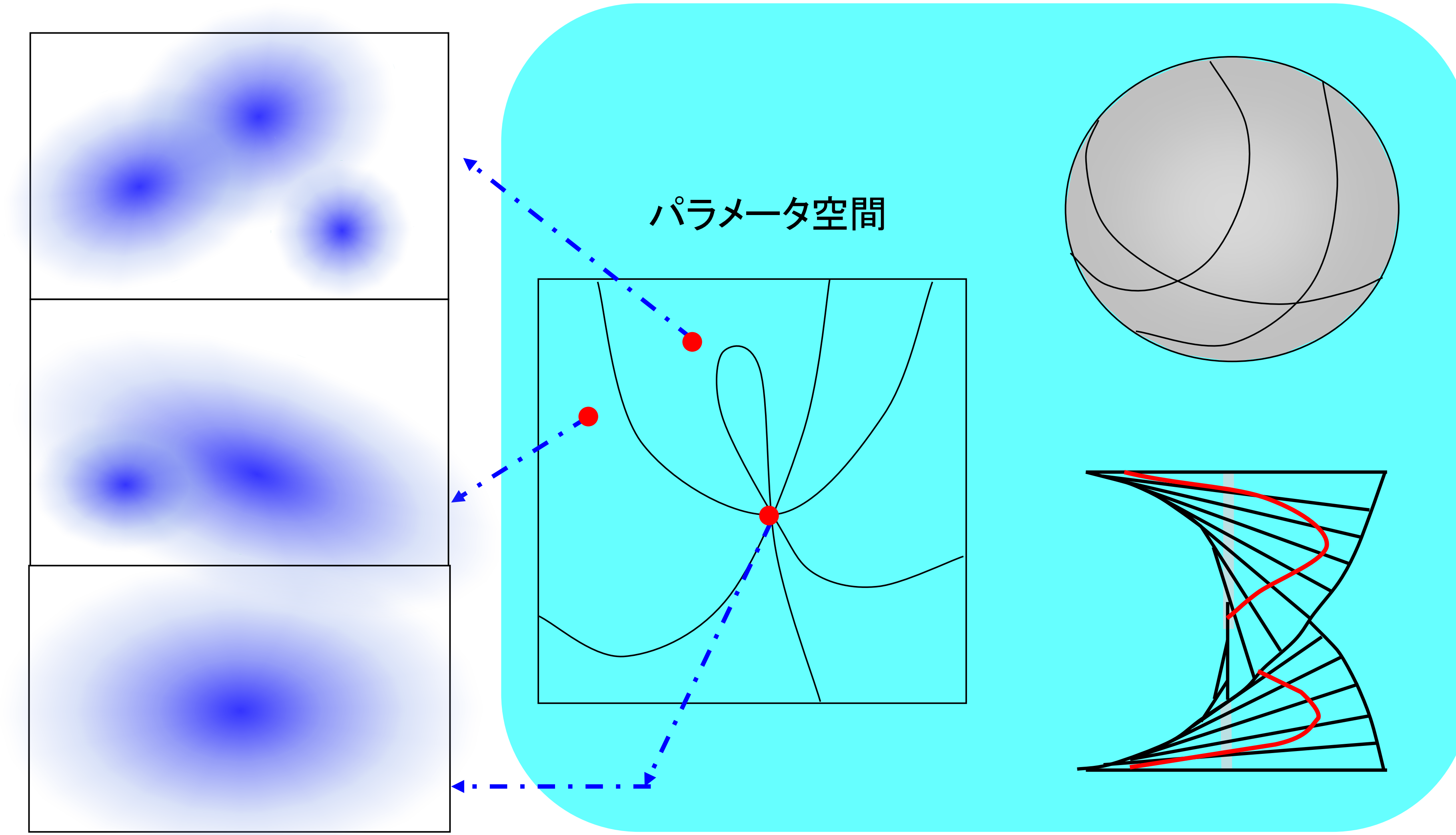
$$E[\,G\,] = \frac{\lambda}{n} + o\left(\frac{1}{n}\right)$$

λ：実対数閾値，n: データの数

一般にλ＜d/2

定数λは（真の分布、モデル、事前分布）の関数

双有理な世界



確率分布の世界

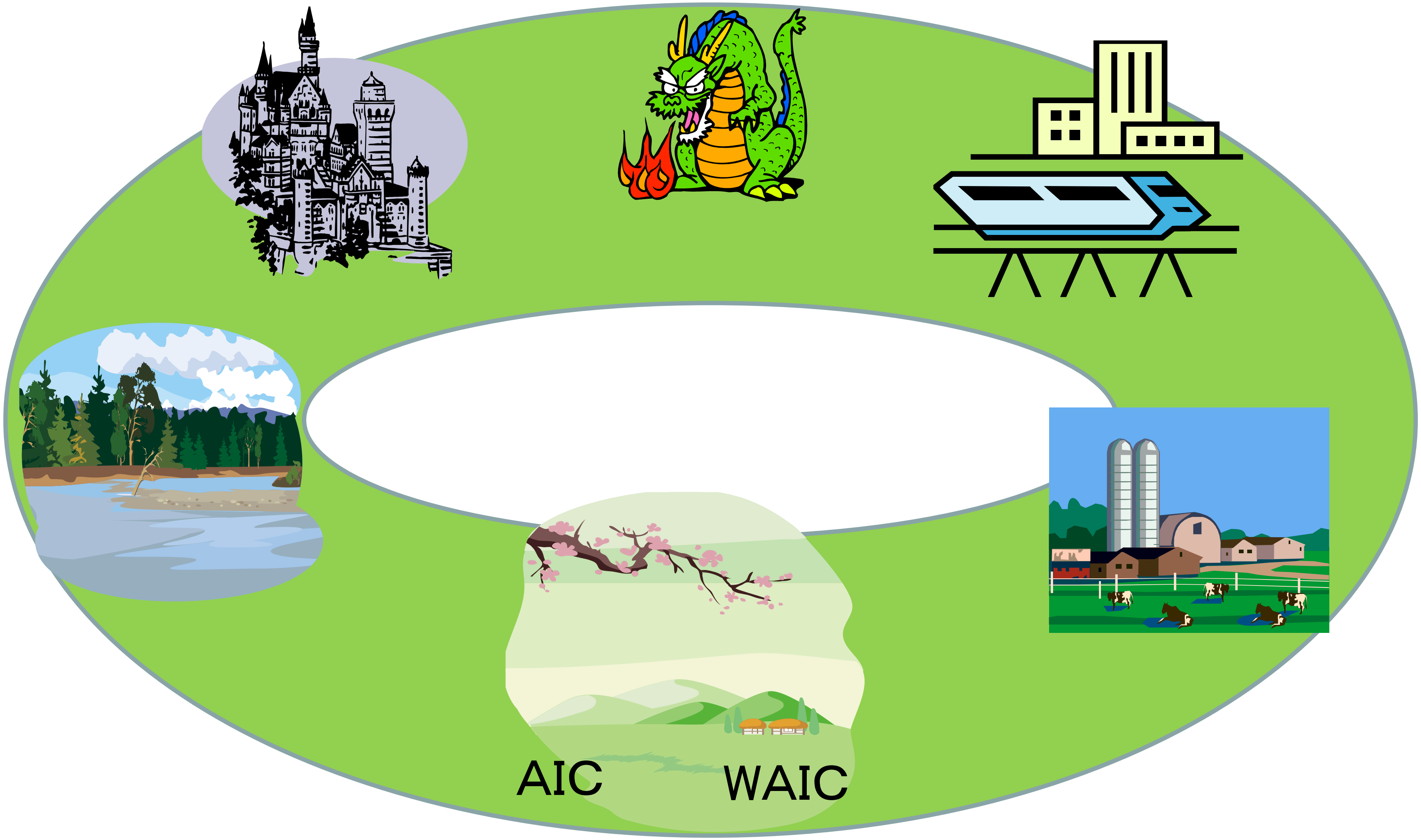
双有理な世界　実対数閾値は不変量

漸近正規性では　解明できなかった　新しい世界

混合正規分布（山崎さん）
ボルツマンマシン（山崎さん）
縮小ランク回帰（青柳さん）
ツリー構造モデル（ズワーニクさん）
確率行列分解（林さん）
変分ベイズ混合正規（渡辺一帆さん）
変分ベイズ汎化誤差（中島さん）
変分ベイズ文脈自由文法（星野さん）
変分ベイズ混合ベルヌーイ（梶さん）
変分ベイズ非負値行列分解（幸島さん）
マルコフ連鎖の交換率（永田さん）
sBIC（ドートンさん）

→ 代数統計学の流れのひとつ

第3話　学習をめぐる旅



統計的モデリングの歴史

伝説

統計モデルは　データの生成過程を深く理解し
人間力・企画力・コミュニケーション力で作るほかなかった。

真の分布は不明であるから

統計モデルの適切さを合理的に調べる方法はないと
考えられていた。

赤池情報量規準 AIC の登場

定理　真の分布が学習モデルで実現可能であり
尤度関数がガウス関数で近似できるならば

$$E[\text{汎化誤差}] = E[\text{学習誤差}] + \frac{d}{n}$$

d : パラメータの数, n: データの数

真の分布がわからなくても　汎化誤差を推測することができる。

AICは人間力だけに頼っていた統計モデリングに数理科学の光をもたらした。

次元の呪い と 機械学習

伝説

あらかじめ作られたモデルでは 音声・画像・言語などの
高次元空間にある大きな複雑さを持つ情報を学習すること
は難しいといわれていた。

深層学習の登場。

機械学習に現れる学習モデルは、階層的な構造を
持つことで「次元の呪い」を克服できる。（1993, バロン）

階層構造を持つモデルでは 事後分布は正規分布で
近似できない。AIC でモデルの評価を行なうことは
できない。

広く使える情報量規準 WAIC

定理　いつでも

E[汎化誤差] = E[学習誤差] + E[事後揺らぎ]

事後揺らぎ = V_w[log p(X|w)]

☆ 高次まで等しいので 事前分布の最適化にも使える。

☆ 真の分布が学習モデルで実現可能で
事後分布が正規分布に等しいときには
事後揺らぎは d/2n と等しくなる

混合正規分布の例

$\mathbf{x} \in R^2$ のモデル

$$p(\mathbf{x}|w) = \sum_{k=1}^{K} a_k \, G(\mathbf{x}-\mathbf{b}_k)$$

事前分布　指数 α のディィリクレ分布

実験では K=8, α =5 を使っている

真の分布は　有限個の混合では作れない

データの数 n=50.

[illegible]

WAIC は 何だったのか

WAICの導出には

特異点論→代数幾何→超関数→経験過程→学習理論

という旅が必要だった。しかし導かれた結果は

E[散逸] = E[揺らぎ]。 （揺動散逸定理）。

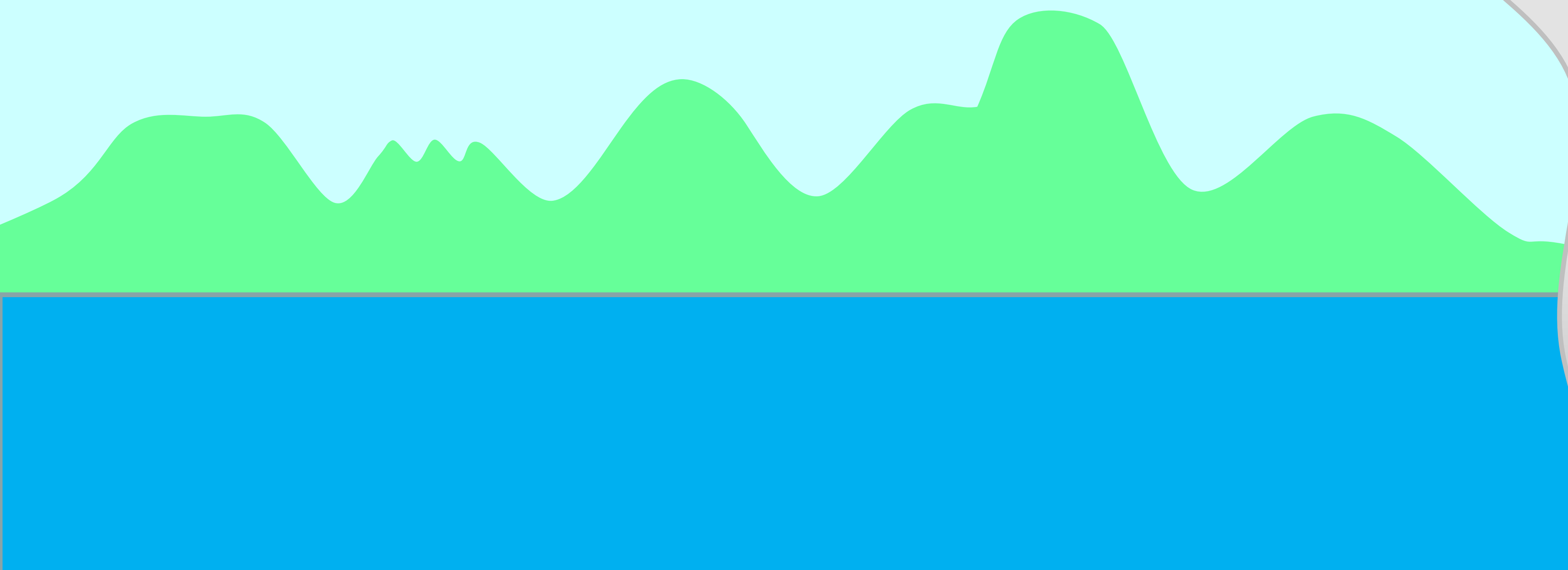
平衡状態（確率分布）では 揺らぎ と 散逸 がつりあっている

第4話　学習理論よ　何処へ

(1) データ科学と機械学習

(2) 学習理論と人間

(3) 学習理論は　自然に帰る　？



(1) データ科学　と　機械学習

データ科学

データ → 個別モデル → 最適設計 → 予測

機械学習

普遍モデル → 学習理論 → データ → 予測

データ科学と機械学習は、最初はまったく違うものだった。

双対の関係

双対構造とは　$f(x) = \langle f, x \rangle = x(f)$

データ科学	データ x に関する研究 データについての深い知見を どのようにしてモデル化に活用するか

機械学習	関数 f に関する研究 変幻自在の普遍モデルと学習理論を いかに作り上げるか

どちらも大切であり State of Art では両者はほとんど同じになる

豊かな未来のために「異なり続ける努力」も大切

(2) 学習理論と人間

宇宙のどこかに地球以外に
知性を持つ生命がある星を
考えてみよう。

そこには数学も物理学も
学習理論もあるだろう。

神経回路網はあるだろうか。

まったく違う形式の学習
モデルもあるかもしれない。

機械学習　と　人間

機械学習は　しばしば　人間の考えとは異なる方向に到達することがある。機械学習のさらなる研究は「理解する」とは何であるのかについての新しい世界に到達するだろう。
「人間力・構想力・コミュニケーション力」 も変わらなくてはならない。学習理論と人間は、どちらも変わっていく。

機械学習さん　　　人間さん

(3) 自然と情報

◎ この伝説が歴史的事実であるかどうかは不明だそうです。

シャノンは、確率分布 $q(x)$ に従う独立な文字列は

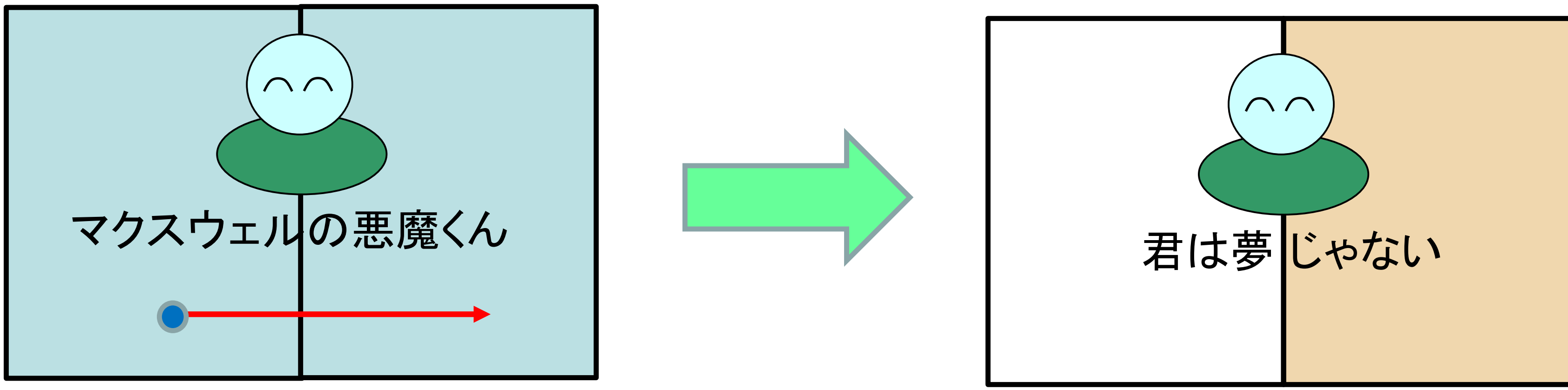
$$S = -\Sigma\ q(x) \log q(x)$$

の長さにまで圧縮できることを発見した。

S は何かと尋ねられたフォンノイマンは
「それはエントロピーです」と答えたという。

それ以来、自然と情報には「よく似ている概念と数理」が
あると思われて来た。

「似ている」ではなく同じだった



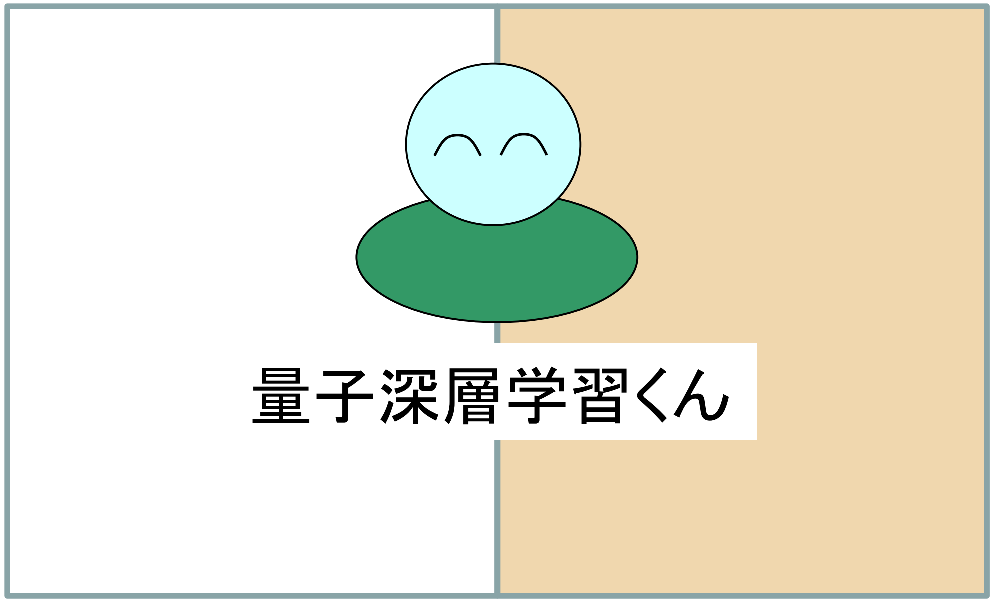
1ビットの情報　＝　物理現象

情報は架空の世界ではなく　実世界の中にあり
情報まで含めた形で熱力学第2法則が成り立つ。

自然のエントロピー　＝　情報エントロピー

☆　田崎晴明,「悪魔」との取りひき-エントロピーをめぐって,
日本物理学会誌, 66(3), 172-173, 2011.

学習理論よ　何処へ



「学習くん」が従う法則は自然科学と無矛盾であるはずである。

学習理論が数理科学として普遍性を追求していくならば
自然科学の法則と　いったん合流するのかもしれない。
もっと未来のための　新しい芽が必要かな。

結論