

ランダムな概念は どう発展してきたか

数学基礎論サマースクール 2012

2012年9月7日 東京工業大学

宮部賢志

京都大学 数理解析研究所

はじめに

ランダムな正当な使い方？

- ❖ ランダムな列とは
「確率的プロセスの実現系列」のことである
(いわゆるサンプルパス)
- ❖ 確率とは何？
- ❖ 循環的な使い方

ランダムのおわけ方

- ❖ 確率 – von Mises, (マルチンゲール,) algorithmic probability, ゲーム論的確率論
- ❖ 予測 – Solomonoff, Hutter, MDL
- ❖ 分類 – Vitányi
- ❖ 乱数

話の流れ

- ❖ ランダム・確率の歴史
(紀元前から1900年頃まで)
- ❖ von Mises と Kolmogorov による確率論
- ❖ ランダムネスの応用
(algorithmic probability, 分類)

ランダム・確率の歴史

テーゼ

- ❖ 予測可能性 = 圧縮可能性
- ❖ 予測限界 = 圧縮限界
- ❖ 理論の限界を語る = ランダムを語る？

古代の偶然論

- ❖ Aristotle B.C. 384-322
- ❖ 自己偶発的な原因を”Tyche”または”automaton”などと呼んだ『自然学』
- ❖ 「偶然的なものは科学的に取り扱えないとみなさなければならぬ」『形而上学』
- ❖ 科学の中での偶然性の研究の否定 『確率論の黎明』
(安藤洋美)より

偶然の否定

- ❖ Augustine (354-430)
- ❖ 「その原因は至るところにある神の手により操作されるもので、その意味で無作為な(random)ものは何もないし、偶運(chance)なるものは存在しない」
『八十三の問題について』
- ❖ 「すべては神の摂理に従う」 『神の国』
- ❖ 17世紀まで確率計算の発生を抑止した原因を作った
『確率論の黎明』(安藤洋美)より

確信の度合い

- ❖ Thomas Aquinas 1225-1274
- ❖ 「聖なる教え(神学)は、哲学者たちが自然理性により真理を知り得た場合には、彼らの権威をも用いる・・・しかし、聖なる教えはかかる権威をいわば教えの外の蓋然的論拠(Argumenta probabilia)として用いるにすぎない。」 『神学大全』
- ❖ 有用性を認めながらも、正当化する必要がない

数学としての確率論の始まり

- ❖ PascalとFermatの往復書簡(1654)
- ❖ 掛け金の配分問題
- ❖ ランダムな概念をどう捉えているか？

Pascalの方法

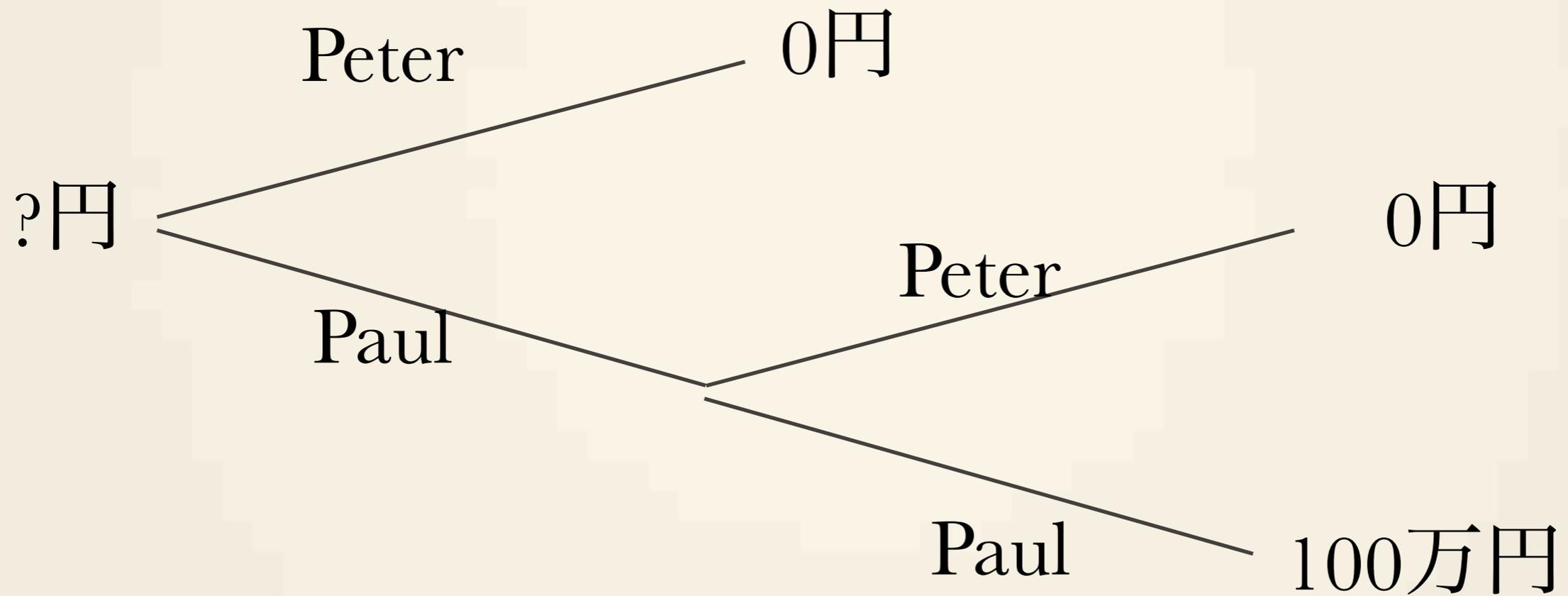
- ❖ Pascal (1623-1662)
- ❖ 「公平な賭け + 帰納法」
により解いた



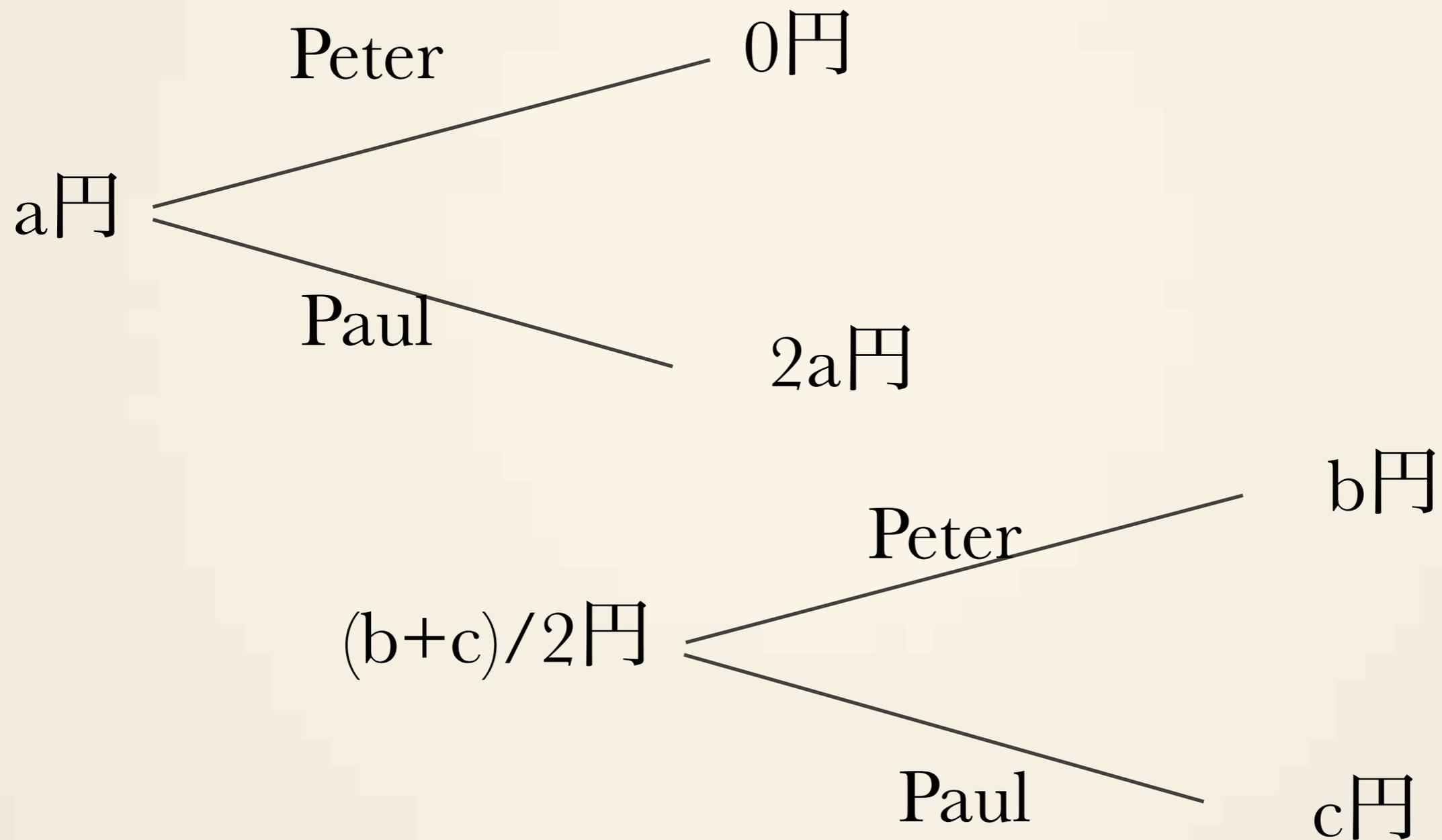
Pascalの方法(1/4)

- ❖ 掛け金100万円, プレイヤーPeterとPaulは勝利までそれぞれあと1点と2点足りない. どう配分すべきか?

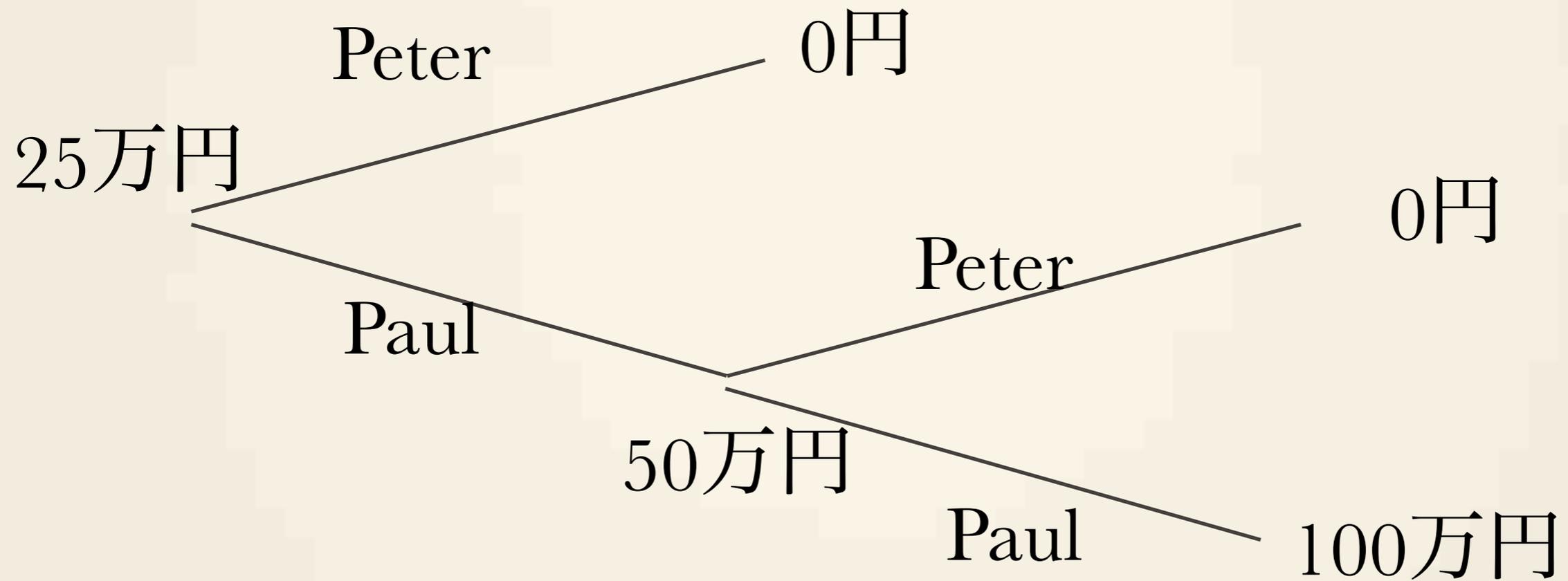
Pascalの方法(2/4)



Pascalの方法(3/4)



Pascalの方法(4/4)



基盤にしているものは？

- ❖ 「同じ条件で争うことが公平である」
という原理
- ❖ von Mises, マルチンゲール,
ゲーム論的確率論, などの流れの大元

Fermatの方法

- ❖ Fermat (1607or08-1665)
- ❖ 「組み合わせ」を基盤に
- ❖ 多くの場合が平等に表れる



Fermatの方法

- ❖ 決まった回数だけプレイを続けるもの
と想像しよう
- ❖ 組み合わせの数によって配分
- ❖ 非常に楽！
- ❖ 公理主義的確率論に似ている？

確率概念の始まり

- ❖ Arnauld (1612-1694)
- ❖ “The Port-Royal Logic”
- ❖ Pascal と交流があった



確率概念の始まり

- ❖ 「10人の人が各人1クラウン賭けるゲームがあり、彼らのうちの一人だけが買って全部をせしめ、他の人々が皆失うとしよう。（中略）彼が1クラウン失い、9クラウンを得ないのは、各々の場合9倍も確からしい。」 “The Port-Royal Logic”
- ❖ G. Shafer (1978) “Non-additive Probabilities in the Work of Bernoulli and Lambert”
- ❖ I. Hacking (1975) “The Emergence of Probability”

Probability と Chance

- ❖ 「公平な賭け」と「組み合わせの数」
- ❖ external evidence and internal evidence (Arnauld)
- ❖ epistemic probability and aleatory probability
- ❖ subjective probability and objective probability ?
- ❖ Bernoulliの弱大数の法則が同一視を後押し

Laplace

- ❖ Laplace (1749-1827)
- ❖ 古典確率の集大成
- ❖ 決定論者 (cf. ラプラスの悪魔) で、確率は我々の無知に起因すると考えた



古典確率

- ❖ 古典確率の集大成
- ❖ 「一定数の同等に可能な場合，すなわちそれらが存在するかどうかについてわれわれが決めかねる程度が同じである場合に帰着させ・・・」 『確率の哲学的試論』

ここまでのまとめ

❖ 古代

ー ランダムは取り扱わない

❖ 1900年まで

ー 「同程度に確からしい」が基盤

von Mises と Kolmogorov による
確率論

Hilbert

- ❖ Hilbert (1862-1943)
- ❖ 現代数学の父
- ❖ 1900年の国際数学者会議での「23の問題」



確率の公理化

- ❖ 1900年の国際数学者会議での「23の問題」
- ❖ 第6問題「物理学は公理化できるか」
- ❖ 確率の厳密な取り扱いとして、幾何学のような公理化を求めた

Kolmogorov

- ❖ Kolmogorov (1903-1987)
- ❖ ロシアの数学者
- ❖ 確率論,
位相幾何学,
アルゴリズム情報理論



公理主義的確率論

- ❖ 『確率論の基礎概念』 (1933) で「公理主義的確率論」を提唱
- ❖ 自身は頻度主義者
- ❖ その使いやすさから確率に対する立場を問わず使われている

確率とは何か

- ❖ 原理A—試行が非常に多くの回数(n 回)繰り返されたとして、その結果事象Aの起こった回数が m 回であるとき、 m と n の比 m/n がほぼ $P(A)$ に等しいと事実上確信できる。
- ❖ 原理B— $P(A)$ が非常に小さい場合には、試行が1回だけ実現したときには事象Aは起こらないと事実上確信できる。

傾向説

- ❖ しかしこれは、わたしたちがその条件について、確率に等しい頻度を持つ連続性を生み出す傾向説、性質、傾向を備えたものとして、具体化しなければならぬことを意味する。そしてそれはまさに、傾向説の主張である。（Popper 1959）
- ❖ 「ランダムである」とは「確率1で起こる事象」

Cournotの架け橋

❖ Antoine Cournot
1801-1877

❖ 物理的に不可能な現象とは、その確率が無限に小さいもののことである。



確率の公理は妥当か？

- ❖ de Finettiらは主観主義の立場から無限加法性に対して批判
- ❖ Keynesの確率論では確率は数値では表現できないとした
- ❖ Dempster-Shafer理論では非加法的な確率を許す
- ❖ imprecise probabilityでは確率を幅で表す

現実への応用のために

- ❖ 既に述べてきたように、数学的確率論の結果を現実の「ランダムな現象」に応用するには、何らかの形で確率の頻度論が必要であろう。そしてその避けられない部分はすでに von Mises の精力的な努力により確立されている。（Kolmogorov 1963）

von Mises (1883-1953)

- ❖ von Mises (1883-1953)
- ❖ 1919年の論文に始まり、
確率の頻度説の数学的定
式化を試みた



確率論の範囲

- ❖ 我々の確率論は「ドイツが未来にリベリアと戦争する確率はあるか？」というような問題とは何の関係もない。
- ❖ 物理の言葉を使うなら、確率論を適用するためには、事実上無限に長い一様な観察がなければならぬと言えるだろう。

確率論の範囲

- ❖ 他のすべての自然科学のように，確率論は観察から始まり，それらを並べ，分類し，そこからある基本的な概念や法則を導き，最後に，一般的な普遍的に適用可能な論理を使って，実験結果と比較することで検定可能な結論を導くのである．すなわち，我々の見方では，確率論は普通の科学であり，テーマとして際立ってはいるが，推論の方法として特別なわけではない．

確率論の範囲

- ❖ ある数学者が「君は私が次の列車に乗り遅れる確率を計算できるか？」と言ってからかうならば、解答を丁重に断らなければならない。それはちょうど「君はあの2つの山の山頂の距離を計算できるか？」と彼が聞かれたら、彼は断るように。つまり、もし適当な距離と角度が分かれば、距離は計算できる。同じように確率が分かるのは、依存する確率が分かっている時だけである。

確率論の範囲

- ❖ 「確率」という言葉は，ランダムな条件を満たす真のcollectiveの極限頻度という意味だけで使う。

von Misesの理論要約

- ❖ 「まずCollectiveありき，そして確率がある。」
- ❖ ランダム概念は，確率が定義できるための条件として使われている。

ここまでのまとめ

❖ 公理主義的確率論

- 確率（モデル）を最初に仮定する
- ランダムが数学と物理の架け橋として使われる

❖ von Misesの理論

- 独立同分布に相当する場合にのみ考える
- ランダムはその条件として使われる

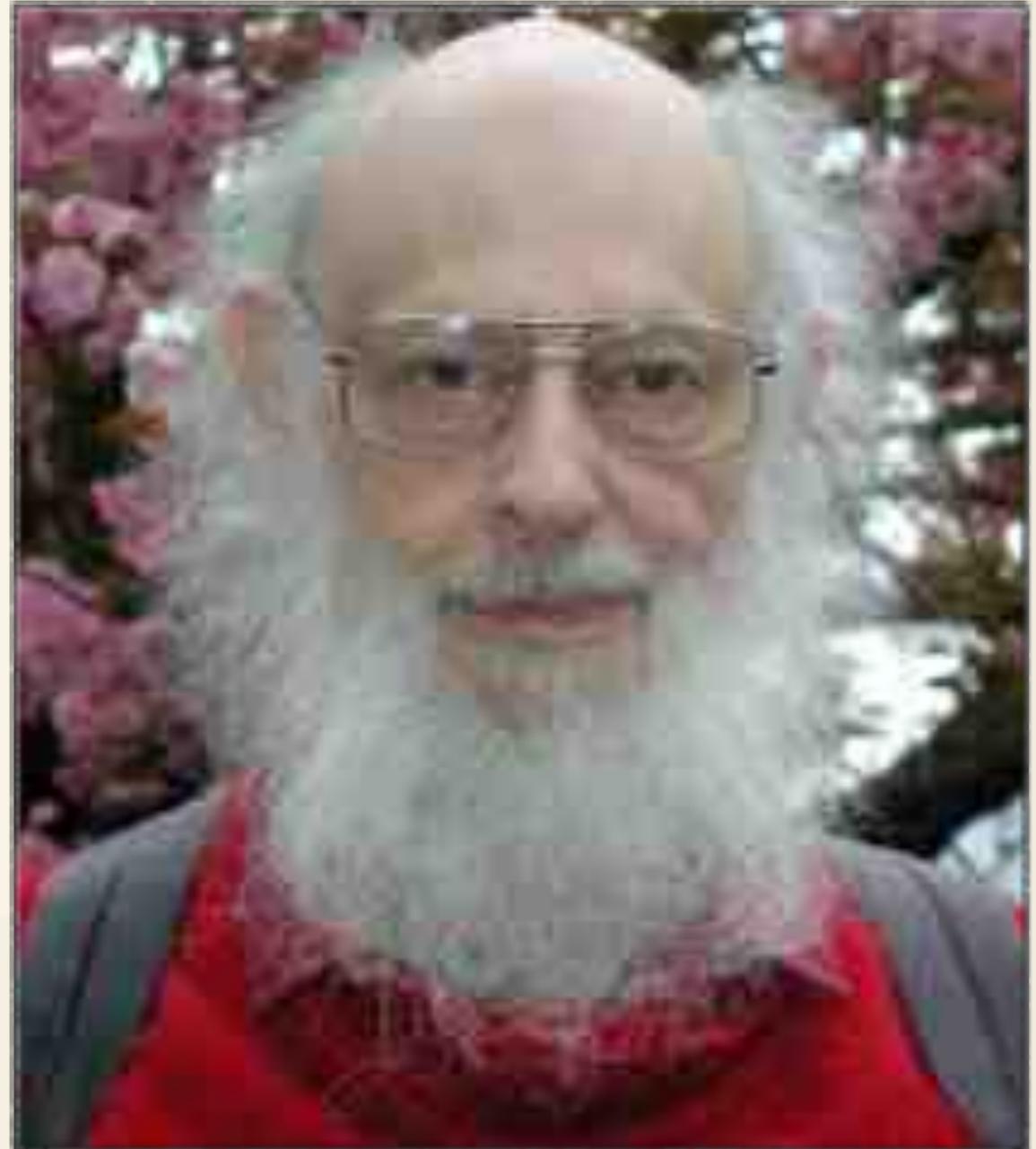
統計の立場から

一般的に考えられている立場	中心人物	ランダムネスから見ると？
頻度説	Neyman, Pearson	計算可能性として頻度のみを考慮している
相対頻度に基礎を置く信頼性の尺度	Fisher	
主観説	Bayesian	計算可能性が考慮されていない

ランダムネスの応用

Solomonoff

- ❖ Solomonoff (1926-2009)
- ❖ algorithmic probability, algorithmic information theory
- ❖ artificial intelligence based on machine learning



解釈

- ❖ Epicurus (B.C.341?-)
“If more than one theory is consistent with the data, keep them all.”
- ❖ William of Ockham (1288-1348)
“Keep the simplest theory consistent with the observations.”
- ❖ 2つの方法の組み合わせ

最善の予測を求めて

定義 (Solomonoff 1964)

関数 M を以下で定義する。

$$M(\sigma) = \sum_{p: \sigma \preceq U(p)} 2^{-|p|}.$$

ここで U は万能マシンである。

半測度

定義

関数 $M : 2^{\omega} \rightarrow \mathbb{R}^+$ が, $M(\epsilon) \leq 1$,

$$M(\sigma) \geq M(\sigma 0) + M(\sigma 1)$$

を満たす時, M を**半測度**と呼ぶ.

半測度 M が以下を満たす時, **最善 (optimal)** であると言う:
任意の半測度 N に対して, ある定数 C が存在し, すべての σ
で, $C \cdot M(\sigma) \geq N(\sigma)$.

最善のc.e.半測度

命題

上記の関数 M は、最善の c.e. 半測度である。

命題

$\{\mu_n\}$ をすべての c.e. 半測度の計算可能な数え上げとする。関数 M を

$$M = \sum_n 2^{-n} \mu_n$$

で定義すると、 M は最善の c.e. 半測度となる。

半測度とマルチンゲール

M を半測度とすると,

$$d(\sigma) = 2^{|\sigma|} M(\sigma)$$

で定義される d はマルチンゲール. d をマルチンゲールとすると,

$$M(\sigma) = 2^{-|\sigma|} d(\sigma)$$

で定義される M は半測度.

半測度による特徴付け

定理 (Levy 1973)

μ を 2^ω 上の計算可能測度. M を任意の最善の c.e. 半測度とする. ある列 $x_{1:\infty}$ が μ -Martin-Löf ランダムであることと, ある定数 c が存在してすべての n で $M(x_{1:n}) \leq c \cdot \mu(x_{1:n})$ となることは同値.

次の値の予測

ある $x \in 2^\omega$ に対して以下のようにおく：

$$M(x_n | x_{<n}) = \frac{M(x_{1:n})}{M(x_{<n})}.$$

この値は,

$$\mu(x_n | x_{<n}) = \frac{\mu(x_{1:n})}{\mu(x_{<n})}$$

と近いだろうか？

Solomonoffの結果

定理 (Solomonoff 1978; Hutter)

計算可能な測度 μ に対して,

$$\frac{M(x_n | x_{<n})}{\mu(x_n | x_{<n})} \rightarrow 1 \quad (n \rightarrow \infty)$$

が μ 確率 1 で成り立つ.

Hellinger 距離を以下で定義する。

$$h_\sigma := \sum_{i \in \{0,1\}} \left(\sqrt{M(\sigma i | \sigma)} - \sqrt{\mu(\sigma i | \sigma)} \right)^2.$$

$$h_\sigma \leq 2 - 2 \sum_{i \in \{0,1\}} \sqrt{M(\sigma i | \sigma) \mu(\sigma i | \sigma)}$$

$$N_\sigma := \sum_{i \in \{0,1\}} \sqrt{M(\sigma i | \sigma) \mu(\sigma i | \sigma)}$$

$$N_\sigma \leq 1 - \frac{h_\sigma}{2} \leq \exp \left(-\frac{h_\sigma}{2} \right).$$

ここで関数 $d : 2^* \rightarrow \mathbb{R}^+$ を以下で定義する.

$$d(\sigma) = \sqrt{\frac{M(\sigma)}{\mu(\sigma)}} \cdot \exp \left(\sum_{i=1}^{|\sigma|-1} h_{\sigma_{1:i}} \right).$$

ここで, M は最善であるから, $M(\sigma)/\mu(\sigma) > c > 0$ と仮定して良い. すると,

$$\begin{aligned}
\sum_{i \in \{0,1\}} \mu(\sigma i) d(\sigma i) &= \sum_{i \in \{0,1\}} \sqrt{M(\sigma i) \mu(\sigma i)} \cdot \exp \left(\sum_{i=1}^{|\sigma|} h_i \right) \\
&= \sum_{i \in \{0,1\}} \sqrt{M(\sigma i | \sigma) \mu(\sigma i | \sigma)} \cdot \exp(h_\sigma) \\
&\quad \times \sqrt{M(\sigma) \mu(\sigma)} \cdot \exp \left(\sum_{i=1}^{|\sigma|-1} h_{\sigma_{1:i}} \right) \\
&\leq \sqrt{M(\sigma) \mu(\sigma)} \cdot \exp \left(\sum_{i=1}^{|\sigma|} h_{\sigma_{1:i}} \right) \\
&= \mu(\sigma) d(\sigma).
\end{aligned}$$

よって, d は μ -優マルチンゲール.

つまり, μ に関してほとんど確実に,

$$\limsup_n d(x_{1:n}) < \infty.$$

これより,

$$\sum_{i=1}^{\infty} h_{x_{1:n}} < \infty.$$

よって,

$$M(x_{1:n} | x_{<n}) - \mu(x_{1:n} | x_{<n}) \rightarrow 0.$$

MLランダムでの非収束

定理 (Hutter and Muchnik 2007)

ある計算可能な測度 μ とある最善の予測 M およびある μ -Martin-Löf ランダムな列に対して,

$$M(x_n | x_{<n}) - \mu(x_n | x_{<n}) \not\rightarrow 1.$$

μ として一様測度を取ることができる。

参考に

- ❖ Vovk (1987) が弱い結果を示す.
- ❖ Li and Vitányi (2nd ed. 1997), Vitányi and Li (2000) に誤りのある証明が載る.
- ❖ Hutter (2003) がその誤りを指摘する.
- ❖ Hutter and Muchnik (2004) が非収束を示す.

主觀確率

For quite some time I felt that the dependence of ALP (Algorithmic Probability) on the reference machine was a serious flaw in the concept, and I tried to find some “objective” universal device, free from the arbitrariness of choosing a particular universal machine. When I thought I finally found a device of this sort, I realized that I really didn't want it - that I had no use for it at all!

Solomonoff (2009)

分類

- ❖ 似ている \Leftrightarrow 似ていない
- ❖ ランダムでない \Leftrightarrow ランダム
- ❖ どれくらいランダムか？
- ❖ どれくらい似ているか？

情報量

文字列 x に対し, x^* で x を出力するプログラムの中で最小のもの (の 1 つ) を表す.

定義

y の x に含まれる **情報量** を

$$I(x; y) = K(x) - K(x|y^*)$$

で定義する.

情報量の基本的性質

命題

定数を除いて、以下が成り立つ。

1. $I(x : y) \geq 0$.
2. $I(x : x) = K(x)$.

情報量の対称性

定理

定数を除いて，以下が成り立つ．

$$K(x, y) = K(x) + K(y|x^*) = K(x) + K(y|\langle x, K(x) \rangle).$$

系

定数を除いて，以下が成り立つ．

$$I(x : y) = K(y) - K(y|x^*) = K(x) + K(y) - K(x, y) = I(y : x).$$

\leq は自明なので逆を示す.

上記の定理を示すのに, 以下の概念を使う. prefix-free マシン M に対し,

$$Q_M(\sigma) = \mu(\llbracket \{\tau : M(\tau) = \sigma\} \rrbracket)$$

とすると,

$$K(\sigma) = -\log Q(\sigma).$$

$\{\nu_s\}$ を U の定義域の計算可能な数え上げとし, σ_s, τ_s を $U(\nu_s) = \langle \sigma_s, \tau_s \rangle$ とする. また

$$W_\sigma = \{s : \sigma_s = \sigma\}$$

とおく. 任意の n と σ に対し,

$$s \in W_\sigma \text{ に対して } \langle |\nu_s| - n, \tau_s \rangle$$

を request として KC 集合を作る. ただしその request の重さは 1 を超えないようにする. その KC 集合から作られた prefix-free マシンを $M_{n,\sigma}$ と呼ぼう.

次のような prefix-free マシン V を考える. ある τ について $U(\nu) = \langle \sigma, \tau \rangle$ となるときに, $V(\nu) = \sigma$ とする. このとき,

$$Q_V(\sigma) = \sum_{\tau} Q(\langle \sigma, \tau \rangle).$$

よって,

$$\sum_{\tau} Q(\langle \sigma, \tau \rangle) \leq Q(\sigma) + O(1).$$

これからある c があって, すべての σ で,

$$2^{K(\sigma)-c} \sum_{\tau} Q(\langle \sigma, \tau \rangle) \leq 1.$$

よって

$$\sum_{s \in W_\sigma} 2^{-|\nu_s| - |\sigma^*| + c} = 2^{K(\sigma) - c} \sum_{\tau} Q(\langle \sigma, \tau \rangle) \leq 1.$$

これより, $M_{K(\sigma) - c, \sigma}$ は関係するすべての request が数え上げられている.

次のような prefix-free マシン M を定義しよう. 神託 ρ に対して, $U(\rho) = \sigma$ となる σ を探し, その後 $M_{|\rho| - c, \sigma}$ を模倣する.

神託が σ^* であれば, M は $M_{K(\sigma) - c, \sigma}$ を模倣する. 任意の τ に対し, $\langle \sigma, \tau \rangle^* = \nu_s$ となる $s \in W_\sigma$ が存在し, $\langle K(\sigma, \tau) - K(\sigma) + c, \tau \rangle$ が request されるから,

$$K(\tau | \sigma^*) \leq K_M(\tau | \sigma^*) \leq K(\sigma, \tau) - K(\sigma) + O(1).$$

相互情報量

定義

$I(x : y)$ を x と y の相互情報量と呼ぶ.

正規圧縮距離

定義 (Vitányi 2006)

x, y の正規圧縮距離を

$$\mathbf{NCD}(x, y) = \frac{C(yx) - \min\{C(x), C(y)\}}{\max\{C(x), C(y)\}}$$

で定義する.

その他

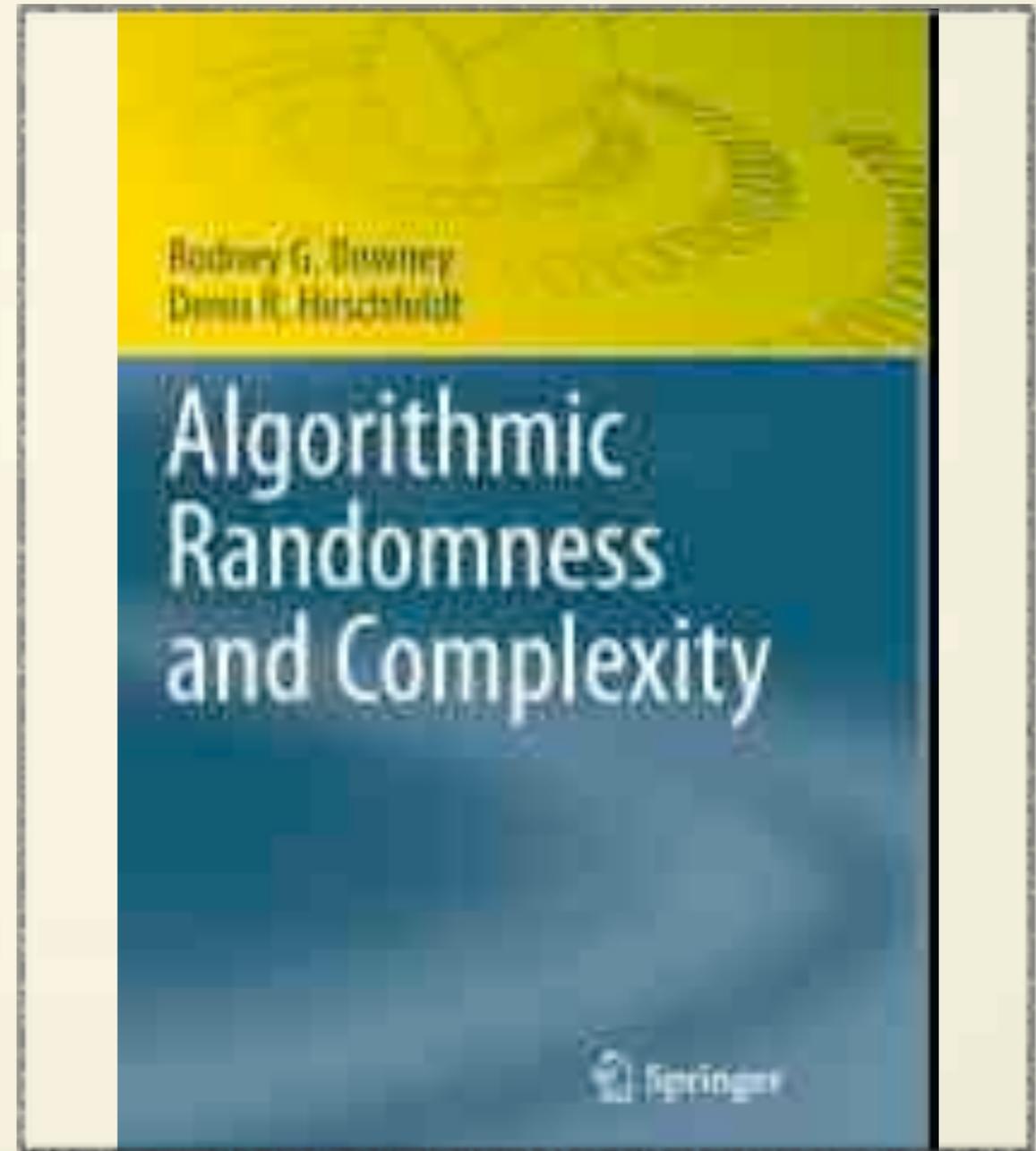
- ❖ 赤池情報量規範, 最小記述長の理論
- ❖ 乱数の解析 (杉田洋先生など)
- ❖ ゲーム論的確率論

まとめ

- ❖ ランダムという一見捉えどころのない概念に，数学的な特徴付けが与えられている。
- ❖ ランダムの概念を通して，一見捉えどころのない様々な概念が捉えられる。様々な理論の基礎として使われることを期待する。

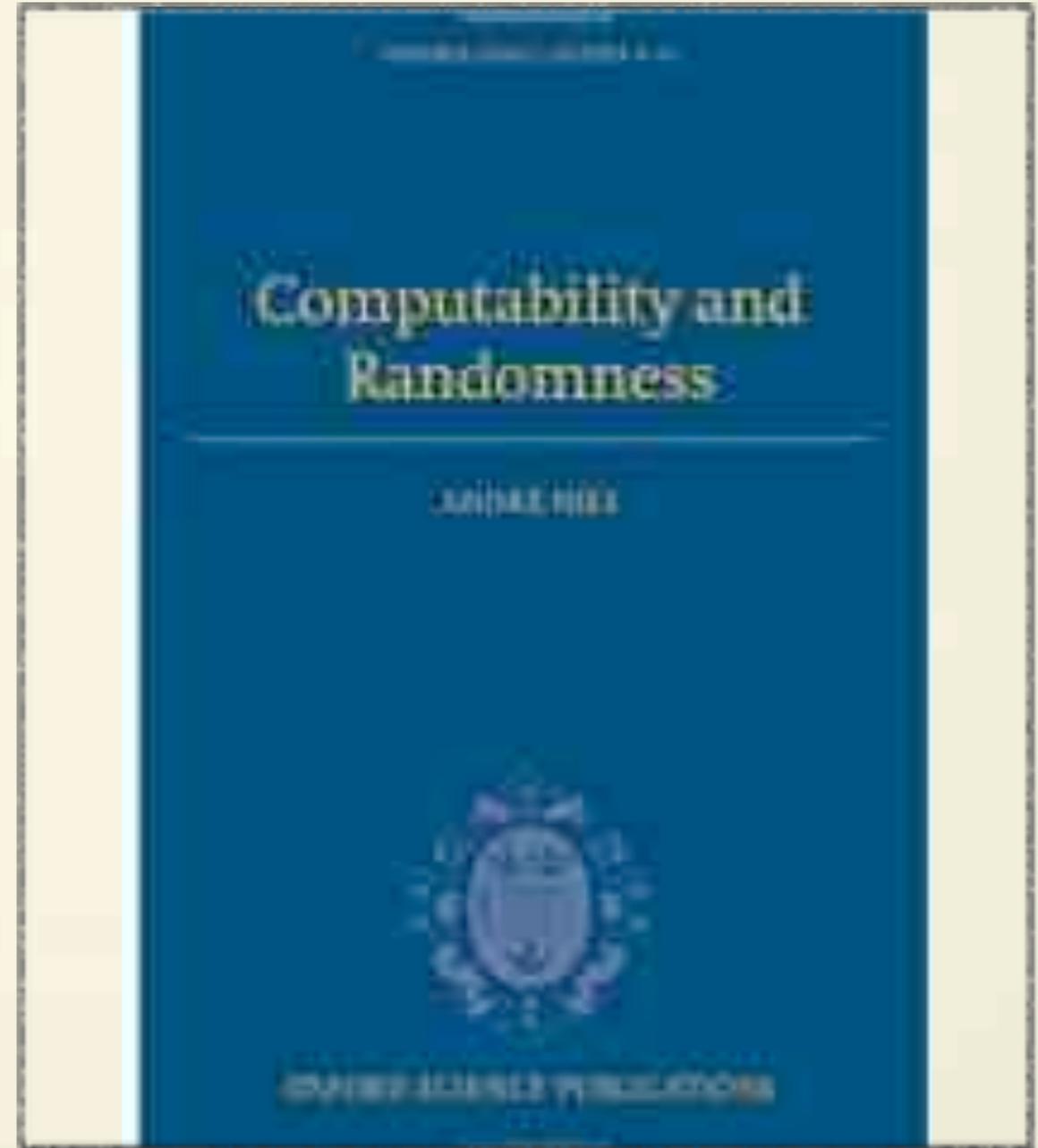
参考書その1

- ❖ Downey & Hirshfeldt
- ❖ “Algorithmic Randomness and Complexity”
- ❖ 2010年出版
- ❖ 800ページ以上！



参考書その2

- ❖ Nies
- ❖ “Computability and Randomness”
- ❖ 2009年出版



参考書その他

- ❖ Li & Vitányi. “An Introduction to Kolmogorov Complexity and Its Applications”
- ❖ Calude. “Information and Randomness”
- ❖ Hutter. “Universal Artificial Intelligence”
- ❖ Shafer & Vovk. “Probability and Finance; It’s Only a Game!”