統計モデルにおける存在論的仮定：対称性と自然種

吉井達哉 (Tatsuya Yoshii)・大塚淳 (Jun Otsuka)

一橋大学・ZEN 大学

「正当化された帰納推論を行うためには、そこで問題となる対象・現象のあり方についての実質的な仮定—存在論的仮定—が必要である」という主張は（十分に広く解釈すれば）もはや現在の科学哲学における共通理解になったと言える [Norton, 2021]。そうした帰納推論を支える存在論的構造として古くから有望視されているのが自然種である [Quine, 1969, Bird and Tobin, 2025]。自然種の定義や構造については現在でも精力的な議論がなされているが、それらのほとんどは「世界の側に自然種的構造があるおかげで、科学における帰納推論が成功する／正当化される」という**存在論ファースト**の議論となっている [Quine, 1969, Boyd, 1991, Kornblith, 1993]。しかし、このような議論は、我々が実際に行なっている帰納推論実践についてあまり多くを教えてはくれない。特にこのような枠組みは、個々の帰納推論の正当化についての問いを、「この特定の性質・種類は世界の自然種的構造と対応しているのか」という過度に形而上学的かつ概して不毛な問いへと誘導してしまう。しかし、自然種についての議論を発火させ加熱させ続けてきたのは、そもそも「適切な帰納推論とは何か」という問いとそれにまつわる哲学的問題であった [Goodman, 1983, Quine, 1969]。そこで本発表ではむしろ、帰納推論の分析からそこに内在する存在論的仮定を取り出す**推論ファースト**のアプローチでこの主題に取り組む。すなわち、正当化された帰納推論を行うために必要な仮定を取り出し、その仮定を分析することで、帰納推論を支える存在論的仮定の構造を解明することを目指す。

そのために我々は、現代科学において中心的な役割を果たす統計的推論と、そこで用いられる統計モデルに着目する。理論統計学者 Peter McCullagh は、適切な統計的推論を行うために統計モデルが満たすべき斉合性条件を、圏論的な**自然性条件**の形で定式化した [McCullagh, 2002]。このような自然性条件を適用することで、帰納推論についての「グルーのパラドクス」に対して一定の分析・解決を与えることができる [Yoshii and Otsuka, 2026]。そこで本発表では McCullagh の枠組みを拡張しつつ、自然性条件を満たすモデルを仮定することが、帰納推論の対象となる事物ないし現象についていかなる存在論的仮定を導くのかを検討する。

自然性条件は特に、モデルが指定する確率割り当てが、統計ユニットに対する様々な変換に対して整合的に変化することを要請する。すなわち、自然性条件を満たすためには、モデルの指定する確率的な関係が、統計ユニットに対する様々な反事実的な撹乱によって崩れないものでなくてはならない。このような意味での対称性—許容される変換のもとでの不変性・共変性—は、物理学をはじめとする形式諸科学において対象の本性を特徴づけると考えられてきた。この意味で、自然性条件を満たす統計モデルを仮定することは、問題となっている対象・現象のあり方についての一定の存在論的仮定を引き受けることに他ならない。本発表では、このような要請が、自然種を反事実的な撹乱に対して頑健な確率的対応関係とみなす Stable Property Cluster 説 [Slater, 2015] や、自然種を、それが整合的に変化することのできる変換の集合—動的対称性構造—によって定義する Dynamical

Kind 説 [Jantzen, 2015] と密接に関連することを指摘する。その結果として、これらの形而上学的に定義・分析されてきた自然種構造が、正当化された帰納推論の不可欠な仮定としてどのように現れるのかを明らかにする。このような試みは従来の自然種論に不足していた、帰納推論実践の側からの視点を提供するものであり、自然種のより十全な理解への重要な足掛かりとなりうる。

さらに本発表は、こうした統計的・圏論的な自然種の見方を、機械学習の「存在論」の分析へと適用する。機械学習モデルは、内部表現という仕方で対象を分類することによって、世界についての「存在論」を構築していると言える。深層学習の成功は、このように学習された存在論の妥当性を示唆する一方で、入力に対して人間にはほとんど気づかれない細工を加えることで機械学習モデルの判断を誤らせる**敵対的攻撃** [Goodfellow et al., 2020] への脆弱性も指摘されている。こうした敵対的攻撃は、機械学習モデルが、我々が想定するのとは異なる対称性—すなわち異なる自然種—を学習していることを示唆する。以上を踏まえ、本発表は、敵対的事例について我々が感じる違和感を、我々と機械との間における存在論的仮定の相違として分析する。

参考文献

[Bird and Tobin, 2025] Bird, A. and Tobin, E. (2025). Natural kinds. In Zalta, E. N. and Nodelman, U., editors, *The Stanford Encyclopedia of Philosophy*. Stanford University, winter 2025 edition.

[Boyd, 1991] Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical studies*, 61(1-2):127–148.

[Goodfellow et al., 2020] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2020). Generative adversarial networks. *Commun. ACM*, 63(11):139 – 144.

[Goodman, 1983] Goodman, N. (1983). *Fact, Fiction, and Forecast*. Harvard University Press.

[Jantzen, 2015] Jantzen, B. C. (2015). Projection, symmetry, and natural kinds. *Synthese*, 192(11):3617–3646.

[Kornblith, 1993] Kornblith, H. (1993). *Inductive Inference and its Natural Ground*. MIT Press.

[McCullagh, 2002] McCullagh, P. (2002). What is a statistical model? *Annals of statistics*, 30(5):1225–1310.

[Norton, 2021] Norton, J. D. (2021). *The Material Theory of Induction*. University of Calgary Press.

[Quine, 1969] Quine, W. V. O. (1969). Natural kinds. In Kim, J. and Sosa, E., editors, *Ontological Relativity and Other Essays*, pages 114–138. Columbia University Press.

[Slater, 2015] Slater, M. H. (2015). Natural kindness. *British Journal for the Philosophy of Science*, 66(2):375–411.

[Yoshii and Otsuka, 2026] Yoshii, T. and Otsuka, J. (2026). A categorical solution to the grue paradox. *The British journal for the philosophy of science*.