Provability logic と Solovay の定理

倉橋　太志

神戸大学 M1

2009 年 11 月 27 日　数学基礎論若手の会

目次

Provability predicate と様相論理

T : Q(Robinson's arithmetic) を含む r.e. theory とする.

Provability predicate

T の provability predicate $\mathrm{Pr}_T(x)$ は以下のように構成される.

- $\mathrm{Proof}_T(y, x)$ は recursive relation
  " y は論理式 x の T における証明である "
  を Q において表現する Δ_1 論理式.
- $\mathrm{Pr}_T(x) \stackrel{\mathrm{def}}{\equiv} \exists y \mathrm{Proof}_T(y, x)$

このとき $\mathrm{Pr}_T(x)$ は Σ_1 論理式である.

Gödel は provability predicate を定義し不完全性定理を証明したが, その証明には次の補題が本質的である.

The diagonalization lemma

x のみを自由変数としてもつ任意の論理式 $\varphi(x)$ に対して

$$\mathbf{Q} \vdash \psi \leftrightarrow \varphi(\ulcorner \psi \urcorner)$$

を満たす論理式 ψ が存在する.

第一不完全性定理 (Gödel 1931)

$\mathsf{Q} \vdash \sigma \leftrightarrow \neg \mathrm{Pr}_T(\ulcorner \sigma \urcorner)$ ととれば

1. T : 無矛盾 $\Longrightarrow T \nvdash \sigma$
2. T : Σ_1-sound $\Longrightarrow T \nvdash \neg\sigma$

第二不完全性定理 (Gödel 1931)

T : 無矛盾 $\Longrightarrow T \nvdash \neg \mathrm{Pr}_T(\ulcorner 0 = 1 \urcorner)$.

これらは provability predicate $\mathrm{Pr}_T(x)$ に関する結果ともいえる.

Derivability conditions (Hilbert-Bernays-Löb)

任意の sentence φ, ψ について

$\mathbf{D1}: T \vdash \varphi \Longrightarrow \mathbf{Q} \vdash \mathrm{Pr}_T(\ulcorner \varphi \urcorner)$

$\mathbf{D2}: \mathbf{Q} \vdash \mathrm{Pr}_T(\ulcorner \varphi \urcorner) \to \mathrm{Pr}_T(\ulcorner \mathrm{Pr}_T(\ulcorner \varphi \urcorner) \urcorner)$

$\mathbf{D3}: \mathbf{Q} \vdash \mathrm{Pr}_T(\ulcorner \varphi \to \psi \urcorner) \to (\mathrm{Pr}_T(\ulcorner \varphi \urcorner) \to \mathrm{Pr}_T(\ulcorner \psi \urcorner)).$

Löb's theorem (Löb 1955)

任意の sentence φ について

$$T \vdash \mathrm{Pr}_T(\ulcorner \varphi \urcorner) \to \varphi \Longrightarrow T \vdash \varphi.$$

Formalized Löb's theorem

任意の sentence φ について

$$\mathbf{Q} \vdash \mathrm{Pr}_T(\ulcorner \mathrm{Pr}_T(\ulcorner \varphi \urcorner) \to \varphi \urcorner) \to \mathrm{Pr}_T(\ulcorner \varphi \urcorner).$$

1933 年 Gödel は

「Provability predicate $\mathrm{Pr}_T(x)$ を 様相演算子 $\Box$ とみた 様相論理の体系をみつけること」

を問題として提起した.

(1933,“ Eine Interpretation des Intuitionistischen Aussagenkalküls ”)

様相論理 (Modal logic):“ 必然性 ”と“ 可能性 ”に関する論理．

$\Box$: 必然演算子 「〜は必然である」．

$\Diamond$: 可能演算子 「〜は可能である」．

$\Box$ と $\Diamond$ は双対の関係にある．

$$\Diamond p \leftrightarrow \neg\Box\neg p$$

$$\Box p \leftrightarrow \neg\Diamond\neg p$$

様相に関してさまざまな世界観を持ちうるため, さまざまな体系が定義されている.

様相論理の言語

- **命題変数:** $p, q, r, \ldots$
- **真理記号:** $\top, \bot$
- **命題演算子:** $\neg, \wedge, \vee, \rightarrow$
- **様相演算子:** $\Box, \Diamond$

様相体系 K

様相体系 K の推論規則と公理は以下の図式である.

- **推論規則:**

  modus ponens $A, A \rightarrow B \vdash B$

  necessitation $A \vdash \Box A$

- **公理:**
  - **全てのトートロジー**
  - $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$

様相論理の諸体系

T $\Box A \to A$

K4 $\Box A \to \Box\Box A$

S4 $\Box A \to A, \Box A \to \Box\Box A$

B $\Box A \to A, A \to \Box\Diamond A$

S5 $\Box A \to A, \Diamond A \to \Box\Diamond A$

GL $\Box(\Box A \to A) \to \Box A$

Gödel の問題

Provability predicate $\mathrm{Pr}_T(x)$ に関する事実と一致する様相体系は存在するのか？

- Kripke semantics (Kripke, 1959,“ A completeness theorem in modal logic ”)
- $\Box(\Box A \to A) \to \Box A$ を公理としてもつ様相体系 GL の研究 (Smiley, 1963,“ The Logical Basis of Ethics ”)
- Kripke completeness theorem for GL (Segerberg, 1971, “ An essay in classical modal logic ”)
- The fixed point theorem for GL (de Jongh and Sambin, 1975)
- Arithmetical completeness theorem for GL (Solovay, 1976,“ Provability interpretations of Modal Logic ”)

Gödel の問題への答え

Provability predicate $\mathrm{Pr}_T(x)$ に関して T の証明できる事実は GL の定理と一致する.

様相体系 GL と Kripke model

様相体系 GL

- **推論規則:**

  modus ponens $A, A \to B \vdash B$

  necessitation $A \vdash \Box A$

- **公理:**
  - **全てのトートロジー**
  - $\Box(A \to B) \to (\Box A \to \Box B)$
  - $\Box(\Box A \to A) \to \Box A$

定理 (de Jongh-Kripke-Sambin)

任意の様相論理式 A に対して

$$\mathbf{GL} \vdash \Box A \to \Box\Box A$$

つまり GL は provability predicate $\mathrm{Pr}_T(x)$ のもつ, これまでに挙げた性質を証明できる.

Derivability conditions

任意の様相論理式 A, B について

D1 : $\mathbf{GL} \vdash A \Longrightarrow \mathbf{GL} \vdash \Box A$

D2 : $\mathbf{GL} \vdash \Box A \to \Box\Box A$

D3 : $\mathbf{GL} \vdash \Box(A \to B) \to (\Box A \to \Box B)$.

Formalized Löb's theorem

任意の様相論理式 A について

$$\mathbf{GL} \vdash \Box(\Box A \to A) \to \Box A.$$

定義 (Arithmetical realization)

以下で帰納的に定義された, 様相論理式 から 算術の sentence への翻訳 $*$ を **arithmetical realization** という.

- 各命題変数 p に対して p^* は算術 (T) のある sentence
- $\top^* \equiv \ 0 = 0, \ \bot^* \equiv \ 0 = 1$
- $(\neg A)^* \equiv \neg A^*$
- $(A \circ B)^* \equiv (A^* \circ B^*)$ ただし $\circ \in \{\wedge, \vee, \rightarrow\}$
- $(\Box A)^* \equiv \mathrm{Pr}_T(\ulcorner A^* \urcorner)$ ただし $\Diamond A \stackrel{\text{def}}{\equiv} \neg\Box\neg A$

Arithmetical soundness theorem

任意の様相論理の sentence A に対して

$$\mathbf{GL} \vdash A \Longrightarrow \forall *, T \vdash A^*$$

Solovay は逆の statement, つまり次の定理を証明した.

Arithmetical completeness theorem (Solovay 1976)

T : sound とする. このとき任意の様相論理の sentence A に対して

$$\forall *, T \vdash A^* \Longrightarrow \mathbf{GL} \vdash A$$

Solovay の定理の証明には Kripke model の概念が必要である.

定義 (Kripke model)

- 空でない集合 W とその上の二項関係 R の対 $\langle W, R\rangle$ を **Kripke frame** という.
- Kripke frame $\langle W, R\rangle$ と, $w \in W$ と命題変数 p の関係 V の対 $\langle W, R, V\rangle$ を **Kripke model** という.
- Kripke model $\mathcal{M} = \langle W, R, V\rangle$ に対して
  - W を $\mathcal{M}$ の **domain**
  - W の元を **world**
  - R を **accessibility relation**
  - V を **valuation**

  という.

定義

任意の様相論理式 A, Kripke model $\mathcal{M} = \langle W, R, V \rangle$, $w \in W$ に対して 関係

$$\mathcal{M}, w \models A$$

を以下で定義する.

- $\mathcal{M}, w \models p \iff wVp$
- $\mathcal{M}, w \models \top \quad \mathcal{M}, w \not\models \bot$
- $\mathcal{M}, w \models A \land B \iff (\mathcal{M}, w \models A$ かつ $\mathcal{M}, w \models B)$
- $\mathcal{M}, w \models A \lor B \iff (\mathcal{M}, w \models A$ または $\mathcal{M}, w \models B)$
- $\mathcal{M}, w \models A \to B \iff (\mathcal{M}, w \models A \Longrightarrow \mathcal{M}, w \models B)$
- $\mathcal{M}, w \models \Box A \iff \forall x \in W (wRx \Longrightarrow \mathcal{M}, x \models A)$
- $\mathcal{M}, w \models \Diamond A \iff \exists x \in W (wRx$ かつ $\mathcal{M}, x \models A)$

定義

R : W 上の二項関係
$\forall i \in \mathbb{N}$ に対して R^i を以下で定義する.

- $R^0 = \{(w, w) : w \in W\}$
- $R^{i+1} = \{(w, y) : \exists x(wR^i x$ かつ $xRy)\}$

Generating submodel theorem

$\mathcal{M} = \langle W, R, V \rangle$, $w \in W$ とする.
$X = \{x : \exists i\ wR^i x\}$ とし, $\mathcal{N} = \langle X, R, V \rangle$ とする.
このとき任意の様相論理式 A に対して

$$\mathcal{M}, w \models A \iff \mathcal{N}, w \models A$$

となる. $\mathcal{N}$ を w で生成される $\mathcal{M}$ の submodel という.

定義 (validity)

- Kripke model $\mathcal{M} = \langle W, R, V \rangle$ で A は valid
$\overset{\text{def}}{\iff} \forall w \in W(\mathcal{M}, w \models A)$
- Kripke frame $\mathcal{F} = \langle W, R \rangle$ で A は valid ($\mathcal{F} \models A$)
$\overset{\text{def}}{\iff} \forall V$: valuation, $\langle W, R, V \rangle$ で A は valid.

定義 (converse wellfoundness)

W 上の二項関係 R が converse wellfounded
$\overset{\text{def}}{\iff} \forall S \subseteq W(S \neq \phi)$,
$\exists w \in S$ s.t. w は R に関する極大元.

定理

任意の Kripke frame $\mathcal{F}$ について

- $\mathcal{F} \models \Box(A \to B) \to (\Box A \to \Box B)$.
- $\mathcal{F} \models A$ かつ $\mathcal{F} \models A \to B \Longrightarrow \mathcal{F} \models B$.
- $\mathcal{F} \models A \Longrightarrow \mathcal{F} \models \Box A$.
- 集合 $\{A : \mathcal{F} \models A\}$ は 命題変数への代入で閉じている.

定理

任意の Kripke frame $\langle W, R \rangle$ について

- $\langle W, R \rangle \models \Box p \to \Box\Box p \Longleftrightarrow R$ は 推移的.
- $\langle W, R \rangle \models \Box(\Box p \to p) \to \Box p \Longleftrightarrow R$ は 推移的かつ converse wellfounded.

Kripke completeness theorem for GL (Segerberg 1971)

GL $\vdash A \iff$

- R : 推移的・converse wellfounded

である任意の Kripke model $\mathcal{M} = \langle W, R, V \rangle$ で A は valid.

Finite model property for GL (Segerberg 1971)

GL $\vdash A \iff$

- W : 有限
- R : 推移的・非反射的

である任意の Kripke model $\mathcal{M} = \langle W, R, V \rangle$ で A は valid.

Solovay の定理の証明

Arithmetical completeness theorem (Solovay 1976)

T : sound とする. このとき任意の様相論理の sentence A に対して

$$\forall *, T \vdash A^* \implies \mathbf{GL} \vdash A$$

Arithmetical completeness theorem の証明の方針

対偶を示す.

(i) GL $\nvdash A$ と仮定して Kripke 反例モデルを構成.

(ii) 反例モデルを反映する算術の sentence を構成.

(iii) その sentence から $T \nvdash A^*$ となるような arithmetical realization * を構成.

(i) について

- GL $\nvdash A$ と仮定する.
- $\exists \mathcal{M} = \langle W, R, V \rangle$: Kripke model s.t.
  - $\exists w \in W$ s.t. $w \not\models A$
  - W : 有限
  - R : 推移的・非反射的
- w で生成される $\mathcal{M}$ の submodel をとる.
  $W = \{1, \cdots, n\}, w = 1, 1Ri \iff 1 < i \leq n$
  としてよい.
- $\mathcal{M}$ を $\mathcal{M}' = \langle W', R', V' \rangle$ に拡張する. つまり
  - $W' = W \cup \{0\}$
  - $R' = R \cup \{(0, i) : 1 \leq i \leq n\}$
  - $iV'p \iff iVp (1 \leq i \leq n)$ で $0V'p$ は任意.

  このとき $\mathcal{M}', 1 \not\models A$.

(iii) について.

次を満たす算術の sentence $S_i (0 \leq i \leq n)$ がつくれたとする.

(1) $0 \leq i < j \leq n \Longrightarrow T \vdash \neg(S_i \wedge S_j)$

(2) $T \vdash S_0 \vee S_1 \vee \cdots \vee S_n$

(3) $iR'j \Longrightarrow T \vdash S_i \rightarrow \neg \mathrm{Pr}_T(\ulcorner \neg S_j \urcorner)$

(4) $i \geq 1 \Longrightarrow T \vdash S_i \rightarrow \mathrm{Pr}_T(\ulcorner \neg S_i \urcorner)$

(5) $i \geq 1 \Longrightarrow T \vdash S_i \rightarrow \mathrm{Pr}_T(\ulcorner \bigvee_{j:iR'j} S_j \urcorner)$

各命題変数 p に対して realization $*$ を $p^* = \bigvee_{i:iV'p} S_i$ とする.

補題

$\forall i (1 \leq i \leq n), \forall B : A$ の subsentence に対して

- $M', i \models B \Longrightarrow T \vdash S_i \rightarrow B^*$
- $M', i \not\models B \Longrightarrow T \vdash S_i \rightarrow \neg B^*$

補題より定理が従う.

証明.

- $\mathcal{M}', 1 \not\models A$ より補題から $T \vdash S_1 \to \neg A^*$.
  $T \vdash \neg \mathrm{Pr}_T(\ulcorner \neg S_1 \urcorner) \to \neg \mathrm{Pr}_T(\ulcorner A^* \urcorner)$.
  (3) より $T \vdash S_0 \to \neg \mathrm{Pr}_T(\ulcorner A^* \urcorner)$.
- $i \geq 1$ について $\mathbb{N} \models S_i$ と仮定する.
  (4) より $T \vdash \neg S_i$ となるため $\mathbb{N} \models \neg S_i$ で矛盾.
  よって $\mathbb{N} \not\models S_i$.
  (2) より $\mathbb{N} \models S_0$.
- $\mathbb{N} \models \neg \mathrm{Pr}_T(\ulcorner A^* \urcorner)$ であり, $T \nvdash A^*$.

□

(2) $T \vdash S_0 \vee S_1 \vee \cdots \vee S_n$
(3) $iR'j \Longrightarrow T \vdash S_i \to \neg \mathrm{Pr}_T(\ulcorner \neg S_j \urcorner)$
(4) $i \geq 1 \Longrightarrow T \vdash S_i \to \mathrm{Pr}_T(\ulcorner \neg S_i \urcorner)$

補題の証明 B の構成に関する induction. $B \equiv \Box C$ のときのみ示す.

- C に対して補題が成り立つと仮定して

$$
\begin{aligned}
\mathcal{M}', i \models B &\Longrightarrow \forall j (iR'j \Longrightarrow \mathcal{M}', j \models C) \\
&\Longrightarrow \forall j (iR'j \Longrightarrow T \vdash S_j \to C^*) \\
&\Longrightarrow T \vdash \bigvee_{j:iR'j} S_j \to C^* \\
&\Longrightarrow T \vdash \mathrm{Pr}_T(\ulcorner \bigvee_{j:iR'j} S_j \urcorner) \to \mathrm{Pr}_T(\ulcorner C^* \urcorner) \\
&\Longrightarrow T \vdash S_i \to B^*
\end{aligned}
$$

$$(5)\quad i \geq 1 \Longrightarrow T \vdash S_i \to \mathrm{Pr}_T(\ulcorner \bigvee_{j:iR'j} S_j \urcorner)$$

Solovay の定理の証明

$$
\begin{aligned}
\mathcal{M}', i \not\models B &\Longrightarrow \exists j (iR'j \,\&\, \mathcal{M}', j \not\models C) \\
&\Longrightarrow \exists j (iR'j \,\&\, T \vdash S_j \to \neg C^*) \\
&\Longrightarrow \exists j (iR'j \,\&\, T \vdash \neg \mathrm{Pr}(\ulcorner \neg S_j \urcorner) \to \neg \mathrm{Pr}(\ulcorner C^* \urcorner)) \\
&\Longrightarrow T \vdash S_i \to \neg B^*
\end{aligned}
$$

(3) $iR'j \Longrightarrow T \vdash S_i \to \neg \mathrm{Pr}_T(\ulcorner \neg S_j \urcorner)$

(ii) について.

- $B(c, a, b)$ は,$c = \ulcorner \varphi(x, y) \urcorner$ のとき以下の内容を表す Σ_1 論理式とする.
  $\exists h(0), \cdots, \exists h(a)$ s.t.
  $\forall x < a$ に対して
  1. $h(0) = 0$
  $$h(x+1) = \begin{cases} i & \text{if } h(x)R'i \ \& \\ & \quad \mathrm{Proof}_T(x, \ulcorner \neg \exists z \forall y \geq z \varphi(y, \bar{i}) \urcorner) \\ h(x) & \text{otherwise} \end{cases}$$
  2. $h(a) = b$
- Diagonalization lemma より

$$T \vdash H(a, b) \leftrightarrow B(\ulcorner H(a, b) \urcorner, a, b)$$

を満たす論理式 $H(a, b)$ が存在する.

- 各 $i (0 \leq i \leq n)$ に対して

$$S_i \stackrel{\text{def}}{\equiv} \exists z \forall x \geq z H(x, \bar{i})$$

- このとき $H(a, b)$ は次の recursive function $h(x) = y$ を T において関数として表現している Σ_1 論理式.

$$\begin{aligned} h(0) &= 0 \\ h(x+1) &= \begin{cases} i & \text{if } h(x) R' i \ \& \ \mathrm{Proof}_T(x, \ulcorner \neg S_i \urcorner) \\ h(x) & \text{otherwise} \end{cases} \end{aligned}$$

この S_i は (1)〜(5) を満たす.

(1) $0 \leq i < j \leq n \Longrightarrow T \vdash \neg(S_i \wedge S_j)$

(2) $T \vdash S_0 \vee S_1 \vee \cdots \vee S_n$

(3) $iR'j \Longrightarrow T \vdash S_i \to \neg \mathrm{Pr}_T(\ulcorner \neg S_j \urcorner)$

(4) $i \geq 1 \Longrightarrow T \vdash S_i \to \mathrm{Pr}_T(\ulcorner \neg S_i \urcorner)$

(5) $i \geq 1 \Longrightarrow T \vdash S_i \to \mathrm{Pr}_T(\ulcorner \bigvee_{j : iR'j} S_j \urcorner)$

(3) $iR'j \Longrightarrow T \vdash S_i \to \neg\mathrm{Pr}_T(\ulcorner \neg S_j \urcorner)$ を示す.

証明.

$iR'j$ として次の議論を T のなかで行う.
$h(x)$ の limit を i と仮定する.
m を $\forall r \geq m$ に対して $h(r) = h(m) = i$ となる最小の数 とする.
$\neg S_j$ が T で証明可能ならば $k > m$ かつ k は $\neg S_j$ の証明の Gödel 数であるような最小の k がとれる.
このとき $h(k)R'j$ & $\mathrm{Proof}_T(\bar{k}, \ulcorner \neg S_j \urcorner)$ より
$h(k+1) = j \neq i$ で矛盾.
したがって $\neg S_j$ は T で証明可能でない. □

Solovay の結果以降の研究

Solovay の結果以降の研究

- Provability logic for Rosser provability predicate(Guaspari and Solovay, 1979)
- Predicate provability logic(Vardanyan, 1986)
- Bimodal and polymodal provability logic(Japaridze, 1988)
- Interpretability logic(Berarducci, 1990)
- Provability logic of bounded arithmetic (Berarducci and Verbrugge, 1993)
- Logic of proofs(Artemov, 1994)
- 証明論への応用 (Beklemishev, 1999)

Solovay の結果以降の研究

- Provability logic for Rosser provability predicate(Guaspari and Solovay, 1979)
- Predicate provability logic(Vardanyan, 1986)
- Bimodal and polymodal provability logic(Japaridze, 1988)
- Interpretability logic(Berarducci, 1990)
- Provability logic of bounded arithmetic (Berarducci and Verbrugge, 1993)
- Logic of proofs(Artemov, 1994)
- 証明論への応用 (Beklemishev, 1999)

Rosser predicate

$$\mathrm{Pr}_T^R(x) \stackrel{\mathrm{def}}{\equiv} \exists y(\mathrm{Proof}_T(y, x) \land \forall z \leq y \neg \mathrm{Proof}_T(z, \neg x))$$

Kreisel の問題

Rosser predicate に対して Derivability conditions のどれが成り立たないのか？

Rosser predicate に対して Solovay の定理と同様の結果は示せないだろうか？

定理 (Guaspari and Solovay 1979)

Rosser predicate に対する体系 R が存在して,

$$\forall *, T \vdash A^* \iff \mathsf{R} \vdash A$$

系

- $\mathsf{T} \nvdash \mathrm{Pr}_T^R(\ulcorner \varphi \urcorner) \to \mathrm{Pr}_T^R(\ulcorner \mathrm{Pr}_T^R(\ulcorner \varphi \urcorner) \urcorner)$
- $\mathsf{T} \nvdash \mathrm{Pr}_T^R(\ulcorner \psi \to \rho \urcorner) \to (\mathrm{Pr}_T^R(\ulcorner \psi \urcorner) \to \mathrm{Pr}_T^R(\ulcorner \rho \urcorner))$

となる sentence φ, ψ, ρ が存在する.

参考文献

- G. Boolos. *The logic of provability*. Cambridge University Press, Cambridge, 1993.
- C. Smoryński. *Self-Reference and Modal Logic*. Springer, New York, 1985.
- G. Japaridze and D. de Jongh. *The Logic of Provability*. Handbook of Proof Theory. North holland, 1998.
- Sergri N. Artemov and Lev D. Beklemishev. *Provability logic*. Handbook of Philosophical Logic. Springer, 2005.
- R. Verbrugge. *Provability logic*. Stanford Encyclopedia of Philosophy, WWW, 2003.