Математическая логика

(mk.cs.msu.ru → Лекционные курсы → Математическая логика (группы 318, 241))

Лекция 10

Аксиоматические теории

Основные свойства теорий
(адекватность интерпретациям,
непротиворечивость, разрешимость, полнота)

Формальная арифметика

Арифметика Пеано

Теорема Гёделя о неполноте

Лектор:
Подымов Владислав Васильевич

E-mail:
valdus@yandex.ru

ВМК МГУ, 2019, весенний семестр

Вступление

$$\varphi: \quad \mathbf{2} + \mathbf{2} = \mathbf{4}$$

Попробуем *доказать* справедливость этого утверждения логическими методами

С одной стороны, φ — *формула логики предикатов*:

- **2** и **4** — константы
- $+^{(2)}$ — функциональный символ
- $=^{(2)}$ — предикатный символ

С другой стороны, компоненты формулы имеют конкретный смысл (*интерпретацию*): число 2, число 4, операция сложения чисел, отношение равенства чисел

А что такое "число N", "операция +" и "отношение ="?

Вступление

$$\varphi: \quad \mathbf{2} + \mathbf{2} = \mathbf{4}$$

Что такое “2” и “4”

Один из способов определения этих чисел выглядит так:

- $\mathbf{4}$ — это число, следующее за $\mathbf{3}$: $\quad \mathbf{s}(\mathbf{3})$
- $\mathbf{3}$ — это $\mathbf{s}(\mathbf{2})$
- $\mathbf{2}$ — это $\mathbf{s}(\mathbf{1})$
- $\mathbf{1}$ — это $\mathbf{s}(\mathbf{0})$
- $\mathbf{0}$ в этом примере особо определять не понадобится

Остановимся на такой сигнатуре логики предикатов:

$$\sigma = \left\langle \{\mathbf{0}\}, \{+^{(2)}, \mathbf{s}^{(1)}\}, \{=^{(2)}\} \right\rangle$$

Утверждение φ записывается в сигнатуре σ так:

$$\mathbf{s}(\mathbf{s}(\mathbf{0})) + \mathbf{s}(\mathbf{s}(\mathbf{0})) = \mathbf{s}(\mathbf{s}(\mathbf{s}(\mathbf{s}(\mathbf{0}))))$$

Вступление

$$\varphi: \quad \mathbf{s}(\mathbf{s}(\mathbf{0})) + \mathbf{s}(\mathbf{s}(\mathbf{0})) = \mathbf{s}(\mathbf{s}(\mathbf{s}(\mathbf{s}(\mathbf{0}))))$$

Что такое "$+$"

Одно из определений операции сложения *целых неотрицательных чисел* выглядит так:
это двуместная операция $+$, такая что для любых чисел x и y верно следующее:

- $\mathrm{x} + 0 = \mathrm{x}$
- $\mathrm{x} + (\mathrm{y} + 1) = (\mathrm{x} + \mathrm{y}) + 1$

Два пункта этого определения можно записать как две формулы:

- $A_{+\mathbf{0}}$: $\forall \mathrm{x}\, (\mathrm{x} + \mathbf{0} = \mathrm{x})$
- $A_{+\mathbf{s}}$: $\forall \mathrm{x}\, \forall \mathrm{y}\, (\mathrm{x} + \mathbf{s}(\mathrm{y}) = \mathbf{s}(\mathrm{x} + \mathrm{y}))$

Вступление

$$\varphi: \quad \mathbf{s}(\mathbf{s}(\mathbf{0})) + \mathbf{s}(\mathbf{s}(\mathbf{0})) = \mathbf{s}(\mathbf{s}(\mathbf{s}(\mathbf{s}(\mathbf{0}))))$$

Что такое "="

Прежде всего, отношение равенства чисел — это **отношение эквивалентности**:

- $A_{r=}$: $\forall \mathrm{x}\ (\mathrm{x} = \mathrm{x})$ (*оно рефлексивно*)
- $A_{s=}$: $\forall \mathrm{x}\ \forall \mathrm{y}\ ((\mathrm{x} = \mathrm{y}) \to (\mathrm{y} = \mathrm{x}))$ (*оно симметрично*)
- $A_{t=}$: $\forall \mathrm{x}\ \forall \mathrm{y}\ \forall \mathrm{z}\ ((\mathrm{x} = \mathrm{y})\,\&(\mathrm{y} = \mathrm{z}) \to (\mathrm{x} = \mathrm{z}))$ (*оно транзитивно*)

В определении равенства чисел содержатся и другие свойства, отличающие равенство от других отношений эквивалентности

Из всех таких свойств в примере понадобится следующее:
если чи́сла x и y равны, то чи́сла, следующие за x и y, также равны

- $A_{=\mathbf{s}}$: $\forall \mathrm{x}\ \forall \mathrm{y}\ ((\mathrm{x} = \mathrm{y}) \to (\mathbf{s}(\mathrm{x}) = \mathbf{s}(\mathrm{y})))$

Вступление

$$\varphi\colon\quad \mathbf{s}(\mathbf{s}(\mathbf{0})) + \mathbf{s}(\mathbf{s}(\mathbf{0})) = \mathbf{s}(\mathbf{s}(\mathbf{s}(\mathbf{s}(\mathbf{0}))))$$
$$\Gamma = \{A_{+\mathbf{0}}, A_{+\mathbf{s}}, A_{r=}, A_{s=}, A_{t=}, A_{=\mathbf{s}}\}$$
$$\sigma = \left\langle \{\mathbf{0}\}, \left\{+^{(2)}, \mathbf{s}^{(1)}\right\}, \left\{=^{(2)}\right\} \right\rangle$$

Покажем, что $\Gamma \models \varphi$:
(для наглядности $\mathbf{s}(\mathbf{0})$, $\mathbf{s}(\mathbf{s}(\mathbf{0}))$, $\mathbf{s}(\mathbf{s}(\mathbf{s}(\mathbf{0})))$ и $\mathbf{s}(\mathbf{s}(\mathbf{s}(\mathbf{s}(\mathbf{0}))))$ записываются как $\mathbf{1}$, $\mathbf{2}$, $\mathbf{3}$ и $\mathbf{4}$ соответственно)

$$\begin{array}{lcll}
A_{+\mathbf{0}} & \models & \mathbf{2}+\mathbf{0}=\mathbf{2} & (\varphi_1) \\
A_{=\mathbf{s}}, \varphi_1 & \models & \mathbf{s}(\mathbf{2}+\mathbf{0})=\mathbf{3} & (\varphi_2) \\
A_{+\mathbf{s}} & \models & \mathbf{2}+\mathbf{1}=\mathbf{s}(\mathbf{2}+\mathbf{0}) & (\varphi_3) \\
A_{t=}, \varphi_2, \varphi_3 & \models & \mathbf{2}+\mathbf{1}=\mathbf{3} & (\varphi_4) \\
A_{=\mathbf{s}}, \varphi_4 & \models & \mathbf{s}(\mathbf{2}+\mathbf{1})=\mathbf{4} & (\varphi_5) \\
A_{+\mathbf{s}} & \models & \mathbf{2}+\mathbf{2}=\mathbf{s}(\mathbf{2}+\mathbf{1}) & (\varphi_6) \\
A_{t=}, \varphi_5, \varphi_6 & \models & \mathbf{2}+\mathbf{2}=\mathbf{4} & (\varphi_7)
\end{array}$$

Следовательно, $\{A_{+\mathbf{0}}, A_{=\mathbf{s}}, A_{+\mathbf{s}}, A_{t=}\} \models \varphi_7$, а значит, $\Gamma \models \varphi$

И как это доказывает, что два плюс два — четыре?

Вступление

$$\varphi\colon\quad \mathbf{2}+\mathbf{2}=\mathbf{4}$$
$$\Gamma=\{A_{+\mathbf{0}}, A_{+\mathbf{s}}, A_{r=}, A_{s=}, A_{t=}, A_{=\mathbf{s}}\}$$
$$\sigma=\left\langle\{\mathbf{0}\}, \left\{+^{(2)}, \mathbf{s}^{(1)}\right\}, \left\{=^{(2)}\right\}\right\rangle$$
$$\Gamma\models\varphi$$

Рассмотрим такую интерпретацию $\mathcal{I}^{\sigma}_{ar}$ сигнатуры σ:

- предметная область — $\mathbb{N}_0=\{0,1,2,\dots\}$
- $\mathbf{0}$, $+$, $\mathbf{s}$ и $=$ оцениваются как число 0, операции сложения чисел и увеличения чисел на единицу и отношение равенства чисел

Формулы в Γ подобраны так, чтобы было верно $\mathcal{I}^{\sigma}_{ar}\models\Gamma$

φ — логическое следствие множества Γ, а значит, $\mathcal{I}^{\sigma}_{ar}\models\varphi$:
если интерпретировать $\mathbf{2}$, $\mathbf{4}$, $+$ и $=$ как числа 2 и 4, операцию сложения чисел и отношение равенства чисел, то формула $\mathbf{2}+\mathbf{2}=\mathbf{4}$ выполняется

Вступление

$$\varphi\colon \quad \mathbf{2} + \mathbf{2} = \mathbf{4}$$
$$\Gamma = \{A_{+\mathbf{0}}, A_{+\mathbf{s}}, A_{r=}, A_{s=}, A_{t=}, A_{=\mathbf{s}}\}$$
$$\sigma = \left\langle \{\mathbf{0}\}, \left\{+^{(2)}, \mathbf{s}^{(1)}\right\}, \left\{=^{(2)}\right\} \right\rangle$$
$$\Gamma \models \varphi$$

Сигнатурой σ задана совокупность понятий,
которые допускается использовать в формулировке высказываний:
константа $\mathbf{0}$, операции $\mathbf{s}$ и $+$, отношение $=$

Множеством Γ задан набор *a priori* верных свойств,
определяющих смысл этих понятий:
утверждений, не требующих доказательств, то есть **аксиом**

Формула φ — это высказывание, справедливость которого
доказывается с использованием справедливости аксиом: **теорема**

Вступление

$$\varphi: \quad \mathbf{2} + \mathbf{2} = \mathbf{4}$$
$$\Gamma = \{A_{+\mathbf{0}}, A_{+\mathbf{s}}, A_{r=}, A_{s=}, A_{t=}, A_{=\mathbf{s}}\}$$
$$\sigma = \left\langle \{\mathbf{0}\}, \left\{+^{(2)}, \mathbf{s}^{(1)}\right\}, \left\{=^{(2)}\right\} \right\rangle$$
$$\Gamma \models \varphi$$

Если множество аксиом (Γ) подобрано “достаточно хорошо”, то оно без изменений может быть использовано при доказательстве многих других *справедливых* теорем

Набор аксиом, вместе с выбранной сигнатурой позволяющий формулировать и доказывать теоремы, называется **аксиоматической теорией**

Если в качестве аксиом и теорем выбраны формулы логики предикатов (первого порядка), то такой теории присваивается название

аксиоматическая теория первого порядка

Аксиоматические теории

Выберем сигнатуру σ алфавита логики предикатов

Теория[1] $\mathcal{T}$ сигнатуры σ — это множество предложений (*сигнатуры* σ)

Формулы, принадлежащие теории $\mathcal{T}$,
называются аксиомами этой теории

Логические следствия теории $\mathcal{T}$
называются теоремами этой теории

Если теория ясна из контекста,
то будем *теоремы теории* называть просто *теоремами*

---

[1] Полное название: аксиоматическая теория первого порядка

Аксиоматические теории

Формула φ общезначима в теории $\mathcal{T}$, если $\forall \widetilde{\mathtt{x}^{\mathtt{n}}}\ \varphi(\widetilde{\mathtt{x}^{\mathtt{n}}})$ — теорема

Другое название: $\mathcal{T}$-общезначима

Обозначение: $\models_{\mathcal{T}} \varphi$

Формула φ противоречива в теории $\mathcal{T}$,

если формула $\neg\varphi$ $\mathcal{T}$-общезначима

Другие названия:

невыполнима в теории $\mathcal{T}$, $\mathcal{T}$-противоречива, $\mathcal{T}$-невыполнима

Обозначение: $|\not\models_{\mathcal{T}} \varphi$[1]

Формула φ выполнима в теории $\mathcal{T}$,

если она не является $\mathcal{T}$-противоречивой

Другое название: $\mathcal{T}$-выполнима

Обозначение: $|\models_{\mathcal{T}} \varphi$[1]

[1] Как и раньше, это обозначение необщеизвестно, его придумал я, чтобы сэкономить место на слайдах

Аксиоматические теории

Достаточно исследовать только одно из трёх свойств (общезначимость, выполнимость, невыполнимость в теории):

Утверждение

формула $\varphi(\widetilde{\mathtt{x}}^{\mathtt{n}})$ $\mathcal{T}$-общезначима

$\psi = \forall \widetilde{\mathtt{x}}^{\mathtt{n}}\ \varphi(\widetilde{\mathtt{x}}^{\mathtt{n}})$

$\varphi(\widetilde{\mathtt{x}}^{\mathtt{n}}) = \neg\psi(\widetilde{\mathtt{x}}^{\mathtt{n}})$ $\qquad$ $\psi(\widetilde{\mathtt{x}}^{\mathtt{n}}) = \neg\varphi(\widetilde{\mathtt{x}}^{\mathtt{n}})$ $\qquad$ предложение ψ $\mathcal{T}$-общезначимо

формула $\psi(\widetilde{\mathtt{x}}^{\mathtt{n}})$ $\mathcal{T}$-невыполнима

$\varphi = \exists \widetilde{\mathtt{x}}^{\mathtt{n}}\ \psi(\widetilde{\mathtt{x}}^{\mathtt{n}})$

$\varphi = \psi$ противоположный ответ $\qquad$ предложение φ $\mathcal{T}$-невыполнимо

формула $\varphi(\widetilde{\mathtt{x}}^{\mathtt{n}})$ $\mathcal{T}$-выполнима

$\psi = \exists \widetilde{\mathtt{x}}^{\mathtt{n}}\ \varphi(\widetilde{\mathtt{x}}^{\mathtt{n}})$

предложение ψ $\mathcal{T}$-выполнимо

Доказательство.Напрямую следует из определений $\mathcal{T}$-выполнимости, $\mathcal{T}$-невыполнимости и $\mathcal{T}$-общезначимости

Аксиоматические теории

Пример:

$$\sigma = \left\langle \{\mathbf{0}\}, \left\{\mathbf{s}^{(1)}, +^{(2)}\right\}, \left\{=^{(2)}\right\} \right\rangle$$
$$\Gamma = \{A_{+\mathbf{0}}, A_{+\mathbf{s}}, A_{r=}, A_{s=}, A_{t=}, A_{=\mathbf{s}}\}$$
$$\varphi\colon \mathbf{2} + \mathbf{2} = \mathbf{4}$$

Γ — теория сигнатуры σ

$\models_\Gamma \varphi$, а значит, предложение φ является Γ-общезначимым

$\mathcal{I}^\sigma_{ar} \models \Gamma$, но $\mathcal{I}^\sigma_{ar} \not\models \neg\varphi$, а значит, $\Gamma \not\models \neg\varphi$, и

- предложение $\neg\varphi$ не является Γ-общезначимым
- предложение φ не является Γ-противоречивым
- предложение φ является Γ-выполнимым

$\models_\Gamma \neg\neg\varphi$, а значит, предложение $\neg\varphi$ является Γ-противоречивым

Проблема общезначимости формул в теории

Проблема общезначимости формул в теории $\mathcal{T}$ формулируется так:

для заданной формулы φ проверить,
является ли она $\mathcal{T}$-общезначимой:

$$\models_{\mathcal{T}} \varphi \; ?$$

Основные свойства теорий

Элементарная теория интерпретации $\mathcal{I}$ сигнатуры σ ($\mathsf{Th}(\mathcal{I})$) — это множество всех предложений сигнатуры σ, истинных в $\mathcal{I}$

Для любой интерпретации $\mathcal{I}$ существует элементарная теория (*по определению*), но этот факт никак не помогает проверять истинность формул в интерпретации $\mathcal{I}$:
$\models_{\mathsf{Th}(\mathcal{I})} \varphi(\widetilde{\mathtt{x}^{\mathtt{n}}})$ — синоним записи $\forall \widetilde{\mathtt{x}^{\mathtt{n}}}\, \varphi \in \mathsf{Th}(\mathcal{I})$, а значит, и $\mathcal{I} \models \varphi$

Основные свойства теорий

В частности, существует и элементарная теория интерпретации $\mathcal{I}_{ar}^{\sigma}$ ($\sigma = \langle\{\mathbf{0}\}, \{\mathbf{s}^{(1)}, +\}, \{=\}\rangle$)

При этом ранее была предложена *какая-то* теория Γ, позволяющая доказывать справедливость *каких-то* арифметических утверждений

Более того, можно **наугад** выписать несколько предложений в сигнатуре σ, объявить их аксиомами, а все логические следствия аксиом — "арифметическими" теоремами, не задумываясь об интерпретации $\mathcal{I}_{ar}^{\sigma}$

Далее будет введено несколько понятий, позволяющих рассуждать о том, чем одни теории "лучше" других:

- **адекватность интерпретациям**
- **непротиворечивость**
- **разрешимость**
- **полнота**

Основные свойства теорий

Теория $\mathcal{T}$ адекватна классу интерпретаций $\mathfrak{I}$, если любая интерпретация из $\mathfrak{I}$ является моделью всех предложений множества $\mathcal{T}$

Например, теория Γ, предложенная ранее при доказательстве того, что *два плюс два — четыре*, адекватна интерпретации $\mathcal{I}^{\sigma}_{ar}$

Другой пример: теория частичных порядков — это теория сигнатуры $\langle \emptyset, \emptyset, \{<\} \rangle$, содержащая две аксиомы:

- аксиома антирефлесивности строгого частичного порядка:
$$\forall \mathrm{x}\ \neg(\mathrm{x} < \mathrm{x})$$
- аксиома транзитивности строгого частичного порядка:
$$\forall \mathrm{x}\ \forall \mathrm{y}\ \forall \mathrm{z}\ ((\mathrm{x} < \mathrm{y})\,\&(\mathrm{y} < \mathrm{z}) \to (\mathrm{x} < \mathrm{z}))$$

Теория частичных порядков адекватна множеству всех интерпретаций, оценивающих символ “$<$” как отношение строгого частичного порядка

Основные свойства теорий

Теория $\mathcal{T}$ непротиворечива, если она адекватна хотя бы одной интерпретации, и противоречива в противном случае

Утверждение. Любая формула φ является $\mathcal{T}$-общезначимой и $\mathcal{T}$-невыполнимой и не является $\mathcal{T}$-выполнимой для любой противоречивой теории $\mathcal{T}$

Доказательство.
Множество предложений $\mathcal{T}$ не имеет ни одной модели
Тогда по *определению логического следствия*:

1. $\mathcal{T} \models \varphi$, то есть формула φ $\mathcal{T}$-общезначима
2. $\mathcal{T} \models \neg\varphi$, то есть формула φ $\mathcal{T}$-противоречива и, следовательно, не является $\mathcal{T}$-выполнимой ▼

Таким образом, все противоречивые теории **абсолютно бессмысленны**

Основные свойства теорий

Теория $\mathcal{T}$ называется разрешимой, если проблема $\mathcal{T}$-общезначимости формул *алгоритмически разрешима*

Содержательно, разрешимость теории означает, что существует способ *автоматически* доказать или опровергнуть **каждую** теорему, которую в принципе можно доказать или опровергнуть

Осталось определить понятие **полноты** теории

Начнём издалека — представим себе, что теория $\mathcal{T}$ разрабатывалась так:

- выбрана **одна** интерпретация $\mathcal{I}$
- в теорию $\mathcal{T}$ включено не очень много аксиом, но столько, чтобы ими в точности описывался смысл всех констант, операций и отношений, задаваемых $\mathcal{I}$

Как строго определить фразу “в точности описывается смысл”?

Основные свойства теорий

"аксиомами теории $\mathcal{T}$ в точности описывается смысл понятий, задаваемых интерпретацией $\mathcal{I}$"

Вариант 1: интерпретация $\mathcal{I}$ является единственной моделью множества $\mathcal{T}$

Интерпретации $\mathcal{I} = \left\langle \mathrm{D}_{\mathcal{I}}, \overline{\mathrm{Const}}, \overline{\mathrm{Func}}, \overline{\mathrm{Pred}} \right\rangle$ и $\mathcal{J} = \left\langle \mathrm{D}_{\mathcal{J}}, \overline{\overline{\mathrm{Const}}}, \overline{\overline{\mathrm{Func}}}, \overline{\overline{\mathrm{Pred}}} \right\rangle$ изоморфны, если существует взаимно-однозначное отображение $\tau : \mathrm{D}_{\mathcal{I}} \to \mathrm{D}_{\mathcal{J}}$, такое что

- $\overline{\overline{\mathbf{c}}} = \tau(\overline{\mathbf{c}})$ $\quad (\mathbf{c} \in \mathrm{Const})$
- $\overline{\overline{\mathbf{f}}}(\tau(d_1), \dots, \tau(d_n)) = \tau(\overline{\mathbf{f}}(d_1, \dots, d_n))$ $\quad (\mathbf{f}^{(n)} \in \mathrm{Func};\ d_1, \dots, d_n \in \mathrm{D}_{\mathcal{I}})$
- $\overline{\overline{\mathrm{P}}}(\tau(d_1), \dots, \tau(d_n)) = \overline{\mathrm{P}}(d_1, \dots, d_n)$ $\quad (\mathrm{P}^{(n)} \in \mathrm{Pred};\ d_1, \dots, d_n \in \mathrm{D}_{\mathcal{I}})$

Основные свойства теорий

"аксиомами теории $\mathcal{T}$ в точности описывается смысл понятий, задаваемых интерпретацией $\mathcal{I}$"

Вариант 1: **интерпретация $\mathcal{I}$ является единственной моделью множества $\mathcal{T}$**

Утверждение. Если $\mathcal{I} \models \mathcal{T}$ и интерпретация $\mathcal{J}$ изоморфна интерпретации $\mathcal{I}$, то $\mathcal{J} \models \mathcal{T}$

Доказательство.Очевидно?
($\mathcal{I}$ и $\mathcal{J}$ отличаются только тем, как именно названы предметы)

Следствие

Любая непротиворечивая теория имеет бесконечно много моделей

Значит, *вариант 1* несостоятелен

Основные свойства теорий

"аксиомами теории $\mathcal{T}$ в точности описывается смысл понятий, задаваемых интерпретацией $\mathcal{I}$"

Вариант 2: **все модели теории $\mathcal{T}$ изоморфны**

Мощность предметной области любой интерпретации можно увеличить так:[1]

- произвольно выберем предмет d
- добавим в предметную область новые предметы: столько, чтобы мощность множества этих предметов была больше мощности исходной предметной области
- доопределим все оценки интерпретации так, чтобы добавленные предметы были неотличимы от d

---

[1] Здесь записано доказательство утверждения о несостоятельности варианта 2, но строгая формулировка, и тем более доказательство, не приводятся. Чтобы записать всё строго, требуется углубиться в теорию множеств — делать это сейчас нет ни времени, ни желания.

Основные свойства теорий

"аксиомами теории $\mathcal{T}$ в точности описывается смысл понятий, задаваемых интерпретацией $\mathcal{I}$"

Вариант 2: **все модели теории $\mathcal{T}$ изоморфны**

Расширенная интерпретация **неизоморфна** исходной, так как не существует биекции между множествами различной мощности

При этом расширенная интерпретация отличается от исходной только тем, что одному исходному предмету d присвоено несколько названий

Значит, *вариант 2* тоже несостоятелен

Основные свойства теорий

“аксиомами теории $\mathcal{T}$ в точности описывается смысл понятий, задаваемых интерпретацией $\mathcal{I}$”

Интерпретации $\mathcal{I}$, $\mathcal{J}$ элементарно эквивалентны, если
$$\mathsf{Th}(\mathcal{I}) = \mathsf{Th}(\mathcal{J})$$

Вариант 3 (подходящий): все модели теории $\mathcal{T}$ элементарно эквивалентны

Утверждение. Любые изоморфные интерпретации элементарно эквивалентны, и при этом существуют неизоморфные элементарно эквивалентные интерпретации

Доказательство. Попробуйте самостоятельно

Основные свойства теорий

Теория $\mathcal{T}$ называется полной, если для любого предложения φ верно хотя бы одно из двух соотношений: $\models_{\mathcal{T}} \varphi$, $\models_{\mathcal{T}} \neg\varphi$

Утверждение. Теория $\mathcal{T}$ полна $\Leftrightarrow$ все её модели элементарно эквивалентны

Доказательство.
($\Rightarrow$): Рассмотрим предложение φ

Если $\models_{\mathcal{T}} \varphi$, то для любой модели $\mathcal{I}$ теории $\mathcal{T}$ верно: $\mathcal{I} \models \varphi$

Иначе $\models_{\mathcal{T}} \neg\varphi$, и для любой модели $\mathcal{I}$ теории $\mathcal{T}$ верно: $\mathcal{I} \models \neg\varphi$

($\Leftarrow$): Рассмотрим модель $\mathcal{I}$ теории $\mathcal{T}$ и предложение φ

Если $\mathcal{I} \models \varphi$, то для любой модели $\mathcal{J}$ теории $\mathcal{T}$ верно $\mathcal{J} \models \varphi$, а значит, $\models_{\mathcal{T}} \varphi$

Иначе $\mathcal{I} \models \neg\varphi$, а значит, для любой модели $\mathcal{J}$ теории $\mathcal{T}$ верно $\mathcal{J} \models \neg\varphi$, и тогда $\models_{\mathcal{T}} \neg\varphi$ ▼

Основные свойства теорий

Пример: теория частичных порядков

$$\mathcal{T}_{<} = \left\{ \begin{array}{c} \forall x\ \neg(x < x) \\ \forall x\ \forall y\ \forall z\ ((x < y)\,\&(y < z) \to (x < z)) \end{array} \right\}$$

неполна:

- $\not\models_{\mathcal{T}_{<}} \exists x\ \exists y\ (x < y)$
  - так как существует частичный порядок, содержащий только несравнимые элементы
- $\not\models_{\mathcal{T}_{<}} \neg\exists x\ \exists y\ (x < y)$
  - так как существует частичный порядок, содержащий пару сравнимых элементов

Формальная арифметика

Остановимся подробнее на том, как может выглядеть теория, предназначенная для доказательства **арифметических** теорем

Для начала остановимся на такой сигнатуре (сигнатуре формальной арифметики):

$$\sigma_{ar} = \left\langle \{\mathbf{0}\}, \left\{\mathbf{s}^{(1)}, +^{(2)}, \times^{(2)}\right\}, \left\{=^{(2)}\right\} \right\rangle$$

$\mathcal{I}_{ar}$ — обозначение "естественной" арифметической интерпретации сигнатуры σ_{ar}, в которой символы чисел, операций и отношений оцениваются как соответствующие числа и арифметические операции и отношения

Формальной арифметикой в широком смысле называется любая теория, адекватная интерпретации $\mathcal{I}_{ar}$

Числами будем называть предметы произвольной модели формальной арифметики

Формальная арифметика

Пример: множество Γ, предложенное в начале лекции для доказательства того, что *два плюс два — четыре*, является формальной арифметикой:

$$\left\{\begin{array}{ll} A_{+\mathbf{0}}: & \forall \mathrm{x}\,(\mathrm{x}+\mathbf{0}=\mathrm{x}) \\ A_{+\mathbf{s}}: & \forall \mathrm{x}\,\forall \mathrm{y}\,(\mathrm{x}+\mathbf{s}(\mathrm{y})=\mathbf{s}(\mathrm{x}+\mathrm{y})) \\ A_{r=}: & \forall \mathrm{x}\,(\mathrm{x}=\mathrm{x}) \\ A_{s=}: & \forall \mathrm{x}\,\forall \mathrm{y}\,((\mathrm{x}=\mathrm{y})\rightarrow(\mathrm{y}=\mathrm{x})) \\ A_{t=}: & \forall \mathrm{x}\,\forall \mathrm{y}\,\forall \mathrm{z}\,((\mathrm{x}=\mathrm{y})\,\&(\mathrm{y}=\mathrm{z})\rightarrow(\mathrm{x}=\mathrm{z})) \\ A_{=\mathbf{s}}: & \forall \mathrm{x}\,\forall \mathrm{y}\,((\mathrm{x}=\mathrm{y})\rightarrow(\mathbf{s}(\mathrm{x})=\mathbf{s}(\mathrm{y}))) \end{array}\right\}$$

Но это *не очень хорошая* арифметика:

- нельзя ни доказать, ни опровергнуть утверждение $\neg(\mathbf{2}+\mathbf{2}=\mathbf{3})$, так как среди моделей теории есть как $\mathcal{I}_{ar}$, так и такая, в которой все числа равны
- отсутствуют аксиомы, описывающие смысл умножения, то есть в моделях теории символ $\times$ может оцениваться как произвольная операция

Формальная арифметика

Другой пример: арифметика Пеано — это множество аксиом

$$\left\{\begin{array}{ll}
A_{\times\mathbf{0}}: & \forall \mathrm{x}\, (\mathrm{x} \times \mathbf{0} = \mathbf{0}) \\
A_{\times\mathbf{s}}: & \forall \mathrm{x}\, \forall \mathrm{y}\, (\mathrm{x} \times \mathbf{s}(\mathrm{y}) = \mathrm{x} \times \mathrm{y} + \mathrm{x}) \\
A_{+\mathbf{0}}: & \forall \mathrm{x}\, (\mathrm{x} + \mathbf{0} = \mathrm{x}) \\
A_{+\mathbf{s}}: & \forall \mathrm{x}\, \forall \mathrm{y}\, (\mathrm{x} + \mathbf{s}(\mathrm{y}) = \mathbf{s}(\mathrm{x} + \mathrm{y})) \\
A_{r=}: & \forall \mathrm{x}\, (\mathrm{x} = \mathrm{x}) \\
A_{s=}: & \forall \mathrm{x}\, \forall \mathrm{y}\, ((\mathrm{x} = \mathrm{y}) \to (\mathrm{y} = \mathrm{x})) \\
A_{t=}: & \forall \mathrm{x}\, \forall \mathrm{y}\, \forall \mathrm{z}\, ((\mathrm{x} = \mathrm{y}) \,\&\, (\mathrm{y} = \mathrm{z}) \to (\mathrm{x} = \mathrm{z})) \\
A_{=+\mathbf{s}}: & \forall \mathrm{x}\, \forall \mathrm{y}\, ((\mathrm{x} = \mathrm{y}) \to (\mathbf{s}(\mathrm{x}) = \mathbf{s}(\mathrm{y}))) \\
A_{=-\mathbf{s}}: & \forall \mathrm{x}\, \forall \mathrm{y}\, ((\mathbf{s}(\mathrm{x}) = \mathbf{s}(\mathrm{y})) \to (\mathrm{x} = \mathrm{y})) \\
A_0: & \forall \mathrm{x}\, \neg(\mathbf{0} = \mathrm{x}) \\
A_{ind}: & \varphi\{\mathrm{x}/\mathbf{0}\} \,\&\, \forall \mathrm{x}\, (\varphi \to \varphi\{\mathrm{x}/\mathbf{s}(\mathrm{x})\}) \to \forall \mathrm{x}\, \varphi
\end{array}\right\}$$

A_{ind} — это схема аксиом индукции: бесконечное множество аксиом, по одной для каждой формулы $\varphi(\mathrm{x})$

А насколько *хороша* арифметика Пеано, и можно ли придумать арифметику *ещё лучше*?

Теорема Гёделя о неполноте

Любая рекурсивно перечислимая[1] формальная арифметика неполна

Набросок доказательства

Докажем более простое утверждение:

любая *конечная* теория с моделью $\mathcal{I}_{ar}$ неполна

От противного предположим, что существует конечная полная теория $\mathcal{T}$, такая что $\mathcal{I}_{ar} \models \mathcal{T}$

Докажем, что в любой такой теории $\mathcal{T}$ необходимо присутствует парадокс лжеца:

существует предложение,
утверждающее, что это предложение ложно

[1] "Множество X рекурсивно перечислимо" = "Существует алгоритм, последовательно вычисляющий все элементы X в процессе бесконечной работы"

Теорема Гёделя о неполноте

Набросок доказательства

Существование парадокса лжеца в формальной арифметике основывается на том, что арифметическими формулами можно описать *всё, что можно вычислить*

Вычислимая функция — это частично определённое отображение $f : \mathbb{N}_0 \to \mathbb{N}_0$, такое что существует реализующий его алгоритм (*машина Тьюринга, алгоритм Маркова, ...*)

(*предполагаю, что понятие вычислимой функции вам знакомо, поэтому подробнее на нём не останавливаюсь*)

График функции $f : \mathbb{N}_0 \to \mathbb{N}_0$ — это множество всех пар чисел (i, j), таких что значение $f(i)$ определно и равно j

Теорема Гёделя о неполноте

Набросок доказательства

Отношение $R \subseteq \mathbb{N}_0^n$ арифметизуемо, если существует формула $\varphi(\widetilde{\mathrm{x}^{\mathrm{n}}})$ в сигнатуре формальной арифметики, такая что

$$\mathcal{I}_{ar} \models \varphi(\widetilde{\mathrm{x}^{\mathrm{n}}})[\widetilde{d^n}] \quad \Leftrightarrow \quad (\widetilde{d^n}) \in R$$

Лемма(*об арифметизации*)

График любой вычислимой функции арифметизуем

(*это утверждение весьма нетривиально, но доказываться не будет*)

Вычислимые функции работают только с числами из $\mathbb{N}_0$, при этом в доказательстве потребуется арифметизация алгоритмов, работающих с формулами

Для этого опишем нумерацию Гёделя: “алгоритмичное” сопоставление числа каждой формуле

Теорема Гёделя о неполноте

Набросок доказательства

Сопоставим натуральное число каждому символу алфавита сигнатуры σ_{ar}:

- $g(\mathbf{0}) = 1, \quad g(+) = 2, \quad g(\times) = 3, \quad g(=) = 4$
- $g(() = 5, \quad g(,) = 6, \quad g()) = 7$
- $g(\&) = 8, \quad g(\vee) = 9, \quad g(\to) = 10, \quad g(\neg) = 11, \quad g(\forall) = 12,$ $g(\exists) = 13$
- $g(\mathtt{x}_1) = 14, \quad g(\mathtt{x}_2) = 15, \quad g(x_3) = 16, \quad \dots$

Сопоставим каждой формуле φ натуральное число (код формулы)

$$g(\varphi) = p_1^{g(\varphi[1])} \times p_2^{g(\varphi[2])} \times \dots \times p_{|\varphi|}^{g(\varphi[|\varphi|])},$$ где

- p_i — i-е простое число
- $\varphi[i]$ — i-й символ в записи формулы φ
- $|\varphi|$ — размер записи формулы φ

Теорема Гёделя о неполноте

Набросок доказательства

Утверждение. **Существует алгоритм, проверяющий, является ли число i, $i \in \mathbb{N}_0$, кодом формулы**

Утверждение. **Существует алгоритм, вычисляющий формулу по её коду**

Утверждение. **Существует алгоритм, вычисляющий код формулы, поданной на вход**

Используя тот же приём со степенями простых чисел, можно аналогичным образом определить код

- конечной последовательности формул
- конечной семантической таблицы
- конечного табличного вывода

Теорема Гёделя о неполноте

Набросок доказательства

В теореме о полноте табличного вывода был описан алгоритм построения успешного вывода $Tab(\varphi)$ для таблицы $\langle\, \mathcal{T} \mid \varphi \,\rangle$, где φ — произвольная $\mathcal{T}$-общезначимая формула

Утверждение. Существует алгоритм, останавливающийся тогда и только тогда, когда на вход подан код какой-либо общезначимой формулы φ, и выдающий в ответ код вывода $Tab(\varphi)$

Как следствие, арифметизуемым будет график такой функции:

$$f_t(i) = \begin{cases} g(Tab(g^{-1}(i))), & \text{если } i \text{ — код } \mathcal{T}\text{-общезначимой формулы} \\ \text{не определено}, & \text{иначе} \end{cases}$$

Теорема Гёделя о неполноте

Набросок доказательства

Это означает, что существует формула $Proof(\mathtt{x}, \mathtt{y})$, такая что

$$\mathcal{I}_{ar} \models Proof(\mathtt{x}, \mathtt{y})[d_1, d_2]$$

$$\Leftrightarrow$$

d_2 — код табличного вывода $Tab(\varphi)$,

где φ — формула, кодом которой является число d_1

Рассмотрим такую формулу $Val(\mathtt{x})$: $\exists \mathtt{y}\ Proof(\mathtt{x}, \mathtt{y})$

$\mathcal{I}_{ar} \models Val[d] \quad \Leftrightarrow \quad d$ — код формулы, истинной в $\mathcal{I}_{ar}$

Так свойство истинности формул арифметики можно выразить на языке самой арифметики, и теперь наконец-таки можно попытаться формализовать парадокс лжеца

Теорема Гёделя о неполноте

Набросок доказательства

Лемма(*о диагонали*). Для любой арифметической формулы $\varphi(\mathtt{x})$ существует арифметическое предложение ψ, такое что

$$\mathcal{I}_{ar} \models (\psi \leftrightarrow \varphi(\mathtt{x}))[g(\psi)]$$

(это второе нетривиальное утверждение, оставленное без доказательства)

Применим лемму о диагонали к формуле $\varphi = \neg Val$:

Существует предложение ψ, такое что

$$\mathcal{I}_{ar} \models (\psi \leftrightarrow \neg Val(\mathtt{x}))[g(\psi)]$$

Предложение ψ — формализация **парадокса лжеца**:

Предложение ψ истинно в $\mathcal{I}_{ar}$ $\Leftrightarrow$ оно ложно в $\mathcal{I}_{ar}$

Теорема Гёделя о неполноте

Итог: если взять любой набор арифметических свойств, с которым *хоть как-нибудь* можно работать (бесконечно перечислять), то обязательно найдётся арифметическая теорема, которую нельзя ни доказать, ни опровергнуть

Следствия

- Никакая конечная формальная арифметика не полна
- Арифметика Пеано неполна
- *Если Вы начнёте перечислять какие угодно конкретные арифметические свойства и включать их в список аксиом, то никогда не получите полную арифметику*
- Элементарная теория интерпретации $\mathcal{I}_{ar}$ не является рекурсивно перечислимой
- Полная формальная арифметика неразрешима

Теорема Гёделя о неполноте

Негативный результат теоремы Гёделя может показаться странным в свете того, что выбран довольно узкий фрагмент арифметики: в арифметике в целом содержится намного больше операций и отношений, чем $+$, $\times$, $\mathbf{s}$ и $=$

Следует иметь в виду, что кажущаяся узость этого фрагмента обманчива: если разрешить использовать только эти операции и отношения, а также число 0 и логические операции, то можно записать в виде формулы логики предикатов арифметическую функцию, реализуемую **любым алгоритмом**
(*с этого начиналось доказательство теоремы Гёделя о неполноте*)