Теорема Вильямса

Сергей Николенко

Теория экономических механизмов — ИТМО, весна 2008

Outline

Bilateral trade

- Начнём с примера bilateral trade.
- Один хочет продать, другой — купить.
- У продавца своё распределение себестоимости C; в частности, $c \in [c_0, c_1]$.
- У покупателя — своё распределение ценности V; в частности, $v \in [v_0, v_1]$.

Постановка задачи

- Распределения всем известны, конкретные стоимости — нет.
- Предположим, что конфликт *может* возникнуть, т.е. $v_0 < c_1$.
- Можно ли построить механизм так, чтобы торговля происходила тогда и только тогда, когда выгодно обоим?

Формально

- Формально говоря, механизм должен определить:
  - p — сколько покупатель заплатит;
  - r — сколько продавец получит.
- Эффективен механизм, если объект продан тогда и только тогда, когда $v > c$.

Теорема о невозможности

Теорема

В вышеописанной задаче не существует механизма, который бы был эффективен, правдив, рационален и у которого в то же время сходился бы бюджет.

- Это называется *теорема Майерсона–Саттертуэйта*.

Доказательство

- Рассмотрим механизм VCG. Он работает в данном случае так (проверьте!):
  - Покупатель объявляет v, продавец объявляет c.
  - Если $v \leq c$, ничего не происходит.
  - Если $v > c$, покупатель платит $\max\{C, v_0\}$, а продавец получает $\min\{v, c_1\}$.

Доказательство

- Механизм правдивый (проверьте!) и эффективный (объект продаётся iff $v > c$).
- Более того, он рационален:
  - у покупателя с ценностью v_0 ожидаемая прибыль равна 0, дальше — больше;
  - у продавца с ценностью c_1 ожидаемая прибыль равна 0, дальше — больше.

Доказательство

- Но вот беда: если $v_0 < c_1$, это значит, что когда вообще есть обмен, $\min\{v, c_1\} > \max\{c, v_0\}$.
- То есть продавец получает строго больше, чем платит покупатель.
- Значит, VCG не может сбалансировать бюджет.

Доказательство

- А любой другой хороший механизм, по теореме об эквивалентности доходности, должен на константу отличаться от VCG.
- Но в VCG продавец с себестоимостью c_1 получает 0, то есть уменьшить доход продавца, сохранив рациональность, не получится.
- И покупатель с ценностью v_0 получает 0, т.е. увеличить платёж, сохранив рациональность, тоже не получится.
- Итого доказали теорему.

Outline

История вопроса

- Про bilateral trade придумали Майерсон и Саттертуэйт (1983).
- А обобщение, которое сейчас буду рассказывать — это статья Williams (1999), «A characterization of efficient, bayesian incentive compatible mechanisms».
- Эта невозможность тоже будет следовать из теоремы об эквивалентности.

Вспоминаем определения

Определение

Квазилинейная функция полезности агента i с типом θ_i имеет вид

$$u_i(o,\theta_i) = u_i(p_i, a, \theta_i) = v_i(a,\theta_i) - p_i,$$

где исход o определяет выбор $a \in \mathcal{K}$ из дискретного множества $\mathcal{K}$ и выплату p_i, производимую агентом.

Агенты с квазилинейными преференциями

- У агента с квазилинейными преференциями есть *функция оценки* (valuation function) $v_i(a, \theta_i)$, $a \in \mathcal{K}$.
- Например, в аукционе, где продаётся одна вещь, $\mathcal{K} = \{0, 1\}$ — агент либо получит эту вещь, либо не получит.
- p_i в этом случае — выплата агента продавцу.

Постановка задачи

- Для начала предположим, что тип агента лежит в интервале $\theta_i \in [\underline{\theta}_i, \overline{\theta}_i] \subset \mathbb{R}$.
- Обозначим через θ_i тип агента, а через θ_i^* — тип, который он говорит.

Постановка задачи

- $U_i(\theta_i^* \mid \theta_i)$ — ожидаемая прибыль (utility) агента i:
$$U_i(\theta_i^* \mid \theta_i) = \boldsymbol{E}_{\boldsymbol{\theta}_{-i}}[u_i(p_i(\theta^*, \boldsymbol{\theta}_{-i}), a(\theta^*, \boldsymbol{\theta}_{-i}), \theta_i)].$$
- U_i складывается из V_i и P_i:
$$V_i(\theta_i^* \mid \theta_i) = \boldsymbol{E}_{\boldsymbol{\theta}_{-i}}[v_i(a(\theta^*, \boldsymbol{\theta}_{-i}), \theta_i)],$$
$$P_i(\theta_i^* \mid \theta_i) = \boldsymbol{E}_{\boldsymbol{\theta}_{-i}}[p_i(\theta^*, \theta_i)],$$
$$U_i(\theta_i^* \mid \theta_i) = V_i(\theta_i^* \mid \theta_i) - P_i(\theta_i^* \mid \theta_i).$$

Постановка задачи

- Тогда правдивость означает, что

$$U_i(\theta_i) = U_i(\theta_i \mid \theta_i) \geq U_i(\theta_i^* \mid \theta_i) \quad \forall \theta_i^*, \theta_i \in \Theta_i.$$

- Рациональность: $U_i(\theta_i) \geq 0$ для всех θ_i.
- Баланс бюджета (ex ante!) означает, что ожидаемая сумма выплат неотрицательна:

$$\boldsymbol{E}\left[\sum_i v_i(a(\theta), \theta_i) - U_i(\theta_i)\right] = \boldsymbol{E}\left[\sum_i p_i(\theta)\right] \geq 0.$$

Groves mechanisms

- Вспомним механизм VCG:

$$M_i^V(\mathbf{x}) = W(\alpha_i, \mathbf{x}_{-i}) - W_{-i}(\mathbf{x}).$$

Упражнение. Доказать, что механизм VCG в терминах квазилинейных преференций выглядит как

$$p_i(\theta) = -\sum_{j \neq i} v_j(a(\theta), \theta_j) + k_i,$$

где k_i — константа (не будем сейчас специфицировать, какая — нас интересует всё семейство механизмов, т.н. Groves mechanisms, которые друг от друга на константу отличаются).

Теорема об огибающей

- Есть в мат. анализе такая теорема об огибающей (envelope theorem).
- Рассмотрим задачу оптимизации $F(a) = \max_x f(x, a)$.
- Тогда при достаточно хороших условиях дифференцируемости
$$\frac{dF(a)}{da} = \left.\frac{\partial f(x^*, a)}{\partial a}\right|_{x^*=x(a)},$$
где $x(a)$ — точка, в которой достигается максимум.
- Иначе говоря, можно продифференцировать f по a и вычислить в точке максимума.

Применяем теорему об огибающей

- Применим её к нашей ситуации:

$$\frac{dU_i(\theta_i)}{d\theta_i} = \left.\frac{\partial U_i(\theta_i^* \mid \theta_i)}{\partial \theta_i}\right|_{\theta_i^*=\theta_i} = \left.\frac{\partial V_i(\theta_i^* \mid \theta_i)}{\partial \theta_i}\right|_{\theta_i^*=\theta_i}.$$

- Иначе говоря, получается, что

$$U_i(\theta_i) = U_i(\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\theta_i} \left.\frac{\partial V_i(\theta_i^* \mid \tau_i)}{\partial \tau_i}\right|_{\theta_i^*=\tau_i} d\tau_i.$$

Применяем теорему об огибающей

- $U_i(\theta_i) = U_i(\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\theta_i} \left. \frac{\partial V_i(\theta_i^* | \tau_i)}{\partial \tau_i} \right|_{\theta_i^* = \tau_i} d\tau_i.$
- Это и даёт нам результат об эквивалентности всех механизмов, т.к. $V_i(\theta_i^* \mid \tau_i)$ зависит только от правила $a(\theta)$ и ценностей агентов v_i, но не от деталей механизма.
- Механизмы Гровса, таким образом, покрывают всё множество «хороших» механизмов. Это и есть основная теорема.

Что такое хорошо

- Осталось понять, что такое «хороший» механизм.
- По идее, в теореме «хороший» должно означать «правдивый и эффективный».
- Но у нас тут ещё какие-то ограничения на дифференцируемость появлялись.
- Вообще говоря, нельзя применять теорему об огибающей к произвольному правдивому и эффективному механизму. Поэтому мы сейчас всё докажем по-другому.

Формулировка теоремы

Теорема

Рассмотрим проблему социального выбора с квазилинейными преференциями. Предположим также, что

- *множества типов Θ_i — связные открытые подмножества $\mathbb{R}^{n_i}$,*
- *ожидаемые interim ценности агентов $V_i(\theta_i^* \mid \theta_i)$ непрерывно дифференцируемы на $\Theta_i \times \Theta_i$ в точках, в которых $\theta_i^* = \theta_i$.*

Тогда механизмы Гровса являются правдивыми и эффективными для этой задачи, и interim ожидаемые ценности агентов $U_i(\theta_i^* \mid \theta_i)$ любого правдивого и эффективного механизма совпадают с ценностями одного из механизмов Гровса.

Что ограничивают ограничения

- Важно понять, что именно ограничивают ограничения. Они на V_i.
- А V_i, как мы уже отмечали, зависит только от свойств *задачи*, но не механизма.
- То есть мы немного ограничиваем класс задач, к которым применима теорема.
- Но при этом класс механизмов остаётся полным — доказываем для *всех* правдивых эффективных механизмов.

Переформулировка теоремы

- Как обычно, a good formula stays for ever. Теорема будет следовать из формулы.

Теорема (Вильямса)

В условиях теоремы Вильямса функция доходности любого правдивого эффективного механизма для любой пары типов $\theta_i, \theta_i^* \in \Theta_i$ имеет вид

$$U_i(\theta_i) = U_i(\theta_i^*) + \int_C D_{\theta_i} V_i(\theta_i^* \mid \theta_i)\Big|_{\theta_i^*=\tau, \theta_i=\tau} d\tau,$$

где C — гладкая кривая от θ_i^* к θ_i внутри Θ_i, $\tau \in \mathbb{R}^{n_i}$.

Доказательство

- Обозначим $\boldsymbol{\rho} \in \mathbb{R}^{n_i}$ — некоторый единичный вектор, $s \in \mathbb{R}$.
- Правдивость гласит, что $\forall \theta_i \in \Theta_i$

$$\begin{aligned} U_i(\theta_i) &\geq U_i(\theta_i + s\boldsymbol{\rho} \mid \theta_i), \\ U_i(\theta_i + s\boldsymbol{\rho}) &\geq U_i(\theta_i \mid \theta_i + s\boldsymbol{\rho}). \end{aligned}$$

- Суммарно:

$$\begin{aligned} & U_i(\theta_i \mid \theta_i + s\boldsymbol{\rho}) - U_i(\theta_i) \leq \\ & \qquad \leq U_i(\theta_i + s\boldsymbol{\rho}) - U_i(\theta_i) \leq \\ & \qquad\qquad \leq U_i(\theta_i + s\boldsymbol{\rho}) - U_i(\theta_i + s\boldsymbol{\rho} \mid \theta_i). \end{aligned}$$

Доказательство

- $U_i(\theta_i \mid \theta_i + s\boldsymbol{\rho}) - U_i(\theta_i) \leq U_i(\theta_i + s\boldsymbol{\rho}) - U_i(\theta_i) \leq U_i(\theta_i + s\boldsymbol{\rho}) - U_i(\theta_i + s\rho \mid \theta_i)$.
- Сократим там P_i слева и справа и разделим на s:

$$\frac{V_i(\theta_i \mid \theta_i + s\boldsymbol{\rho}) - V_i(\theta_i)}{s} \leq \\ \leq \frac{U_i(\theta_i + s\boldsymbol{\rho}) - U_i(\theta_i)}{s} \leq \\ \leq \frac{V_i(\theta_i + s\boldsymbol{\rho}) - V_i(\theta_i + s\boldsymbol{\rho} \mid \theta_i)}{s}.$$

Доказательство

- $\frac{V_i(\theta_i|\theta_i+s\rho)-V_i(\theta_i)}{s} \leq \frac{U_i(\theta_i+s\rho)-U_i(\theta_i)}{s} \leq \frac{V_i(\theta_i+s\rho)-V_i(\theta_i+s\rho|\theta_i)}{s}$.
- Устремим $s \to 0$. По условию о дифференцируемости V_i, левая часть сходится к производной $V_i(\tau_i^* \mid \tau_i)$ по направлению ρ в точке $\tau_i^* = \tau_i = \theta_i$.

Доказательство

- Правая часть раскладывается на
$$\frac{V_i(\theta_i + s\boldsymbol{\rho}) - V_i(\theta_i)}{s} - \frac{V_i(\theta_i + s\boldsymbol{\rho} \mid \theta_i) - V_i(\theta_i)}{s}.$$
- Первое слагаемое сходится к производной $V_i(\tau_i)$ по τ_i по направлению $\boldsymbol{\rho}$ в $\tau_i = \theta_i$.
- Второе слагаемое — к производной $V_i(\tau_i^* \mid \tau_i)$ по τ_i^* по направлению $\boldsymbol{\rho}$ в $\tau_i^* = \tau_i = \theta_i$.
- А вся правая часть — к производной $V_i(\tau_i^* \mid \tau_i)$ по τ_i по направлению $\boldsymbol{\rho}$ в $\tau_i^* = \tau_i = \theta_i$.
- Значит,
$$D_{\theta_i} U_i(\theta_i) = D_{\theta_i} V_i(\theta_i^* \mid \theta_i)\Big|_{\theta_i^* = \tau, \theta_i = \tau}.$$
- Отсюда следует теорема, т.к. производная по предположению непрерывна.

Обсуждение

- Это очень показательный метод доказательства.
- Собственно, это развитие исходной идеи Майерсона в максимальной (или близкой к тому) общности.
- Видно, что откуда берётся во всех таких теоремах: нужно взять изменение (приращение $s\rho$) и посмотреть, что от него изменится.

Давайте применим

- Давайте теперь применим теорему Вильямса.
- Мы бы хотели рациональные механизмы создавать.
- Посмотрим, когда это получится.

Теорема

Теорема

Рассмотрим проблему социального выбора с квазилинейными предпочтениями и типами из интервалов $\Theta_i = [\underline{\theta}_i, \overline{\theta}_i]$. *Тогда в предположениях теоремы Вильямса минимальная субсидия, которая требуется рациональному, правдивому и эффективному механизму, равна*

$$\min\left\{0, -(N-1)\boldsymbol{E}_\theta\left[\sum_{i=1}^{N} v_i(a(\theta), \theta_i)\right] + \sum_{i=1}^{N} U_i(\underline{\theta}_i)\right\}.$$

Теорема

Теорема

В частности, рациональные, правдивые и эффективные механизмы со сбалансированным бюджетом существуют тогда и только тогда, когда

$$(N-1)\boldsymbol{E}_{\theta}\left[\sum_{i=1}^{N} v_i(a(\theta),\theta_i)\right] \leq \sum_{i=1}^{N} U_i(\underline{\theta}_i).$$

Доказательство

- По теореме Вильямса, достаточно рассмотреть механизмы Гровса.
- Для них ожидаемая сумма трансферов

$$\boldsymbol{E}_\theta\left[\sum_{i=1}^{N} p_i(\theta)\right] = -\boldsymbol{E}_\theta\left[\sum_{i=1}^{N}\sum_{j\neq i} v_j(a(\theta),\theta_j)\right] + \sum_{i=1}^{N} k_i =$$
$$= -(N-1)\boldsymbol{E}_\theta\left[\sum_{i=1}^{N} v_i(a(\theta),\theta_i)\right] + \sum_{i=1}^{N} k_i.$$

- По рациональности, $U_i(\underline{\theta}_i) \geq k_i$ для всех i.
- Отсюда и получается утверждение теоремы.

Итоги

- За минувшие лекции мы уже фактически прошлись по всей классической теории дизайна механизмов.
- Уж точно по всему тому, за что Нобелевские премии давали; не считая, конечно, экономических, т.е. практических приложений (а без этого, конечно, не дадут).
- Отныне будем заниматься более узкими вещами: онлайн-аукционами, например.

Спасибо за внимание!

- Lecture notes и слайды будут появляться на моей homepage:
  http://logic.pdmi.ras.ru/~sergey/index.php?page=teaching
- Присылайте любые замечания, решения упражнений, новые численные примеры и прочее по адресам: sergey@logic.pdmi.ras.ru, snikolenko@gmail.com
- Заходите в ЖЖ smartnik.