

1, 2, 3. Формулировка задачи оптимизации поиска равновесных объемов и цен в сетевом аукционе поставщиков и потребителей одного товара с ограничениями на передачу. Определение равновесных цен в сетевом аукционе. Финансовый баланс в сетевом аукционе с ограничениями на передачу.

Постановка задачи:

c_{Π} — ценовая заявка потребителей $\Pi \in \Pi$ (готовность купить 1 единицу товара по цене не выше c_{Π})

$c_{\Gamma}, \Gamma \in \Gamma$ — ценовые заявки генераторов (готовы продать 1 единицу товара по цене не ниже c_{Γ})

Найти равновесные цены и объемы x_{Π} и x_{Γ} , которые являются **индивидуально рациональными** стратегиями участников.

Математическая модель

Функция благосостояния рынка:

$$\sum_{\Pi \in \Pi} c_{\Pi} x_{\Pi} - \sum_{\Gamma \in \Gamma} c_{\Gamma} x_{\Gamma} \rightarrow \max_{x_{\Gamma}, x_{\Pi}} \\ \sum x_{\Pi} - \sum x_{\Gamma} = 0, x_{\Pi} \in [0, \overline{x_{\Pi}}], x_{\Gamma} \in [0, \overline{x_{\Gamma}}]$$

Равновесие:

$$L(x, \lambda) = \sum_{\Pi} (c_{\Pi} - \lambda) x_{\Pi} + \sum_{\Gamma} (\lambda - c_{\Gamma}) x_{\Gamma} \rightarrow \max \sim \sum_{\Pi} \max_{x_{\Pi}} (c_{\Pi} - \lambda) x_{\Pi} + \sum_{\Gamma} \max_{x_{\Gamma}} (\lambda - c_{\Gamma}) x_{\Gamma}$$

$$c_{\Pi} - \lambda - \pi_{\Pi}^{+} + \pi_{\Pi}^{-} = 0, \Pi \in \Pi \\ \lambda - c_{\Gamma} - \pi_{\Gamma}^{+} + \pi_{\Gamma}^{-}, \Gamma \in \Gamma$$

Финансовый баланс:

$$\lambda \left(\sum_{\Pi} x_{\Pi} \right) - \lambda \left(\sum_{\Gamma} x_{\Gamma} \right) = 0, \text{ экономический смысл множителя лагранжа } \lambda \text{ равновесная цена на рынке}$$

Территориально распределенный аукцион:

В задаче фигурируют генераторы (поставщики) в узле $s: \Gamma \in s: \Gamma_s$ и потребители в узле $s: \Pi \in s: \Pi_s$.

Задача: найти равновесие с учетом переток товара между узлами и ограничений на пропускную способность сети.

Формулировка задачи:

$$(1) g(x_{\Pi}, x_{\Gamma}) = \sum_{\Pi \in \Pi} c_{\Pi} x_{\Pi} - \sum_{\Gamma \in \Gamma} c_{\Gamma} x_{\Gamma} \rightarrow \max \\ (2) \sum_{r \in B(s)} f_{rs} - \sum_{k \in A(s)} f_{sk} = \sum_{\Pi \in \Pi_s} x_{\Pi} - \sum_{\Gamma \in \Gamma_s} x_{\Gamma} \quad \forall s = \overline{1, N} \Rightarrow h(x_{\Pi}, x_{\Gamma}, f) \\ (3) -\underline{f}_{rs} \leq f_{rs} \leq \overline{f}_{rs} \quad (r, s) \Rightarrow k(f) \\ (4) 0 \leq x_{\Gamma} \leq V_{\Gamma}, \Gamma \in \Gamma \Rightarrow \overline{u}(x_{\Gamma}), \underline{u}(x_{\Gamma}) \\ (5) 0 \leq x_{\Pi} \leq V_{\Pi}, \Pi \in \Pi \Rightarrow \overline{v}(x_{\Pi}), \underline{v}(x_{\Pi})$$

x_{Γ} — объем поставщика (генератора) $\Gamma \in \Gamma$, V_{Γ} — максимальный объем;

x_{Π} — объем потребления $\Pi \in \Pi$, V_{Π} — максимальный объем.

c_{Γ} — целевая заявка поставщика (генератора) $\Gamma \in \Gamma$

c_{Π} — целевая заявка потребителя $\Pi \in \Pi$

$s = \overline{1, N}$ — индекс узла.

$A(s)$ — множество ветвей выходящих из узла $s \quad \{r | \exists (s, r)\}$

$B(s)$ — множество ветвей входящих в узел $s \quad \{r | \exists (r, s)\}$

f_{rs} — переменная потока по ветви (r, s) . Если $f_{rs} > 0$, то поток из r в s , если $f_{rs} < 0$, то поток из s в r .

- (1) — целевой функцией является функция благосостояния рынка
 (2) — материальный баланс (финансовый баланс) в узле s (ограничения баланса спроса и предложения в узле s сети)
 (3) — Ограничения на пропускную способность сети
 (4) и (5) — ограничения поставки и потребления.

Решая задачу оптимизации с помощью функции Лагранжа вводим следующие множители Лагранжа (неотрицательные):

- (2) λ_s
 (3) $\sigma_{rs}^+, \sigma_{rs}^-$ (цена пропускной способности ветви (r, s))
 (4) π_r^-, π_r^+
 (5) π_π^-, π_π^+

После решения задачи оптимизации (поиск равновесных объемов и цен) получим:

λ_s^* — равновесная цена в узле s .

x_π^*, x_r^* — объемы участников аукциона.

Для билета 2.

Выпишем функцию Лагранжа:

$$L = g(x_\pi, x_r) - \sum_s \lambda_s h(x_\pi, x_r, f) - \sum \sigma_{rs}^+ \bar{k}_{rs} + \sum \sigma_{rs}^- \underline{k}_{rs} - \sum \pi_r^+ \bar{u}_r + \sum \pi_r^- \underline{u}_r - \sum \pi_\pi^+ \bar{v}_\pi + \sum \pi_\pi^- \underline{v}_\pi$$

$$\frac{\partial L}{\partial x_r} = -c_r + \lambda_s - \pi_r^+ + \pi_r^- = 0, r \in \Gamma_s$$

$$\frac{\partial L}{\partial x_\pi} = c_\pi - \lambda_s - \pi_\pi^+ + \pi_\pi^- = 0, \pi \in \Pi_s$$

$$\lambda_s = c_r + \pi_r^+ - \pi_r^-, r \in \Gamma_s \quad c_\pi = \lambda_s + \pi_\pi^+ - \pi_\pi^-, \pi \in \Pi_s$$

Если максимальный объем $\Rightarrow \pi_r^- = 0 \Rightarrow \lambda_s$ — цена и поставщик имеет дополнительную прибыль.

Если заявка потребителя удовлетворена полностью $\Rightarrow \pi_\pi^- = 0 \Rightarrow$ потребитель платит меньше, чем $\bar{c}_\pi$.

$$\frac{\partial L}{\partial f_{rs}} = -\lambda_s + \lambda_r + \sigma_{rs}^+ - \sigma_{rs}^-$$

$$\frac{\partial L}{\partial f_{rs}} = 0 = \begin{cases} \lambda_s = \lambda_r, & -\underline{f}_{rs} < f_{rs} < \bar{f}_{rs} \\ \lambda_s = \lambda_r + \sigma_{rs}^+ > \lambda_r, & f_{rs} = \bar{f}_{rs}, \sigma_{rs}^+ > 0 \\ \lambda_s = \lambda_r - \sigma_{rs}^- < \lambda_r, & f_{rs} = -\underline{f}_{rs}, \sigma_{rs}^- > 0 \end{cases}$$

Таким образом узловая цена не убывает в направлении потока и строго возрастает, если ограничение насыщено (то есть достигает предельных ограничений на поток из узла s в узел r).

Для билета 3.

Финансовый баланс в узле s (материальный баланс).

$$\sum_{\pi \in \Pi_s} x_\pi - \sum_{r \in \Gamma_s} x_r = \sum_{r \in B(s)} f_{rs} - \sum_{k \in A(s)} f_{sk}$$

Домножаем уравнение баланса в узле s на λ_s

$$\lambda_s \left(\sum_{\pi \in \Pi_s} x_\pi - \sum_{r \in \Gamma_s} x_r \right) = \left(\sum_{r \in B(s)} f_{rs} - \sum_{k \in A(s)} f_{sk} \right)$$

Просуммируем правую и левую части по s :

$$\sum_s \lambda_s \left(\sum_{\Pi \in \Pi_s} x_{\Pi} \right) - \sum_s \left(\sum_{\Gamma \in \Gamma_s} x_{\Gamma} \right) = \text{сумма денег, заплаченных потребителями} - \text{сумма денег, полученных ген-}$$

$$\sum_s \lambda_s \left(\sum_{r \in B(s)} f_{rs} - \sum_{k \in A(s)} f_{sk} \right) = \sum_{(i,j) \in E} f_{i,j} (\lambda_j - \lambda_i) = \sum_{(i,j) \in E} \left(\sigma_{i,j}^+ \overline{f_{ij}} + \sigma_{ij}^- \underline{f_{ij}} \right) \geq 0,$$

так как

$$f_{ij}(\lambda_j - \lambda_i) = \begin{cases} 0, & \lambda_i = \lambda_j \\ \sigma_{ij}^+ \overline{f_{ij}}, & f_{ij} = \overline{f_{ij}} \\ \sigma_{ij}^- \underline{f_{ij}}, & f_{ij} = -\underline{f_{ij}} \end{cases}$$

Билет 4

Формулировка задачи выбора состава включенного генерирующего оборудования(ВСВГО). Нарушение условий индивидуальной рациональности в задаче с бинарными переменными, моделирующими состояние оборудования

Выбор состава генерирующего оборудования оказывает определяющее влияние на ценообразование в конкурентном рынке электроэнергии. Это продиктовано особенностями производства электроэнергии на современных электростанциях и вытекающими из этих особенностей процедурами планирования. Длительность цикла краткосрочного планирования составляет сутки. На предстоящие сутки оператор зоны централизованного управления, исходя из прогнозируемых условий почасового потребления, должен запланировать мощность электростанций, которые будут нести нагрузку для покрытия прогнозируемого потребления, а также будут находиться в резерве. Определяет состав работающего оборудования. Включение/выключение блоков может занимать некоторое число времени и нести издержки

Процесс краткосрочного планирования в Единой энергосистеме (ЕЭС) можно разделить на три этапа. Обозначим через X операционные сутки, в течение которых происходит производство и поставка электроэнергии. На первом этапе, в сутки $X - 2$, при помощи формализованной процедуры ВСВГО системным оператором ЕЭС осуществляется планирование включения/отключения генераторов от сети. Далее, в сутки $X - 1$, производится аукцион ценовых заявок поставщиков и потребителей электроэнергии и определяется плановые графики работы генерирующего оборудования на сутки X . Плановый график работы корректируется на балансирующем рынке (БР) в сутки X каждые два часа.

Рассмотрим задачу ВСВГО с интервалами планирования $t = \overline{1, T}$. Обозначим, через $s_{ti}, s_{ti}^u, s_{ti}^d \in \{0, 1\}$ индикаторы работы пуска и остановки для генератора $i \in B$ в момент времени t , где B - множество генераторов в система. $p_{ti}^g \geq 0$ - производство мощности генератора i в момент времени t . В общем виде задача выбора состава оборудования может быть сформулирована как задача минимизации стоимости работы, пусков и остановки генерирующего оборудования.

$$\sum_{t=1}^T \sum_{i \in B} \{s_{ti}^u \cdot \hat{C}_{ti}^u + s_{ti}^d \cdot \hat{C}_{ti}^d + GenCost(s_{ti}, p_{ti}^g)\} \rightarrow \min_{s_{ti}, s_{ti}^u, s_{ti}^d, p_{ti}^g} \quad (1)$$

s_{ti} - бинарные индикаторы (0,1) обозначающие включенное или нет состояние блока.

$\hat{C}_{ti}^u$ - стоимость пуска $\hat{C}_{ti}^d$ - стоимость остановки

В качестве критерия стоимости работы включенного генерирующего оборудования в задаче ВСВГО в настоящее время используется стоимость включенной мощности (стоимость электроэнергии плюс стоимость резерва, размещенного на включенных в сеть генераторах)

$$GenCost(s_{ti}, p_{ti}^g) = \hat{C}_{ti} \cdot \hat{P}_{ti}^{max} \cdot s_{ti}, \quad t = \overline{1, T}, i \in B \quad (2)$$

Использование данного критерия позволяет минимизировать величину включенного резерва мощности в пределах задаваемого Системным оператором норматива.

При наличии Общесистемных ограничения баланса генерации и потребления с учетом потерь

в сети

$$\sum_{i \in B} p_{ti}^g = \hat{P}_t^c + LOSS_t(p_t^g), t = \overline{1, T}, i \in B \quad (3)$$

Ограничения на максимальные потоки мощности по линиям электропередачи

$$Flow_{tk}(p_t^g) \leq \hat{F}_{tk}^{max}, t = \overline{1, T}, k = \overline{1, N_t^{sec}} \quad (4)$$

где N_t^{sec} - общее число групп линий электропередачи и трансформаторов, по которым задаются ограничения максимального поток мощности. Так же по мимо общесистемных ограничений могут быть использованы следующие технологические ограничения:

- Индикатор пуска и остановки

$$s_{ti}^u = \min(0, s_{ti} - s_{t-1,i}), s_{ti}^d = \min(0, s_{t-1,i} - s_{t,i}), t = \overline{1, T}, i \in B \quad (5)$$

- Диапазон работы включенных генераторов.

$$s_{ti} \hat{P}_{ti}^{min} \leq p_{ti}^g \leq s_{ti} \hat{P}_{ti}^{max}, t = \overline{1, T}, i \in B \quad (6)$$

- Минимальное число генераторов во включенном состоянии

$$\sum_{i \in B_k} s_{ti} \geq \hat{M}_{tk}^{on} - w_{tk}^{on}, t = \overline{1, T} \quad (7)$$

- Минимальное и максимальное производство мощности

$$\hat{P}_t^{gmin} \leq \sum_{i \in B_k} p_{ti}^g \leq s_{ti} \hat{P}_{tk}^{gmax} \quad (8)$$

5. Мат и фин балансы. Описание одного из экономических агентов в однопродуктовой модели.

A - множество экон. агентов, P - продукты (то, что производится), R - ресурсы.

Пусть $A \in A, i \in P \cup R$

$$0 \approx \frac{dQ_i^A(t)}{dt} = X_i^A(t) - C_i^A(t) - V_i^A(t) - J_i^A(t) - \sum_{B \in A, B \neq A} X_i^{AB}(t),$$

где Q_i^A - запас ресурса i у агента A , все остальные переменные ≥ 0 : X_i^A - производство, C_i^A - конечное потребление, V_i^A - затраты на пр-во, J_i^A - капитальные затраты, X_i^{AB} - передача продукта.

Пусть 1 агент производит один продукт, т.е. в уст. системе запас товара у производителя не изменяется.

$Y_i^A(t) = X_i^A(t) - V_i^A(t)$ - чистый выпуск, $Y_i^A(t) \leq M_i^A(t)$ - производственная мощ-ть.

$dM^A(t) = -\mu^A M^A(t) + \frac{J^A}{b^A}$, μ^A - коэффициент износа, b^A - вложение.

$Y_i^A = C_i^A + J_i^A + \sum_{B \in A, B \neq A} X_i^{AB}$ - ур-е баланса для стабил. предприятия.

Чтобы выйти на макроуровень, надо перейти к фин. эквивалентам:

$X_i^{AB}(t) \sim \Phi_i^{BA}(t)$ - платежи. Аналогично делается переход к другим переменным.

$Y_p^A = C_p^A + J_p^A + \sum_{B \in A} \Phi_p^{BA}(t)$, т.е. добавочная ст-ть = потреб. расходы +

инвестиции в пр-во. X_p^A - б-ем пр-ва, V_p^A - мат. затраты. $\frac{dk^A}{dt} = J_p^A - \mu^* k^A$, k^A - амортизация.

Пусть $N \subseteq A$ - подмн-во агентов гос-ва.

$Y_p^N + I_p^N = C_p^N + J_p^N + E_p^N$, т.е. ВВП + Импорт = Конеч. потребление + Накопление + Экспорт.

Это - **основное макроэкономическое тождество** или **макроэкономический баланс**.

Производство (P) в однопродуктовой модели (ОМ):

$a \in P$ - производитель. Затраты: L^a - труд, Y^a - продукт от a . Φ_Y^{Ta} - доход от продажи. $\Phi_L^{aH} = sL^a$ - зарплата, s - ставка з/п. $\Phi_s^{aT} = pS^a$ -

инвестиции, p - цена на продукт.

$\frac{dN^a}{dt} = \Phi_Y^{Ta} - \Phi_S^{aT} - \Phi_L^{aT} - T^{aH} - T^{ag} + \frac{dL^{Ba}}{dt} - \frac{dL^{aB}}{dt} - R^{aB}$ - фин баланс производителя.

Прирост кассовых остатков = чистый доход - валовые инвестиции - фонд з/п - дивиденды - налоги + прирост ссуд банков - прирост остатков расчетных счетов - процентные платежи по ссудам. $Y = \sum_{a \in P} Y^a, L =$

$$\sum_{a \in P} L^a.$$

Домохозяйство (H) в ОМ:

C^a - покупает, платит $\Phi_C^{aT} = pC^a$ (потребительские расходы). $\sum_{a \in H} C^a = C$

- все, что потребили дом. хоз-ва. З/п: $\Phi_L^{Pa} = sL^a, \sum_{a \in H} L^a = \sum_{a \in P} L^a$ - все

работают на пр-ве.

$$\frac{dN^a}{dt} = \Phi_L^{Pa} - \Phi_C^{aT} + T^{Ba} + T^{Pa} - T^{ag} + T^{ga} - \frac{dL^{aB}}{dt} + R^{Ba}$$

Прирост наличных = зп на пр-ве - потреб расходу + дивиденды банков + производителей - налоги + пособия и зп госслужащих - прирост сбережений + процентные платежи по вкладам.

Государство (G) в ОМ:

$$\frac{dN^a}{dt} = -\Phi_G^{gT} + T^{cg} + T^{Bg} + T^{Pg} + T^{Hg} - T^{gH} - \frac{dL^{cg}}{dt} - R^{gc}$$

Прирост кассовых остатков = - оплата гос закупок + прибыль ЦБ + налоги от банков + от производителей + от населения - пособия - прирост внут госдолга - процентные платежи.

Коммерческий банк (B) в модели ОМ:

$$\frac{dN^a}{dt} = -T^{aH} - T^{ag} - \frac{dL^{aP}}{dt} + \frac{dL^{Pa}}{dt} + \frac{dL^{Ha}}{dt} - \frac{dL^{aB}}{dt} + \frac{dL^{Ba}}{dt} + \frac{dL^{ca}}{dt} - \frac{dL^{ac}}{dt} + R^{Pa} - R^{aH} - (R^{Ba} - R^{aB}) - R^{ac}$$

Прирост кассовых остатков = - дивиденды - налоги - прирост ссуд произв + прирост расчет счетов произв + хозяйств - прирост ссуд другим банкам + прирост ссуд ЦБ - прирост резервов в ЦБ + проценты от произв - проценты населению - (сальдо процентов по межбанк операциям) - проценты по ссудам ЦБ

Динамические межотраслевые модели. Понятия траектории, стационарной траектории, динамического равновесия.

Динамическая модель Неймана

Модель Неймана – линейная производственная модель, которая учитывает фактор времени.

существующие технологии линейны, т.е. для выпуска продукции в объеме x_j необходимо и достаточно произвести

затраты в объемах $x_j a_{ij}$, $i = \overline{1, n}$, всех отраслей. Этот факт

Вводятся следующие предположения:

- Матрица затрат A и матрица выпусков B постоянны

Более общий подход к описанию функционирования многоотраслевой экономики предоставляет модель Неймана. Предположим, что в экономике все продукты выпускаются в результате конечного набора производственных процессов вида (a^j, b^j) , $j = \overline{1, m}$, где вектор $a^j = (a_{1j}, \dots, a_{nj})$ описывает затраты, вектор $b^j = (b_{1j}, \dots, b_{nj})$ – выпуск товаров при функционировании j -го процесса с единичной интенсивностью. Обозначим $A = (a_{ij})$, $B = (b_{ij})$. Тогда технология производства в экономике задается парой неотрицательных матриц (A, B) , которые называются, соответственно, **матрицей затрат** и **матрицей выпуска**.

Обозначим через $x = (x_1, \dots, x_n)$ вектор выпусков (интенсивностей).

- экономика функционирует в течение T периодов времени;
- в каждый период $[t-1, t]$ для производства применяется один из процессов множества $C = \{(y, z) \mid \exists x \geq 0: y = Ax, z = Bx\}$, характеризующихся вектором интенсивностей $x(t)$;
- на производство в периоде $[t, t+1]$ могут быть потрачены только те продукты, которые были произведены в период $[t-1, t]$ (модель является замкнутой), т.е. $Ax(t) \leq Bx(t-1)$. Значение величины начальных запасов $Bx(0)$ считается. Последовательность векторов $x(1), x(2), \dots, x(T) \stackrel{\text{def}}{=} \{x(t)\}$, удовлетворяющая этому неравенству, будем называть **планом**.
- производственные затраты осуществляются (и, значит, должны оплачиваться) раньше, чем реализуется произведенная за счет этих затрат продукция:

все товары имеют цены, изменяющиеся от периода к периоду: $p^i(t-1)$ – цена единицы продукции в период $[t-1, t]$, $p(t-1) = (p^1(t-1), \dots, p^n(t-1))$, $p(t-1) \geq 0$. В таком случае доход производственного процесса (a^j, b^j) в период $[t-1, t]$ составляет $\langle p(t), b^j \rangle - \langle p(t-1), a^j \rangle$. Поскольку рассматриваемая экономическая система является замкнутой, то в ней должно выполняться условие сохранения денежной массы. Это означает, что ни один из процессов не должен иметь положительного дохода: $p(t-1)A \geq p(t)B$.

Явная форма записи:

$$p(t-1)Ax(t) = p(t)Bx(t), \quad t = 1, 2, \dots, T \quad (2.10)$$

$$p(t)Bx(t-1) = p(t)Ax(t), \quad t = 1, 2, \dots, T. \quad (2.11)$$

Определение. Траектория $\{x(t)\}$ называется **стационарной**, если существует число $\nu > 0$, такое что $x(t) = \nu x(t-1)$ (или $x(t) = \nu^t x(0)$).

Для того чтобы траектория интенсивностей $\{x(t)\} = \{\nu^t x\}$ была стационарной в модели Неймана, необходимо и достаточно, чтобы $\nu Ax \leq Bx$. Аналогично, траектория цен $\{p(t)\} = \{\mu^{-t} p\}$ стационарна тогда и только тогда, когда $\mu pA \geq pB$.

Для стационарных траекторий интенсивностей и цен условия неизменности денежной массы (2.10) и (2.11) принимают, соответственно, вид: $\mu pAx = pBx$, $\nu pAx = pBx$.

Определение. Модель Неймана находится в состоянии динамического равновесия, описываемого параметрами (ν, μ, x, p) , где числа $\nu > 0$, $\mu > 0$, а x, p – неотрицательные векторы, если

$$\nu Ax \leq Bx, \quad \mu pA \geq pB, \quad \mu pAx = pBx, \quad \nu pAx = pBx. \quad (2.12)$$

Из системы (2.12) следует, что $\nu = \mu \stackrel{def}{=} \alpha$, т.е. темп роста интенсивности производства и темп падения цен совпадают.

Определение. Невырожденным положением равновесия в модели Неймана называется тройка (α, x, p) , где $\alpha > 0$, $x \in \mathbb{R}_+^m$, $p \in \mathbb{R}_+^n$, удовлетворяющая соотношениям:

$$\alpha Ax \leq Bx \quad (2.13)$$

$$\alpha pA \geq pB \quad (2.14)$$

$$pAx > 0. \quad (2.15)$$

Модель динамического межотраслевого баланса

Как и в статической модели Леонтьева, рассмотрим экономику, производящую и потребляющую n типов товаров (разделенную на n соответствующих чистых отраслей), совокупный запас которых описывает вектором $x = (x_1, \dots, x_n)$. Матрица Леонтьева A задает технологические затраты каждой из n чистых отраслей при выпуске единицы продукции. Через ξ_j обозначим производственную мощность j -й отрасли (максимально возможный выпуск, обусловленный наличием основных фондов). Тогда $\xi = (\xi_1, \dots, \xi_n)$ – вектор максимально возможного выпуска. Для увеличения выпуска каждой отрасли необходимо увеличить свои мощности, при этом должны быть затрачены продукты других отраслей. Через η_j обозначим желаемое приращение мощности

j -й отрасли, тогда вектор $(\eta_j d_{1j}, \dots, \eta_j d_{nj})$ определяет затраты всех производимых товаров, необходимые для увеличения мощности на желаемую величину η_j . Таким образом, если $\eta = (\eta_1, \dots, \eta_n)$ – вектор, определяющий желаемое увеличение мощности во всей экономике, то суммарные материальные затраты составят $D\eta$, где $D = (d_{ij})$, $i = \overline{1, n}$, $j = \overline{1, n}$.

Предположим, что для выпуска единицы продукта j -й отрасли требуется l_j трудовых затрат, тогда $l = (l_1, \dots, l_n)$ – вектор трудовых затрат. Величина L определяет общее количество занятых в экономике, $c = (c_1, \dots, c_n)$ – вектор потребления на одного рабочего (его заработная плата в натуральном выражении).

Теперь выпишем динамическую систему, описывающую развитие экономики в течение некоторого конечного числа периодов времени $t = 1, 2, \dots, T$, в которой общее число занятых, объем производства, мощности и затраты на создание новых мощностей меняются со временем.

Затраты на производство, строительство и оплату труда не превосходят объема производства:

$$Ax(t) + D\eta(t) + L(t)c \leq x(t); \quad (2.18)$$

объем производства не превосходит имеющейся мощности:

$$x(t) \leq \xi(t-1); \quad (2.19)$$

новые мощности создаются на основе имеющихся с учетом желаемого прироста:

$$\xi(t) \leq \xi(t-1) + \eta(t); \quad (2.20)$$

трудовые ресурсы ограничены:

$$\langle l, x(t) \rangle \leq L(t). \quad (2.21)$$

Кроме того, $(x(t), \xi(t), \eta(t), L(t)) \geq 0$, $t = 1, 2, \dots, T$, значение $\xi(0)$ – определяет объем производственных мощностей перед началом функционирования экономики и считается заданным.

Описание экономической системы с помощью модели (2.18)–(2.21) определяется **траекторией** $(x(t), \xi(t), \eta(t), L(t))$, $t = 1, 2, \dots, T$, удовлетворяющей указанной системе неравенств. В общем случае существует бесконечного много траекторий, удовлетворяющих системе (2.18) – (2.21), поэтому необходимо ввести критерий оценки качества траектории, который обеспечил бы возможность выбора какой-то одной, в некотором смысле «наилучшей». В общем случае таким критерием является зависящий от времени и значений координат траектории целевой функционал. Например, можно рассмотреть конечное (терминальное) состояние экономики $(x(T), \xi(T), \eta(T), L(T))$ и в качестве целевого функционала выбрать функцию

$$\langle c_1, x(T) \rangle + \langle c_2, \xi(T) \rangle + \langle c_3, \eta(T) \rangle + c_4 L(T), \quad (2.22)$$

где c_1, c_2, c_3 – n -мерные векторы, c_4 – число.

Тогда оптимизационная задача для модели динамического межотраслевого баланса формулируется следующим образом: при заданном начальном векторе производственных мощностей $\xi(0)$ и заданных коэффициентах терминального функционала (c_1, c_2, c_3, c_4) среди всех траекторий, удовлетворяющих системе (2.18)–(2.21), найти ту, которая доставляет максимум суммы (2.22).

Модель (2.18)–(2.22) можно переписать в следующем виде:

$$\begin{aligned} \max \langle \tilde{c}, \tilde{x}(T) \rangle \\ \tilde{A}\tilde{x}(t) \leq \tilde{B}\tilde{x}(t-1), \\ \tilde{x}(t) \geq 0, \quad t = 1, 2, \dots, T, \quad \tilde{x}(0) = (0, 0, \xi(0), 0) - \text{задано.} \end{aligned} \quad (2.23)$$

Здесь

$$\tilde{A} = \begin{pmatrix} A - E & 0 & D & c \\ E & 0 & 0 & 0 \\ 0 & E & -E & 0 \\ l & 0 & 0 & -1 \end{pmatrix}, \quad \tilde{B} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & E & 0 & 0 \\ 0 & E & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\tilde{x} = (x, \xi, \eta, L), \quad \tilde{c} = (c_1, c_2, c_3, c_4).$$

Билет 7

Модель Вальраса. Модель динамического равновесия вальрасовского типа, существование равновесных траекторий.

Много потребителей

Много производителей максимизируют прибыль (разница между выручкой и затратами)

Рынок товаров с совершенной конкуренцией

Замкнутая экономическая система (полученная производителями прибыль распределяется между потребителями)

Каждый участник в отдельности не влияет на цену

N типов товаров

L потребителей с функцией дохода $K(p) = (b^i, p) + I(p)$, где b^i начального запаса товаров и дополнительный доход $I_i(p)$ и функцией спроса $\Phi_i(p)$.

M производителей, каждый имеет производственный процесс $(x, y) \in \Gamma$. Прибыль производителя $\langle y - x, p \rangle$, если p – вектор цен. Цель производителя максимизация прибыли: $\langle y - x, p \rangle = \max_{(x', y') \in \Gamma} \langle y' - x', p \rangle$. $y - x$ – чистый выпуск. $Y = \{y - x \mid x \in \mathbb{R}_+^n, y \in F(x)\}$ – технологическое множество.

$\Psi(p) = \{y \in Y \mid \langle y, p \rangle = \max_{y' \in Y} \langle y', p \rangle\}$ – функция предложения.

$X \subseteq \mathbb{R}_+^n$ каждый вектор $x \in X$ соответствует набору продуктов производства. Если $y \in Y$, $y \geq 0$, то $y = 0$ Условие отсутствия рога изобилия.

Назовем **совокупным технологическим множеством** сумму

$Y = \sum_{k=1}^m Y_k$ и **функцией совокупного предложения** производственного сектора сумму $\Psi_0(p) = \sum_{k=1}^m \Psi_k(p)$.

Пусть $\bar{\Psi}_0(p) = \left\{ y \in Y \mid \langle y, p \rangle = \max_{y' \in Y} \langle y', p \rangle \right\}$ – множество производственных процессов, оптимальных с точки зрения всего производственного сектора. Несложно показать, что производственные процессы, оптимальные с точки зрения всего производственного сектора, оптимальны и с точки зрения каждого производителя. Обратное также верно, т.е. выполнено равенство:

$$\bar{\Psi}_0(p) = \sum_{k=1}^m \Psi_k(p) = \Psi_0(p).$$

Определение. Набор $(y_1^*, \dots, y_m^*; x_1^*, \dots, x_l^*; p^*)$ неотрицательных векторов называется конкурентным равновесием, если

$$y_k^* \in \Psi_k(p^*), \quad k = 1, \dots, m, \quad (3.2)$$

$$x_i^* \in \Phi_i(p^*), \quad i = 1, \dots, l, \quad (3.3)$$

и выполняются соотношения баланса спроса и предложения:

$$\sum_{k=1}^m y_k^* + \sum_{i=1}^l b^i \geq \sum_{i=1}^l x_i^*, \quad (3.4)$$

$$\left\langle p^*, \sum_{k=1}^m y_k^* + \sum_{i=1}^l b^i \right\rangle = \left\langle p^*, \sum_{i=1}^l x_i^* \right\rangle. \quad (3.5)$$

Определение. Вектор p^* называется вектором равновесных цен, многозначное отображение $\Phi(p) = \sum_{i=1}^l \Phi_i(p)$ – функцией совокупного спроса, отображение $\Psi(p) = b + \sum_{k=1}^m \Psi_k(p)$, где $b = \sum_{i=1}^l b^i$, – функцией совокупного предложения.

$$\langle p, x \rangle \leq \langle p, y \rangle, \quad x \in \Phi(p), \quad y \in \Psi(p).$$

Закон Вальраса:

Определение. Набор (y^*, x^*, p^*) называется конкурентным равновесием если $x^* \in \Phi(p^*)$, $y^* \in \Psi(p^*)$, $x^* \leq y^*$ и $\langle p^*, x^* \rangle = \langle p^*, y^* \rangle$.

Модель Эрроу-Дебре:

1. $K_i(p) = \langle p, b^i \rangle + \sum_{j=1}^m a_{ij} \langle p, y_j \rangle$, где b^i – начальный запас товаров, a_{ij} – доля доходов j -го производителя, которую получает i -й потребитель, $\sum_{i=1}^l a_{ij} = 1$ для любого j , y_j – чистый выпуск j -го производителя. Таким образом, мы предполагаем, что капитал потребителя складывается из дохода от продажи начального запаса и участия в прибыли производителей.
2. Множество $X_i \subseteq \mathbb{R}_+^n$, на котором определена ф-ия полезности $u_i(x)$, выпукло, замкнуто и неограниченно. Если $\{x^k\} \in X_i$ и $x_j^k \rightarrow \infty$, то $x_j^k \rightarrow \infty$ для всех j .
3. Функция полезности $u_i(x)$ непрерывна и вогнута на X_i .
4. Для любого i найдется такой вектор $\bar{x}^i \in X_i$, что $\bar{x}^i < b^i$ (ненулевые начальные запасы всех товаров).
5. Потребитель ненасыщаем.
6. Множество чистых выпусков k -го производителя Y_k компактно и $0 \in Y_k$, $k = 1, \dots, m$.
7. Совокупное технологическое множество $Y = \sum_{k=1}^m Y_k$ выпукло.

Билет 8

Описание модели Рамсея. Магистральное свойство сбалансированного роста.

Односекторная экономика, совокупность независимых производителей, выпускает однородный продукт

Уравнения модели:

$$\left\{ \begin{array}{l} Y(t) = C(t) + I(t), C(t) \geq 0, I(t) \geq 0 \\ Y(t) = M(t)f\left(\frac{L(t)}{M(t)}\right) \\ \frac{dM(t)}{dt} = -\mu M(t) + \frac{I(t)}{b} \\ \frac{dN(t)}{dt} = nL(t) \rightarrow L(t) = L_0 e^{nt} \\ s(t) = \frac{I(t)}{Y(t)} \\ C(t) = (1 - s(t))Y(t) \\ M(0) = M_0; L(0) = L_0 \end{array} \right.$$

$Y(t)$ - Весь произведенный продукт в единицу времени

$C(t)$ – Часть произведённого продукта, отданная потребителям

$I(t)$ - Часть произведённого продукта, отданная производителям (новые произв. мощн)

F - производственной функцией определяет зависимость объема производства от количества затрачиваемых факторов производства, $f(x)$ – функция загрузки мощности

$M(t)$ – масштабирует распределение мощностей (производственные мощности)

μ - темп выбытия (мощности стареют и эта часть постоянно убирается из процесса)

B – приростная фондоемкость

$L(t)$ – рабочая сила. Состоит из постоянной доли населения, которое растёт с темпом n

$s(t)$ – норма накопления (определяет какая доля дохода идёт на накопление, а какая на потребление)

(3.16). Перейдем к новым переменным – фондовооруженности

труда $k(t) = \frac{bM(t)}{L(t)}$ и потреблению на душу населения

$$c(t) = \frac{C(t)}{L(t)}. \text{ Тогда } \frac{Y(t)}{L(t)} = \frac{M(t)}{L(t)} f\left(\frac{L(t)}{M(t)}\right) = \frac{k(t)}{b} f\left(\frac{b}{k(t)}\right) \stackrel{\text{def}}{=} \tilde{f}(k(t))$$

и

Обозначив для краткости $\lambda = n + \mu$, получим, что модель (3.12)-

(3.16) принимает вид:

$$\frac{dk(t)}{dt} = s(t)\tilde{f}(k(t)) - \lambda k(t), \quad (3.17)$$

$$c(t) = (1 - s(t))\tilde{f}(k(t)), \quad (3.18)$$

$$k(0) \stackrel{\text{def}}{=} k_0 = \frac{bM_0}{L_0}, \quad 0 \leq s(t) \leq 1. \quad (3.19)$$

Сначала рассмотрим самый простой случай. Пусть доля накопления является постоянной: $s(t) \equiv s = const$. Найдем стационарные (равновесные) точки уравнения (3.17), т.е. такие значения $\bar{k}$, что при $k_0 = \bar{k}$ решением уравнения (3.17) будет функция $k(t) \equiv \bar{k}$. Для того чтобы найти все такие $\bar{k}$, надо найти все решения уравнения $\frac{dk(t)}{dt} = 0$, или

$$s\tilde{f}(k) - \lambda k = 0. \quad (3.20)$$

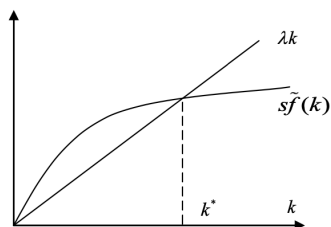


Рис. 4

Решение $k=0$ существует всегда поскольку для неоклассической пр. функция $f(0)=0$. Второго решения нет, если прямая всегда выше кривой, но это невозможно из-за условий Инады на произв. ф-ию –при стремлении к нулю у неё бесконечно большая производная; прямая не может быть всегда ниже кривой из-за второго условия Инады на бесконечности: она растёт медленнее аргумента.

Условия Инады:

$$\lim_{x_i \rightarrow 0} F'_{x_i} = \infty, \quad \lim_{x_i \rightarrow \infty} F'_{x_i} = 0$$

Поскольку пр. ф-ия вогнута, то существует ровно одно пересечение с прямой.

Для всех точек $k < k^*$ верно: $k'(t) > 0$ т.е. ф-ия возрастает, для $k > k^*$ обратное, сл-но

k^* - устойчивое равновесие, к которому стремится любое решение.

Если $k = k^*$ для модели (3.17)-(3.19), то для модели (3.12)-(3.16) имеем:

$$M(t) = \frac{k(t)L(t)}{b} = \frac{1}{b} k^* L_0 e^{nt}, \quad Y(t) = L(t) \tilde{f}(k(t)) = \tilde{f}(k^*) L_0 e^{nt},$$

$$C(t) = (1-s) \tilde{f}(k^*) L_0 e^{nt}, \quad J(t) = s L_0 e^{nt}.$$

Таким образом, все переменные растут с одинаковым темпом. Такую ситуацию называют **режимом сбалансированного роста**. Для описанной модели он обладает тем свойством, что к нему сходятся все траектории модели при постоянной норме накопления.

Добавим критерий оптимальности траектории - уровень потребления на одного трудящегося

$$c(s) = \tilde{f}(k^*(s)) - \lambda k^*(s).$$

$$\frac{d}{ds} (\tilde{f}(k^*(s)) - \lambda k^*(s)) = 0 \quad \text{или} \quad (\tilde{f}'(k^*) - \lambda) \frac{dk^*(s)}{ds} = 0.$$

Из условия (3.20) в силу того, что $\lim_{k \rightarrow \infty} \frac{\tilde{f}(k)}{k} = 0$, $\frac{dk^*(s)}{ds} > 0$, необ-

ходимое условие максимума имеет вид $\tilde{f}'(k^*) = \lambda = n + \mu$. Это соотношение носит название **правила золотого накопления, или правила Солоу**.

Рассмотрим вариант неконстантной нормы накопления

Динамическая оптимальность:

С учетом начальной системы необходимо найти максимум функционала $\langle y \rangle = \frac{1}{T} \int_{t_0}^{t_0+T} \frac{C(t)}{N(t)} dt$

В качестве фазовой переменной берем $k(t) = \frac{bM(t)}{N(t)}$, а в качестве управляющего воздействия $s(t) = \frac{I(t)}{Y(t)}$, с учетом $Y(t) = N(t)\tilde{f}(k(t))$.

$$\text{Тогда } k' = \frac{bM'N - bMN'}{N^2} = \frac{bM'}{N} - \frac{bMN'}{N^2} = k \left(\frac{M'}{M} - \frac{N'}{N} \right) = k \left(\frac{I}{bM} - (n + \mu) \right) = k \left(\frac{sY}{bM} - (n + \mu) \right) = k \left(\frac{sN\tilde{f}(k)}{bM} - (n + \mu) \right) = s(t)\tilde{f}(k(t)) - (n + \mu)k(t)$$

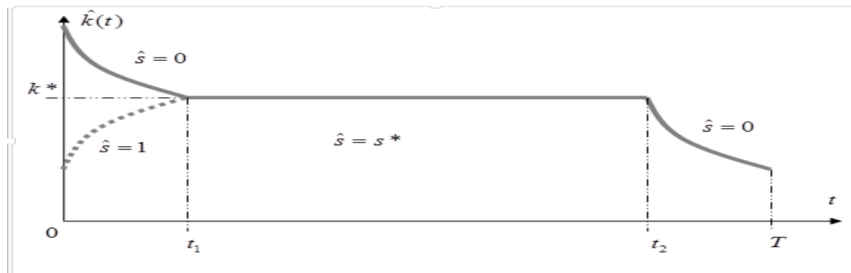
$$C(t) = Y(t) - I(t) = Y(t)(1 - s(t)) = N(t)\tilde{f}(k(t))(1 - s(t))$$

Приходим к следующей задаче оптимального управления:

$$\begin{cases} \int_0^T \tilde{f}(k(t))(1 - s(t)) dt \rightarrow \max \\ s(t) \in [0; 1] \\ k(0) = \frac{bM_0}{N_0} \\ \frac{dk}{dt} = s(t)\tilde{f}(k(t)) - (n + \mu)k(t) \end{cases}$$

Гамильтониан данной системы: $H(k, q, s) = q \left(s(t)\tilde{f}(k(t)) - (n + \mu)k(t) \right) + \tilde{f}(k(t))(1 - s(t))$

При достаточно большом горизонте планирования график фазовой переменной имеет следующий вид:



Магистральное свойство сбалансированного роста:

Если при фиксированном начальном условии k_0 увеличивать горизонт планирования T , то все большая часть оптимальной траектории будет совпадать со сбалансированным ростом, а значение функционала на оптимальной траектории будет приближаться к значению душевого потребления на оптимальном сбалансированном росте. Это явление — слабая зависимость оптимальных траекторий от начальных условий (а фактически и от конкретного вида функционала) при больших горизонтах планирования — называется магистральным свойством.

Вопрос 9

Существование равновесия в модели Неймана.

Модель Неймана описана в начале 6-го билета

Определение. Невырожденным положением равновесия в модели Неймана называется тройка (α, x, p) , где $\alpha > 0$, $x \in \mathbb{R}_+^m$, $p \in \mathbb{R}_+^n$, удовлетворяющая соотношениям:

$$\alpha Ax \leq Bx \quad (2.13)$$

$$\alpha pA \geq pB \quad (2.14)$$

$$pAx > 0. \quad (2.15)$$

Теорема. Пусть в модели Неймана $A \geq 0$, $B \geq 0$ и в матрице A нет нулевых столбцов, в матрице B нет нулевых строк. Тогда существует решение системы (2.13)-(2.15).

Условия этой теоремы означают, что нет производственных процессов, которые ничего не тратят, и всякий процесс производит продукт.

Доказательство. Рассмотрим задачу линейного программирования

$$\begin{aligned} \min u \\ (A - \lambda B)x - ue^n \leq 0, \langle x, e^m \rangle = 1, x \geq 0. \end{aligned} \quad (2.16)$$

Здесь λ – числовой параметр, x , u – переменные, e^n – n -мерный единичный вектор, e^m – m -мерный единичный вектор. Функция $u(\lambda)$, являющаяся значением задачи (2.16) в зависимости от параметра λ , обладает следующими свойствами:

- 1) $u(\lambda)$ непрерывна на всей числовой оси,
- 2) $u(0) > 0$,
- 3) $\lim_{\lambda \rightarrow \infty} u(\lambda) = -\infty$,
- 4) $u(\lambda)$ монотонно убывает.

Из этих свойств следует, что существует $\bar{\lambda} > 0$, для которого $u(\bar{\lambda}) = 0$. Обозначим через $\bar{x}$ соответствующее решение задачи (2.16). Тогда $A\bar{x} \leq \bar{\lambda} B\bar{x}$, $\bar{x} \geq 0$.

Рассмотрим задачу, двойственную к (2.16):

$$\begin{aligned} \max v \\ p(A - \lambda B) - ve^n \geq 0, \langle p, e^n \rangle = 1, p \geq 0. \end{aligned} \quad (2.17)$$

По теореме двойственности $v(\bar{\lambda}) = 0$. Обозначим через $\bar{p}$ соответствующее решение задачи (2.17). Тогда $\bar{p}A \geq \bar{\lambda} \bar{p}B$, $\bar{p} \geq 0$. Таким образом, тройка $(\bar{\lambda}^{-1}, \bar{x}, \bar{p})$ является положением равновесия в модели Неймана. Однако оно может быть вырожденным.

Для указания невырожденного положения равновесия рассмотрим множество решений уравнения $u(\lambda) = 0$. Из свойств

функции $u(\lambda)$ нетрудно заключить, что это множество является отрезком $[\lambda_0, \lambda_1]$, причем $\lambda_0 > 0$. Покажем, что существует невырожденное равновесие, соответствующее числу $\alpha = \lambda_1^{-1}$.

Среди всех решений $\bar{p}$ задачи (2.17) выберем p^* , для которого вектор p^*A обладает наибольшим числом положительных компонент. Теперь покажем, что существует вектор x^* , который образует вместе с указанными α и p^* невырожденное положение равновесия. Допустим противное. Тогда из того, что $(A - \lambda_1 B)x \leq 0$ и $x \geq 0$ следует, что $p^*Ax \leq 0$. По теореме о линейных неравенствах [3] это означает, что существует неотрицательный вектор q , удовлетворяющий условию $q(A - \lambda_1 B) \geq p^*A$.

Вектор q обладает следующими свойствами:

$$(qA)_j > \lambda_1(qB)_j, \text{ для } j \notin \{j | (pA)_j = 0\},$$

$$(qA)_j = \lambda_1(qB)_j = 0, \text{ для } j \in \{j | (pA)_j = 0\},$$

которые означают, что существует $\Delta > 0$ такой, что $qA \geq (\lambda_1 + \Delta)B$, т.е. вектор $(q, 0)$ допустим для задачи (2.17) при $\lambda = \lambda_1 + \Delta$. Следовательно, $u(\lambda_1 + \Delta) = v(\lambda_1 + \Delta) \geq 0$, что противоречит свойствам функции $u(\lambda)$ и максимальности числа λ_1 . Тем самым показано, что числу λ_1 отвечает невырожденное положение равновесия $(\lambda_1^{-1}, x^*, p^*)$, и теорема доказана.

Приведенное доказательство является конструктивным, поскольку оно указывает способ нахождения числа λ_1 , определяющего темп роста экономики в модели Неймана.

Отметим связь между состояниями равновесия модели Неймана и решениями задач (2.16) и (2.17).

Теорема. Тройка (α, x, p) является положением равновесия (быть может, вырожденного) в модели Неймана тогда и только тогда, когда $u(\alpha^{-1}) = 0$, а пары $(x, 0)$ и $(p, 0)$ являются, соответственно, решениями пары двойственных задач (2.16) и (2.17) при $\lambda = \alpha^{-1}$.

Определение. Число α называется темпом роста.

Определение. Число λ_0 называется **числом Неймана**, число λ_1 – **числом Фробениуса** модели Неймана (A, B) .

Числа $\alpha = \lambda_0^{-1}$ и $\beta = \lambda_1^{-1}$ определяют, соответственно, максимально и минимально возможные темпы роста по стационарной траектории.

Отметим имеющуюся аналогию между собственными числами матрицы Леонтьева и темпами роста модели Неймана с точки зрения продуктивности.

Определение. Модель Неймана (A, B) **продуктивна**, если система $Bx - Ax \geq c$, $x \geq 0$, имеет решение при любом $c \geq 0$.

Теорема. Модель Неймана продуктивна тогда и только тогда, когда ее число Фробениуса меньше 1.

10. Модель парной линейной регрессии. Метод наименьших квадратов. Теорема Гаусса- Маркова.

Модель парной регрессии

Рассмотрим модели регрессии. В них присутствует объясняемая (зависимая) переменная y и набор объясняющих (независимых) переменных $x_1, x_2, \dots, x_n$. Связь между ними выражается следующим образом: $y = f(x_1, x_2, \dots, x_n, \beta_1, \beta_2, \dots, \beta_m)$, где f — некоторая функция, а $\beta_1, \beta_2, \dots, \beta_m$ — набор параметров. В зависимости от вида функции $f(x, \beta)$ модели делятся на линейные и нелинейные.

Рассмотрим набор двух переменных $X_t, Y_t, t = \overline{1, n}$. Можно отображать (X_t, Y_t) точками на плоскости. **Задача:** подобрать $Y = f(X)$ из параметрического семейства $f(X, \beta)$ наилучшим образом описывающую зависимость Y от X .

Подобрать функцию означает выбрать β . Примером параметрического семейства является модель парной линейной регрессии: $f(X, \beta) = a + bX$.

Определим следующие меры отклонения полученных значений от реальных значений:

1. $F = \sum_{t=1}^n (Y_t - f(X_t, \beta))^2$ — сумма квадратов отклонений.
2. $F = \sum_{t=1}^n |Y_t - f(X_t, \beta)|$ — сумма модулей отклонений.
3. $F = \sum_{t=1}^n g(Y_t - f(X_t, \beta))$, где g — мера с которой отклонение входит в F .

Метод наименьших квадратов (МНК).

Пусть $f(x, \beta)$ — линейная функция с параметрами α и β : $Y = f(X, a, b) = a + bX$. Метод наименьших квадратов позволяет подобрать наилучшие параметры путем решения задачи оптимизации:

$$F = \sum_{t=1}^n (Y_t - f(X_t, \beta))^2 \rightarrow \min$$

Записав необходимые условия экстремума (производная равна 0), и решив систему уравнений, получаем оптимальные параметры:

$$\hat{a} = \frac{1}{n} \sum_{t=1}^n Y_t, \quad \hat{b} = \frac{n \sum X_t Y_t - \sum X_t \sum Y_t}{n \sum X_t^2 - (\sum X_t)^2} = \frac{Cov(X, Y)}{Var(X)}$$

Линейная регрессионная модель с двумя переменными.

Добавим к постановке задачи статистические свойства данных. Запишем уравнение зависимости Y_t от X_t , добавив в него случайную величину ε_t . Получим регрессионное уравнение:

$$Y_t = a + bX_t + \varepsilon_t, t = \overline{1, n}.$$

В этом уравнении X_t — неслучайная детерминированная величина, называется регрессор. Y_t, ε_t — случайные величины.

Случайная ошибка может возникать:

- Если модель является упрощением действительности, и в ней опускаются некоторые параметры.
- Если возможны неточности в измерении данных, используемых в модели.

Можно перечислить основные гипотезы линейной регрессионной модели с двумя переменными:

Основные гипотезы:

1. $Y_t = a + bX_t + \varepsilon_t, t = 1, \dots, n$, - спецификация модели.
2. X_t - детерминированная величина; вектор $(X_1, \dots, X_n)'$ не коллинеарен вектору $i = (1, \dots, 1)'$.
3. (a) $E\varepsilon_t = 0, E(\varepsilon_t)^2 = V(\varepsilon_t) = \sigma^2$ - не зависит от t ($E\varepsilon = 0, V(\varepsilon) = \sigma^2 I_n$)
 (b) $E(\varepsilon_t \varepsilon_s) = 0$ при $t \neq s$, некоррелированность ошибок для разных наблюдений.
 Часто добавляется условие:
 (c) Ошибка $\varepsilon_t, t = 1, \dots, n$, имеют совместное нормальное распределение:
 $\varepsilon_t \sim N(0, \sigma^2)$

Условие **3(a)** – независимость дисперсии ошибок от номера наблюдения называется **гомоскедастичностью**, случай когда это условие не выполняется называют **гетероскедастичностью**.

Условие **3(b)** указывает на некоррелированность ошибок для разных наблюдений. Оно часто нарушается, когда имеем дело с временными рядами. В случае, когда это условие не выполняется, говорят об **автокорреляции ошибок**.

Теорема Гаусса-Маркова

В модели линейной регрессии с двумя параметрами оценки параметров $\hat{a}$ и $\hat{b}$, полученные методом наименьших квадратов, имеют наименьшую дисперсию в классе всех линейных несмещенных оценок.

Исходя из теоремы, полученные МНК оценки параметров наилучшие для модели (в смысле наименьшей дисперсии). Остается оценить оставшийся параметр σ^2 .

Оценка дисперсии ошибок.

Найдем значение регрессионной функции с посчитанными коэффициентами: $\hat{Y}_t = \hat{a} + \hat{b}X_t$. Тогда, остатком регрессии будет называться следующая случайная величина: $e_t = Y_t - \hat{Y}_t$.

Между оценкой σ^2 и суммой квадратов ошибки регрессии e_t существует связь, называемая несмещенной оценкой дисперсии ошибок:

$$s^2 = \widehat{\sigma^2} = \frac{1}{n-2} \sum_{t=1}^n e_t^2$$

Если дисперсия ошибок σ^2 известна, то значения коэффициентов $\hat{a}$ и $\hat{b}$ можно найти по формулам из теоремы Гаусса-Маркова.

На практике, как правило, σ^2 неизвестна, и, заменив σ^2 на s^2 , можем получить оценки дисперсий коэффициентов $\hat{a}$ и $\hat{b}$:

$$\widehat{Var}(\hat{b}) = s^2 \frac{1}{\sum (X_t - \bar{X})^2} \quad \widehat{Var}(\hat{a}) = s^2 \frac{\sum X_t^2}{\sum (X_t - \bar{X})^2}$$

Эти формулы используются для нахождения стандартных отклонений оценок коэффициентов регрессии: $s_{\hat{b}} = \sqrt{\widehat{Var}(\hat{b})}$, $s_{\hat{a}} = \sqrt{\widehat{Var}(\hat{a})}$

11. Модель множественной регрессии. Метод наименьших квадратов. Теорема Гаусса-Маркова.

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t, \quad t = 1, \dots, n$$

где x_{tp} — значения регрессора x_p в наблюдении t , а $x_{t1} = 1, t = 1, \dots, n$.

1. $y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t, \quad t = 1, \dots, n$ — спецификация модели.
2. $x_{t1}, \dots, x_{tk}$ — детерминированные величины. Векторы $x_s = (x_{1s}, \dots, x_{ns})', s = 1, \dots, k$ линейно независимы в R^n .
3. а. $E\varepsilon_t = 0, E(\varepsilon_t^2) = V(\varepsilon_t) = \sigma^2$ — не зависит от t .
б. $E(\varepsilon_t \varepsilon_s) = 0$ при $t \neq s$ — статистическая независимость (некоррелированность) ошибок для разных наблюдений. Часто добавляется следующее условие.
с. Ошибки $\varepsilon_t, t = 1, \dots, n$ имеют совместное нормальное распределение: $\varepsilon_t \sim N(0, \sigma^2)$.

В этом случае модель называется *нормальной линейной регрессионной*.

Метод наименьших квадратов

Целью метода является выбор вектора оценок $\hat{\beta}$, минимизирующего сумму квадратов остатков e_t (необъясненную часть дисперсии).

$$e = y - \hat{y} = y - X\hat{\beta},$$
$$ESS = \sum e_t^2 = e'e \rightarrow \min_{\hat{\beta}}$$

Выразим $e'e$ через X и β :

$$e'e = y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}.$$

Необходимые условия минимума ESS получаются дифференцированием по вектору $\hat{\beta}$

$$\frac{\partial ESS}{\partial \hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0,$$

откуда, учитывая обратимость матрицы $X'X$ находим оценку метода наименьших квадратов:

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y.$$

Теорема Гаусса-Маркова

Для модели 1-Заб оценка $\hat{\beta}$, полученная с помощью метода наименьших квадратов, является лучшей линейной несмещенной оценкой.

Доказательство:

1) $\hat{\beta}'$ — несмещенная оценка β

$$E\hat{\beta} = E[(X'X)^{-1}X'y] = (X'X)^{-1}X'Ey = (X'X)^{-1}X'E(X\beta + \varepsilon) = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'E\varepsilon = \beta.$$

Пусть $A = (X'X)^{-1}X'$

Тогда $\widehat{\beta} = Ay$

Пусть b — любая другая несмещенная оценка β . Тогда можно представить b в виде

$$b = (A + C)y, C \text{ } k \times n$$

Так как b — несмещенная оценка β , то $\beta = Eb = (A + C)Ey = (A + C)E(X\beta + \varepsilon) = (A + C)X\beta + (A + C)E\varepsilon = (A + C)X\beta = (AX + CX)\beta = (I + CX)\beta \Rightarrow CX = 0$.

2) Вычислим матрицу ковариаций $\widehat{\beta}$:

$$V(\widehat{\beta}) = V(Ay) = AV(y)A' = A\sigma^2 I A' = \sigma^2 (X'X)^{-1} X'X (X'X)^{-1} = \sigma^2 (X'X)^{-1}$$

$$3) V(b) = V(\widehat{\beta}) + \sigma^2 CC', CC' \geq 0.$$

$$\Rightarrow V(b) \geq V(\widehat{\beta}).$$

12. Понятие временного ряда. Понятие строго стационарного временного ряда. Условия стационарности временного ряда в широком смысле.

Под временным рядом понимается последовательность наблюдений значений некоторой переменной, произведенных через равные промежутки времени. Если принять длину такого промежутка за единицу времени (год, квартал, день и т.п), то можно считать, что последовательные наблюдения $x_1, \dots, x_n$ произведены в моменты $t = 1, \dots, n$.

Ряд $x_t, t = 1, \dots, n$ называется **строго стационарным** (или **стационарным в узком смысле**), если для любого m ($m < n$) совместное распределение вероятностей случайных величин $X_{t_1}, \dots, X_{t_m}$, такое же как и для $X_{t_1+\tau}, \dots, X_{t_m+\tau}$, при любых $t_1, \dots, t_m$ и τ , таких, что $1 \leq t_1, \dots, t_m \leq n$ и $1 \leq t_1 + \tau, \dots, t_m + \tau \leq n$, т.е функция плотности при сдвиге во времени не изменится

$$P\{x_{t_1}, \dots, x_{t_m}\} = P\{x_{t_1+\tau}, \dots, x_{t_m+\tau}\}.$$

- $E(X_t) \equiv \mu$
- $D(X_t) \equiv \sigma^2$
- $Cov(X_t, X_{t+\tau}) = \gamma(\tau)$ для любых t и τ .

Ряд, для которого выполнены указанные три условия, называют **стационарным в широком смысле** (**слабо стационарным**, **стационарным второго порядка** или **ковариационно стационарным**).

Если ряд является стационарным в широком смысле, то он не обязательно является строго стационарным. В то же время, и строго стационарный ряд может не быть стационарным в широком смысле просто потому, что у него могут не существовать математическое ожидание и/или дисперсия. (В отношении последнего примером может служить случайная выборка из распределения Коши.) Кроме того, возможны ситуации, когда указанные три условия выполняются, но, например $E(X_t^3)$ зависит от t .

13. Нестационарный процесс авторегрессии – интегрированного скользящего среднего $ARIMA(p,d,q)$.

Рассмотрим нестационарный ряд случайного блуждания. Его уравнение имеет вид: $x_t = x_{t-1} + \varepsilon_t$. Если взять первую разность, равную $\Delta x_t = x_t - x_{t-1} = (1 - L)x_t$, то уравнение сведется к следующему: $\Delta x_t = y_t = \varepsilon_t$. То есть в первых разностях ряд станет стационарным.

Рассмотрим ряд вида: $x_t = \alpha + \beta t + \gamma t^2 + \varepsilon_t$.

Его плотностью детерминированная часть (тренд) является параболической функцией времени. Очевидно, что этот ряд нестационарный. Математическое ожидание этого процесса зависит от времени.

Проверим: Приводится ли такой ряд взятием последовательных разностей к стационарному?

Если мы возьмем первую последовательную разность, то получим: $\Delta x_t = \alpha + \beta t + \gamma t^2 + \varepsilon_t - \alpha - \beta(t-1) - \gamma(t-1)^2 \varepsilon_{t-1} = \beta + 2\gamma t - \gamma + (\varepsilon_t - \varepsilon_{t-1})$.

Степень полинома, описывающего тренд, понизилась на единицу. Если провести взятие второй разности, то останется $\Delta^2 x_t = \Delta x_t - \Delta x_{t-1} = 2\gamma + (\varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2})$, то есть получим стационарный процесс.

НЕОБЯЗАТЕЛЬНЫЙ АБЗАЦ(Правда, видно, как в уравнение начинает "проникать" скользящее среднее. Полученный двукратным взятием разностей стационарный процесс является процессом $MA(2)$. Но, по крайней мере, взятием последовательных разностей исходный ряд с квадратичным трендом приводится к стационарному виду..)

Бокс и Дженкинс на основании этого свойства предложили выделить класс нестационарных рядов, которые взятием последовательных разностей можно привести к стационарному виду, а именно к виду $ARMA$.

Если ряд после взятия d последовательных разностей приводится к стационарному, то этот ряд носит название $ARIMA(p, d, q)$.

$ARIMA$ - процесс авторегрессии - интегрированного скользящего среднего. При этом p - параметр AR -части, d - степень интеграции, и q - это параметр MA -части.

В операторном виде процесс $ARIMA(p, d, q)$ записывается как: $\alpha_p(L)\Delta^d x_t = \beta_q(L)\varepsilon_t$.

Или по-другому $\underbrace{\alpha_p(L)(1-L)^d}_{p+d} x_t = \beta_q(L)\varepsilon_t$.

Этот процесс x_t нестационарный, потому что здесь не выполняется условие, что все корни характеристического уравнения по модулю меньше единицы.

Но, если обозначить $(1-L)^d x_t = y_t$, то y_t - это стационарный процесс.

14. Подход Бокса-Дженкинса построения модели типа $ARIMA(p,d,q)$ по реализации временного ряда.

Бокс и Дженкинс предложили следующий подход, состоящий из нескольких этапов, к выбору модели типа $ARIMA$ по наблюдаемой реализации временного ряда.

Для начала вспомним модель типа $ARIMA$:

$ARIMA$ – процесс авторегрессии интегрированного скользящего среднего. При этом p – параметр AR -части, d – степень интеграции и q – параметр MA -части.

В операторном виде процесс $ARIMA(p, d, q)$ записывается как: $\alpha_p(L)\Delta^d x_t = \beta_q(L)\varepsilon_t$.

Или по-другому $\underbrace{\alpha_p(L)(1-L)^d}_{p+d} x_t = \beta_q(L)\varepsilon_t$.

А этапы следующие:

Этап I. Идентификация модели

1. Установить порядок интеграции d , то есть добиться стационарности ряда, взяв достаточное количество последовательных разностей. Другими словами, "остационарить" ряд. Используются все описанные ранее средства визуальный анализ коррелограммы визуальный анализ автокорреляционной и частной автокорреляционной функций, тест на единичный корень. $d \leq 2$ – обычно.
2. После этого мы получаем временной ряд y_t , которому нужно подобрать уже $ARMA(p, q)$. Исходя из поведения автокорреляционной и частной автокорреляционной функций, установить параметры p и q . Рекомендуется, чтобы $p + q \leq 3$.

Этап II. Оценивание модели

Оценивание коэффициентов $\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q$ при условии, что мы уже знаем p и q .

Этап III. Диагностика модели

Стандартная для эконометрического подхода процедура. По остаткам осуществляется тестирование или диагностика построенной модели.

Этап IV. Использование модели

Использование модели, в основном, для прогнозирования будущих значений временного ряда.

Бокс и Дженкинс применили этот подход ко многим временным рядам, которые были известны в то время, как к финансовым, так и к макроэкономическим.

Бокс и Дженкинс смогли построить модели типа $ARIMA$ для всех

исследуемых рядов и установили, в частности, что практически все экономические процессы описываются моделями с относительно небольшими величинами параметров p и q , а параметр d обычно не превышает 2. К тому же оказалось, что точность прогнозирования по моделям *ARIMA* оказалась выше, чем давали в то время экономические модели.

Если исследуемый ряд *нестационарный*, то его автокорреляционная функция не будет убывать.

Если ряд *стационарен*, то, начиная с какого-то номера, теоретические автокорреляции будут убывать. Поэтому можно рассчитать их оценки – выборочные автокорреляции, и посмотреть, убывают они или нет. Если ряд окажется стационарным, перейти к определению параметров p и q . Если нет, то надо построить ряд первых разностей и проверить на стационарность его.

Рассмотрим, **например**, процесс $x_t = \alpha x_{t-1} + \epsilon_t, \epsilon_t \sim N(0, \sigma^2)$.

При $\alpha > 1$ взятие первой разности не поможет сделать ряд стационарным. Это нестационарный ряд взрывного типа, оценки его "автокорреляционной" функции растут с увеличением сдвига во времени.

Если же $\alpha = 1$, то ряд представляет собой случайное блуждание и после взятия первой разности он станет стационарным.

Переходим к оцениванию по выборке статистических характеристик исследуемого процесса в предположении его эргодичности. В качестве оценки математического ожидания применяется статистика $\bar{x} = \frac{1}{T} \sum_{t=1}^T x_t$, то есть обычное среднее по выборке.

В качестве оценки дисперсии процесса обычно принимается следующая величина

$$S^2 = \frac{1}{T} \sum_{t=1}^T (x_t - \bar{x})^2$$

Обратите внимание делитель не $T - 1$, а как привычно для обработки независимых наблюдений, а T .

Для оценки коэффициента теоретической автокорреляции используем $\hat{\rho}_k = \frac{1}{TS^2} \sum_{t=k+1}^T (x_t - \bar{x})(x_{t-k} - \bar{x})$.

Это означает, что для расчета выборочных ковариаций используется одинаковый делитель – T , что, в случае независимых наблюдений, дает смещенные оценки соответствующих теоретических ковариаций. Если бы мы требовали несмещенности оценок, то у нас бы были разные делители при различных k . Кроме того, при увеличении k уменьшается объем выборки, пригодный для расчета оценки соответствующего коэффициента корреляции. На практике, по рекомендации, идущей еще от Бокса и Дженкинса, не рассчитываются оценки $\hat{\rho}_k$ для $k > \frac{T}{4}$.

15. Статистические свойства оценок по методу наименьших квадратов параметров парной линейной регрессии.

Рассмотрим задачу “наилучшей” аппроксимации набора наблюдений $X_t, Y_t, t = 1, \dots, n$ линейной функцией $f(X) = a + bX$ в смысле минимизации функционала

$$F = \sum_{t=1}^n (Y_t - (a + bX_t))^2 \rightarrow \min, \forall a, b$$

Запишем необходимые условия экстремума:

$$\frac{dF}{da} = -2 \sum_{t=1}^n (Y_t - a - bX_t) = 0,$$

$$\frac{dF}{db} = -2 \sum_{t=1}^n X_t (Y_t - a - bX_t) = 0,$$

Или

$$\sum_{t=1}^n (Y_t - a - bX_t) = 0,$$

$$\sum_{t=1}^n X_t (Y_t - a - bX_t) = 0.$$

Раскроем скобки и получим *стандартную форму нормальных уравнений* (для краткости опустим индексы суммирования у знака суммы):

$$an + b \sum X_t = \sum Y_t,$$

$$a \sum X_t + b \sum X_t^2 = \sum X_t Y_t.$$

Решением этой системы будут $\hat{a}, \hat{b}$

$$\hat{a} = \frac{1}{n} \sum Y_t - \frac{1}{n} \sum X_t \hat{b} = \bar{Y} - \bar{X} \hat{b},$$

$$\hat{b} = \frac{n \sum X_t Y_t - \sum X_t \sum Y_t}{n \sum X_t^2 - (\sum X_t)^2} = \frac{Cov(X, Y)}{Var X}$$

Замечание.

Из первого уравнения следует

$$\bar{Y} = \hat{a} + \hat{b} \bar{X}$$

То есть уравнение прямой линии $Y = \hat{a} + \hat{b}X$, полученное в результате минимизации функционала, проходит через точку $(\bar{X}, \bar{Y})$. Здесь через $\bar{X}, \bar{Y}$ обозначены выборочные средние значения переменных X_t, Y_t :

$$\bar{X} = \frac{1}{n} \sum X_t, \quad \bar{Y} = \frac{1}{n} \sum Y_t,$$

$$Cov(X, Y) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}), \quad Var(X, Y) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Уравнения в отклонениях

Обозначим через $x_t = X_t - \bar{X}$, $y_t = Y_t - \bar{Y}$ отклонения от средних по выборке значений X_t и Y_t (можно проверить, что $\bar{x} = \bar{y} = 0$).

Решим теперь ту же задачу: подобрать линейную функцию $f(x) = a + bx$ минимизирующую функционал

$$F = \sum_{t=1}^n (y_t - (a + bx_t))^2 \rightarrow \min.$$

Из геометрических соображений ясно, что решением задачи будет та же прямая на плоскости (x, y) , что и для исходных данных X_t и Y_t . В самом деле в силу $\bar{Y} = \hat{a} + \hat{b}\bar{X}$ переход от X, Y к отклонениям x, y означает лишь перенос начала координат в точку $(\bar{X}, \bar{Y})$. Вычисления, которые необходимо проделать для решения задачи, вполне аналогичны предыдущим (заменой X, Y на x, y). Сделав замену X_t, Y_t на x_t, y_t в равенствах для $\hat{a}, \hat{b}$ учитывая, что $\bar{x} = \bar{y} = \frac{1}{n} \sum x_t = \frac{1}{n} \sum y_t = 0$, получим

$$\hat{a} = 0$$

$$\hat{b} = \frac{\sum x_t y_t}{\sum x_t^2} = \frac{\sum (X_t - \bar{X})(Y_t - \bar{Y})}{\sum (X_t - \bar{X})^2}.$$

Таким образом, получили другое выражение для углового коэффициента $\hat{b}$.

16. Проверка гипотезы $H_0 : b = b_0$. Доверительные интервалы для коэффициентов регрессии. Коэффициент детерминации.

Основные гипотезы, лежащие в основе линейной регрессионной модели (возможно из можно не писать в билете):

Основные гипотезы:

1. $Y_t = a + bX_t + \varepsilon_t, t = 1, \dots, n$, - спецификация модели.
2. X_t - детерминированная величина; вектор $(X_1, \dots, X_n)'$ не коллинеарен вектору $i = (1, \dots, 1)'$.
3. (a) $E\varepsilon_t = 0, E(\varepsilon_t)^2 = V(\varepsilon_t) = \sigma^2$ - не зависит от t ($E\varepsilon = 0, V(\varepsilon) = \sigma^2 I_n$)
(b) $E(\varepsilon_t \varepsilon_s) = 0$ при $t \neq s$, некоррелированность ошибок для разных наблюдений.
Часто добавляется условие:
(c) Ошибка $\varepsilon_t, t = 1, \dots, n$, имеют совместное нормальное распределение:
 $\varepsilon_t \sim N(0, \sigma^2)$

Свойства МНК-оценок

МНК-оценки коэффициентов регрессии $\hat{a}, \hat{b}$ имеют совместное нормальное распределение:

$$\hat{a} \sim N\left(a, \sigma^2 \frac{\sum X_t^2}{n \sum x_t^2}\right), \quad \hat{b} \sim N\left(b, \sigma^2 \frac{1}{\sum x_t^2}\right), \text{ где } x_t = X_t - \bar{X}, \bar{X} = \frac{\sum X_t}{n}$$

Отсюда получаем, что $\hat{b} - b \sim N(0, \sigma_b^2)$, где $\sigma_b^2 = \frac{\sigma^2}{\sum x_t^2}$

$$\widehat{Var}(\hat{b}) = s_b^2 = s^2 \frac{1}{\sum (X_t - \bar{X})^2} = \frac{s^2}{\sum x_t^2} \Rightarrow \frac{\hat{b} - b}{s_b} \sim N(0, 1)$$

Учитывая, что распределение оценки дисперсии ошибок $s^2 = \widehat{\sigma^2}$ имеет распределение:

$$\frac{(n-2)s^2}{\sigma^2} \sim \chi^2(n-2), \text{ получим}$$

$$\frac{s}{\sigma} \sim \sqrt{\frac{1}{n-2} \chi^2(n-2)} \Rightarrow \{\text{по определению статистики Стьюдента}\} \Rightarrow t = \frac{(\hat{b} - b)/\sigma_b}{s/\sigma} \sim t(n-2), \text{ где } t(n-2) \text{ — Т-статистика Стьюдента с } (n-2)$$

степенями свободы.

Так как $\frac{\sigma_{\hat{b}}}{\sigma} = \frac{s_{\hat{b}}}{s}$ получим $t = \frac{\hat{b}-b}{s_{\hat{b}}} \sim t(n-2) (*)$

Аналогично показываем, что $t = \frac{\hat{a}-a}{s_{\hat{a}}} \sim t(n-2)$.

Статистику (*) можно использовать для проверки гипотезы $H_0 : b = b_0$ против альтернативной гипотезы $H_1 : b \neq b_0$.

Пусть H_0 верна, тогда $t = \frac{\hat{b} - b}{s_{\hat{b}}} \sim t(n-2)$

Гипотеза H_0 принимается при $|t| \leq t_{\alpha}(n-2)$.

Другими словами:

»

$100\% \cdot (1 - \alpha)$ -ый доверительный интервал для b равен:

$[\hat{b} - t_e s_{\hat{b}}, \hat{b} + t_e s_{\hat{b}}]$, где t_e — точка t -распределения с $(n-2)$ степенями свободы при уровне значимости α .

(Хз надо ли это писать в билете) Если гипотеза нормальности ошибок не выполняется — **3(с)**, то при некоторых условиях регулярности на поведение X_t при росте n оценки $\hat{a}, \hat{b}$ имеют асимптотически нормальное распределение, то есть

$$\hat{a} \rightarrow N\left(a, \sigma^2 \frac{\sum X_t^2}{n \sum x_t^2}\right), \quad \hat{b} \rightarrow N\left(b, \sigma^2 \frac{1}{\sum x_t^2}\right)$$

Коэффициент детерминации R^2

Рассмотрим вариацию значения Y_t вокруг среднего значения $\bar{Y}$: $\sum (Y_t - \bar{Y})^2$.

Разобьем вариацию на две части: объяснению и необъяснённую.

Пусть $\hat{Y}_t = \hat{a} + \hat{b}X_t$ — предсказанные значения Y_t .

Тогда $Y_t - \bar{Y} = (Y_t - \hat{Y}_t) + (\hat{Y}_t - \bar{Y})$ и вариация Y_t записывается в виде:

$$\sum (Y_t - \bar{Y})^2 = \sum (Y_t - \hat{Y}_t)^2 + \sum (\hat{Y}_t - \bar{Y})^2 + 2 \sum (Y_t - \hat{Y}_t)(\hat{Y}_t - \bar{Y}) \sim$$

$$TSS = ESS + RSS + 2 \sum (Y_t - \hat{Y}_t)(\hat{Y}_t - \bar{Y})$$

TSS – вся дисперсия. ESS – необъясненная часть дисперсии. RSS – объясненная часть дисперсии.

Третье слагаемое равно 0, так как

$$\sum e_t(\hat{Y}_t - \bar{Y}) = \sum e_t(\hat{a} + \hat{b}X_t - \bar{Y}) = \{x_t = X_t - \bar{X}\} = (\hat{a} + \hat{b}\bar{X} - \bar{Y}) \sum e_t + \hat{b} \sum e_t x_t = 0$$

(вектор остатков регрессии e_t ортогонален

константе i - единичному вектору и вектору x)

Коэффициентом детерминации называется $R^2 = 1 - \frac{ESS}{TSS} = \frac{RSS}{TSS}$.

$R^2 \in [0, 1]$, если $R^2 = 0$, то регрессия ничего не дает, $R^2 = 1$ — точная подгонка.

17. Статистические свойства оценок по методу наименьших квадратов параметров множественной регрессии. Коэффициент детерминации и скорректированный коэффициент детерминации.

Модель множественной регрессии (многомерная регрессивная модель)

$$y_t = \beta_1 + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t, \quad t = \overline{1, n} \quad \text{или} \\ y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t, \quad t = \overline{1, n},$$

где x_{tp} — значение регрессора x_p в наблюдении t , а $x_{t1} = 1, t = \overline{1, n}$. С учетом этого замечания различаются модели со свободным членом и без свободного члена.

Если выполнены следующие гипотезы, то модель называется **нормальной линейной регрессионной**:

1. $y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t, \quad \overline{1, n}$ — спецификация модели;
2. $x_{t1}, \dots, x_{tk}$ — детерминированные величины. Вектор $x_s = (x_{1s}, \dots, x_{ns})^T, s = \overline{1, k}$ линейно независимые в R^n ;
3. $\mathbb{E}\varepsilon_t = 0, \mathbb{E}(\varepsilon_t^2) = \text{Var}(\varepsilon_t) = \sigma^2$ — не зависит от t ;
4. $\mathbb{E}(\varepsilon_t \varepsilon_s) = 0$ при $t \neq s$ — статистическая независимость (некоррелированность) ошибок для разных наблюдений;
5. Ошибка $\varepsilon_t, t = \overline{1, n}$ имеют совместное нормальное распределение: $\varepsilon_t \sim N(0, \sigma^2)$.

Матричный вид гипотез

Пусть y — вектор-столбец $(y_1, \dots, y_n)^T, \beta = (\beta_1, \dots, \beta_k)^T$ — вектор коэффициентов; $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ — вектор ошибок.

$$X = \begin{bmatrix} x_{11} & \dots & x_{1k} \\ \dots & \dots & \dots \\ x_{n1} & \dots & x_{nk} \end{bmatrix} - n \times k \text{ матрица объясняющих переменных.}$$

Столбцами матрицы X являются $n \times 1$ векторы регрессоров $x_s = (x_{1s}, \dots, x_{ns})^T, s = \overline{1, k}$

1. $y = X\beta + \varepsilon$ — спецификация модели;

2. X — детерминированная матрица, имеет максимальный ранг k ;
3. $\mathbb{E}(\varepsilon) = 0, \text{Var}(\varepsilon) = \varepsilon\varepsilon^T = \sigma^2 I_n$;
4. $\varepsilon \sim N(0, \sigma I_n)$, то есть ε — нормально распределенный случайный вектор со средним 0 и матрицей ковариация $\sigma^2 I_n$ (нормальная линейная регрессионная модель).

Теорема Гаусса-Маркова для множественной регрессии

Предположим, что выполнены гипотезы 1-3.

Тогда оценка методом наименьших квадратов $\hat{\beta}_{OLS} = (X^T X)^{-1} X^T y$ является наиболее эффективной (в смысле наименьшей дисперсии) оценкой в классе линейных (по y) несмещённых оценок (Best Linear Unbiased Estimator, BLUE).

Статистические свойства МНК оценок

Введем вектор прогнозируемых значений $\hat{y} = X\hat{\beta} = X(X^T X)^{-1} X^T y$

Введем вектор остатков регрессии:

$$e = y - \hat{y} = y - X\hat{\beta} = (I - X(X^T X)^{-1} X^T)y = My, \text{ где } M = (I - X(X^T X)^{-1} X^T) = (I - N).$$

Вычислим мат. ожидание и матрицу ковар. e :

$$\mathbb{E}(e) = (I - X(X^T X)^{-1} X^T)\mathbb{E}(y) = (I - X(X^T X)^{-1} X^T)X\beta = X\beta - X\beta = 0.$$

$$\text{Var}(e) = \text{Var}(My) = M\text{Var}(y)M^T = M\sigma^2 I M^T = \sigma^2 M$$

Оценка дисперсии ошибок:

$$\mathbb{E}(e^T e) = \text{tr}(\text{Var}(e)) = \sigma^2 \text{tr}(I_n - N) = (n - k)\sigma^2 \text{ (tr — след матрицы)}.$$

Следовательно $s^2 = \hat{\sigma}^2 = \frac{e^T e}{n-k} = \frac{\sum e_i^2}{n-k}$ — несмещенная оценка дисперсии ошибок, то есть $\mathbb{E}s^2 = \sigma^2$.

Распределение:

$$\frac{e^T e}{\sigma^2} \sim \chi^2(n - k) \text{ или } (n - k) \frac{s^2}{\sigma^2} \sim \chi^2(n - k)$$

Оценки $\hat{\beta}$ и s^2 независимы в предположении нормальной линейной множественной регрессионной модели

$$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T (X\beta + \varepsilon) = \beta + (X^T X)^{-1} X^T \varepsilon = \beta + A\varepsilon$$

Вектор $\hat{\beta}$ и e имеют совместное многомерное нормальное распределение
 $\Rightarrow$ нужно доказать их некоррелированность

$$AM = (X^T X)^{-1} X^T (I - X(X^T X)^{-1} X^T) = 0$$

Тогда (так как $\mathbb{E}e = 0$): $Cov(\hat{\beta}, e) = \mathbb{E}((\hat{\beta} - \beta)e^T) = \mathbb{E}(A\epsilon\epsilon^T M) = \sigma^2 AM = 0$

Так как s^2 является функцией $e \Rightarrow$ оценки $\hat{\beta}$ и s^2 независимы.

Коэффициент детерминации R^2 и скорректированный коэффициент детерминации R_{adj}^2

Разобьем вариацию $\sum (y_t - \bar{y})^2$ на две части: объясняемую и не объясняемую.

$$\sum (y_t - \bar{y})^2 = \sum (y_t - \hat{y}_t)^2 + \sum (\hat{y}_t - \bar{y})^2 + 2 \sum (y_t - \hat{y}_t)(\hat{y}_t - \bar{y})$$

То же самое в векторной форме:

$$(y - \bar{y}i)^T (y - \bar{y}i) = (y - \hat{y})^T (y - \hat{y}) + (\hat{y} - \bar{y}i)^T (\hat{y} - \bar{y}i) + 2(y - \hat{y})^T (y - \bar{y}i), \text{ где } i \text{ — единичный вектор.}$$

Третье слагаемое равно нулю:

$$(y - \hat{y})^T (y - \bar{y}i) = e^T (X\hat{\beta} - \bar{y}i) = e^T X\hat{\beta} - \bar{y}e^T i = 0, \text{ так как } e^T X = 0, e^T i/n = 0.$$

$$\text{Поэтому: } \|y - \bar{y}i\|^2 = \|y - \hat{y}\|^2 + \|\hat{y} - \bar{y}i\|^2 (TSS + ESS + RSS)(*)$$

TSS – вся дисперсия. ESS – необъясненная часть дисперсии. RSS – объясненная часть дисперсии.

$$\text{Положим: } y_* = y - \bar{y}i, \quad \hat{y}_* = \hat{y} - \bar{y}i \Rightarrow y_*^T y_* = e^T e + \hat{y}_*^T \hat{y}_*$$

$$\text{Коэффициентом детерминации называют } R^2 = 1 - \frac{ESS}{TSS} = \frac{RSS}{TSS} = 1 - \frac{e^T e}{y_*^T y_*} = \frac{\hat{y}_*^T \hat{y}_*}{y_*^T y_*}, \quad R^2 \in [0, 1].$$

R^2 корректно определен, только если вектор $i = (1, \dots, 1)^T$ принадлежит линейной оболочке векторов $x_1, \dots, x_k$.

Свойства R^2 :

1. R^2 возрастает при добавлении еще одного регрессора.
2. R^2 изменяется даже при простейшем преобразовании зависимой переменной, поэтому сравнивать по значению R^2 можно только регрессии с одинаковыми зависимыми переменными.

Попыткой устранить эффект, связанный с ростом R^2 при возрастании числа регрессоров, является коррекция R^2 на число регрессоров.

Скорректированным (adjusted) R^2 называется

$$R_{adj}^2 = 1 - \frac{e^T e / (n - k)}{y_*^T y_* / (n - 1)}$$

Заметим, что нет никакого существенного оправдания такого способа коррекции (то есть можно по другому корректировать, главное устранять недостаток возрастания при увеличении числа регрессоров).

Свойства скорректированного R^2

1. $R_{adj}^2 = 1 - (1 - R^2) \frac{n-1}{n-k}$
2. $R^2 \geq R_{adj}^2, k > 1$.
3. $R_{adj}^2 \leq 1$, но может принимать значения < 0 .

18. Проверка гипотезы о линейном ограничении общего вида $H_0 : H\beta = r$.

Линейное ограничение общего вида $H_0 : H\beta = r$. Пусть H — $q \times k$ матрица, β — $k \times 1$ вектор коэффициентов, r — $q \times 1$ вектор.

Число ограничений не превосходит числа параметров и ограничения линейно независимы, т.е. $q \leq k$ и матрица H имеет максимальный ранг: $\text{rank}(H) = q$.

Пример. В качестве примера рассмотрим следующие матрицы H, r для $k = 3, q = 2$:

$$H\beta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} = r.$$

Это условие соответствует системе двух линейных ограничений

$$\begin{cases} \beta_1 = 2 \\ \beta_2 - \beta_3 = 0 \end{cases}$$

$$F = \frac{(H\hat{\beta} - r)'(H(X'X)^{-1}H')^{-1}(H\hat{\beta} - r)/q}{e'e/(n-k)} \sim F(q, n-k).$$

Если справедлива гипотеза $H_0 : H\beta - r = 0$, то статистика F не должна принимать слишком больших значений, а именно, с 95%-вероятностью $F < F_c(q, n-k)$, где $F_c(q, n-k)$ есть 95%-квантиль распределения Фишера $F(q, n-k)$.

Другой вид F -статистики:

$$F = \frac{(\hat{\beta} - \beta)'H'(H(X'X)^{-1}H')^{-1}H(\hat{\beta} - \beta)/q}{e'e/(n-k)} \sim F(q, n-k)$$

Условие $F < F_c(q, n-k)$ задает 95%-доверительную область для коэффициентов β (выпуклую).

В случае $H = I$ статистика F выглядит следующим образом:

$$F = \frac{(\hat{\beta} - \beta)'(X'X)(\hat{\beta} - \beta)/k}{e'e/(n-k)} \sim F(k, n-k).$$

$H_0 : \beta_{k-q+1} = \beta_{k-q+2} = \dots = \beta_k = 0$. Гипотеза является, конечно, частным случаем общей линейной гипотезы $H\beta = r$.

$$F = \frac{(e'e - e'e)/q}{e'e/(n-k)} = \frac{(\text{ESS}_R - \text{ESS}_{UR})/q}{\text{ESS}_{UR}/(n-k)} \sim F(q, n-k).$$

Здесь ESS_R — сумма квадратов остатков "короткой" (*restricted*) регрессии; ESS_{UR} — сумма квадратов остатков "длинной" (*unrestricted*) регрессии.

F -статистику можно выразить через коэффициенты детерминации R^2 для "короткой" и "длинной" регрессий:

$$F = \frac{(R_{UR}^2 - R_R^2)/q}{(1 - R_{UR}^2)/(n-k)} \sim F(q, n-k)$$

При $F < F_c(q, n-k) \Rightarrow H_0$ о "короткой" регрессии принимается, т.е. использование q дополнительных регрессоров ничего не добавляет к качеству оценивания модели.

19. Модели процессов авторегрессии и скользящего среднего: AR(p), MA(q) и ARMA(p,q). Условия стационарности этих процессов.

Процесс авторегрессии порядка p (AR(p)) определяется соотношениями

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + \varepsilon_t, \quad a_p \neq 0,$$

где ε_t - процесс белого шума с $D(\varepsilon_t) = \sigma_\varepsilon^2$.

$Cov(X_{t-s}, \varepsilon_t) > 0$ для всех $s > 0$.

При рассмотрении процессов авторегрессии и некоторых других моделей удобно использовать **оператор запаздывания L (lag operator)**, который воздействует на временной ряд и определяется соотношением

$$LX_t = X_{t-1}$$

Если оператор запаздывания применяется k раз, что обозначается как L^k , то это дает в результате

$$L^k X_t = X_{t-k}$$

Тогда процесс авторегрессии p -го порядка перепишется в виде:

$$a(L)X_t = \varepsilon_t$$

где

$$a(L) = 1 - (a_1 L + a_2 L^2 + \dots + a_p L^p).$$

Для того, чтобы такой процесс был стационарным, все корни алгебраического уравнения $a(z) = 0$ должны лежать вне единичного круга, т.е. $|z| > 1$.

Процесс скользящего среднего порядка q ($MA(q)$) определяется соотношениями:

$$X_t = \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \dots + b_q\varepsilon_{t-q} + \mu, \quad b_q \neq 0,$$

Используя оператор запаздывания L , $MA(q)$ можно переписать в следующем виде:

$$X_t - \mu = b(L)\varepsilon_t,$$

где $b(L) = 1 + b_1L + b_2L^2 + \dots + b_qL^q$.

Всякий стационарный процесс $AR(p)$ эквивалентен $MA(\infty)$.

Если корни $b(z) = 0$ лежат вне единичного круга, то ряд обратим, т.е. $MA(q)$ эквивалентен $AR(\infty)$.

Смешанный процесс авторегрессии – скользящего среднего (процесс авторегрессии с остатками в виде скользящего среднего) (ARMA(p,q)) определяется соотношениями:

$$X_t = a_1X_{t-1} + a_2X_{t-2} + \dots + a_pX_{t-p} + \varepsilon_t + b_1\varepsilon_{t-1} + \dots + b_q\varepsilon_{t-q}, \quad a_p \neq 0, \quad b_q \neq 0, \quad \varepsilon_t - \text{белый шум}$$

В операторной форме это соотношение имеет вид

$$\begin{aligned} a(L) &= 1 - (a_1L + a_2L^2 + \dots + a_pL^p) \\ b(L) &= 1 + b_1L + b_2L^2 + \dots + b_qL^q \\ a(L)X_t &= b(L)\varepsilon_t, \end{aligned}$$

т.е. $a(L)$ и $b(L)$ имеют тот же вид, что и в определенных ранее моделях $AR(p)$ и $MA(q)$.

Процесс стационарен, если все корни уравнения $a(z) = 0$ лежат вне единичного круга $|z| \leq 1$

Если все корни уравнения $b(z) = 0$ лежат вне единичного круга $|z| \leq 1$ (**условие обратимости**), то существует эквивалентное представление процесса X_t в виде процесса авторегрессии бесконечного порядка $AR(\infty)$.

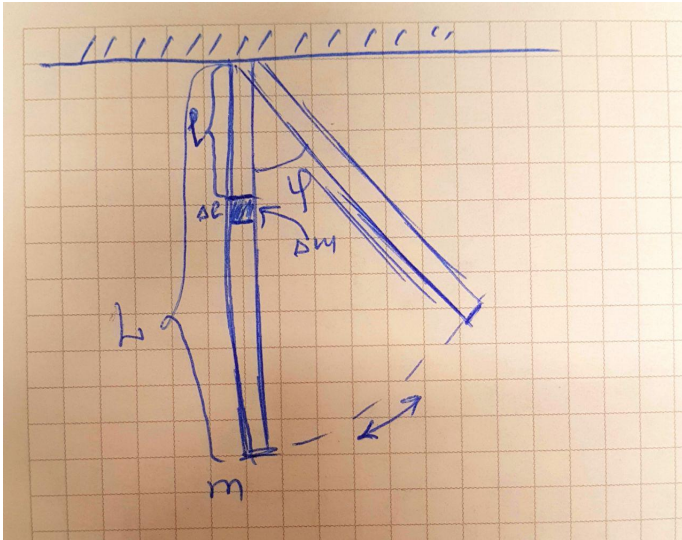
Билет 20

Напишите уравнение Лагранжа для массивного стержня длины L и массы m с закрепленным концом, колеблющегося в вертикальной плоскости, и оцените частоту его колебаний.

Уравнение Лагранжа:

$$\frac{\partial L}{\partial q_i} = \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i}, \forall i = 1, \dots, N$$

где q_i – обобщенные координаты точек тела. Соответственно $L(q, \dot{q}, t)$ – ф-я Лагранжа.



Разобьем стержень на малые части массой Δm длиной Δl , $\Delta m = \rho S \Delta l$. Тогда каждую часть можно представить как математический маятник, и для него:

$$U = \Delta mgh = \Delta mgl(1 - \cos\varphi)$$

И соответственно ф-я Лагранжа:

$$\Delta L = \frac{\Delta m(l\dot{\varphi})^2}{2} - \Delta mgl(1 - \cos\varphi) = \rho S \Delta l \left(\frac{(l\dot{\varphi})^2}{2} - mgl(1 - \cos\varphi) \right)$$

Чтобы получить ф-ю Лагранжа для целого стержня, надо просуммировать ΔL по всем частям (т.е. проинтегрировать):

$$L = \int_0^l \rho S \left(\frac{(l\dot{\varphi})^2}{2} - mgl(1 - \cos\varphi) \right) dl = \rho S \left(\frac{L^3 \dot{\varphi}^2}{3 \cdot 2} - \frac{gL^2}{2}(1 - \cos\varphi) \right) = \frac{m(L\dot{\varphi})^2}{3 \cdot 2} - \frac{mgl}{2}(1 - \cos\varphi)$$

Подставляем это в уравнение Лагранжа, получаем:

$$m \frac{L^2}{3} \ddot{\varphi} = -mg \frac{L}{2} \sin\varphi$$

Отсюда ур-е колебания маятника:

$$\ddot{\varphi} + \frac{3}{2} \frac{g}{L} \sin\varphi = 0$$

Частота колебаний $\sqrt{\frac{3}{2} \frac{g}{L}}$

Напишите функцию Лагранжа для замкнутой системы из N взаимодействующих материальных точек и выведите закон сохранения энергии из гипотезы однородности времени.

Рассмотрим систему взаимодействующих материальных точек. Если точки системы взаимодействуют только друг с другом (на систему не действуют внешние силы), система является замкнутой.

Механическая система характеризуется обобщёнными координатами q и обобщёнными скоростями $\dot{q}$. Механической системе ставится в соответствие функция Лагранжа — лагранжиан, зависящая от обобщённых координат и скоростей, и, возможно, непосредственно от времени — $L(q, \dot{q}, t)$.

Выражение для функции Лагранжа для замкнутой системы:

$$L = \sum_{i=1}^N \frac{m_i v_i^2}{2} - U_B(r_1, \dots, r_N),$$

где первое слагаемое — кинетическая энергия, а второе слагаемое — потенциальная.

Система уравнений Лагранжа:

$$\frac{d}{dt} \frac{\delta L}{\delta \dot{q}_i} - \frac{\delta L}{\delta q_i} = 0, i = 1, \dots, n$$

Однородность времени означает, что функция Лагранжа замкнутой системы не зависит явно от времени. Тогда полная производная функции Лагранжа по времени будет иметь вид:

$$\frac{dL}{dt} = \sum_{i=1}^N \frac{\delta L}{\delta q_i} \dot{q}_i + \sum_{i=1}^N \frac{\delta L}{\delta \dot{q}_i} \ddot{q}_i \quad (1)$$

Используя уравнения Лагранжа отсюда получаем:

$$\frac{dL}{dt} = \sum_{i=1}^N \left(\frac{d}{dt} \frac{\delta L}{\delta \dot{q}_i} \right) \dot{q}_i + \sum_{i=1}^N \frac{\delta L}{\delta \dot{q}_i} \ddot{q}_i = \sum_{i=1}^N \frac{d}{dt} \left(\frac{\delta L}{\delta \dot{q}_i} \dot{q}_i \right)$$

Отсюда получаем:

$$\frac{d}{dt} \left(\sum_{i=1}^N \frac{\delta L}{\delta \dot{q}_i} \dot{q}_i - L \right) = 0$$

Следовательно:

$$\sum_{i=1}^N \frac{\delta L}{\delta \dot{q}_i} \dot{q}_i - L = const$$

Величина $E = \sum_{i=1}^N \frac{\delta L}{\delta \dot{q}_i} \dot{q}_i - L$ (2) остается неизменной при движении системы. Механические системы, энергии которых сохраняются, называются консервативными.

Подставляя в (2) выражение для функции Лагранжа для замкнутой системы, получаем:

$$E = \sum_{i=1}^N \frac{m_i v_i^2}{2} + U_B(r_1, \dots, r_N)$$

Отсюда видно, что энергия системы представляет собой сумму кинетической энергии, зависящей от скоростей, и потенциальной, зависящей только от координат точек системы.

Билет 22

Покажите, что функция $(\frac{e^x/x^*}{x/x^*})^m (\frac{e^y/y^*}{y/y^*})^\alpha$ является первым интегралом для системы ОДУ Лотки-Вольтерра $\frac{dx}{dt} = \alpha x - \beta xy$, $\frac{dy}{dt} = k\beta xy - my$, описывающей сосуществование хищников y и жертв x , где x^* , y^* - стационарное решение этой системы.

Рассмотрим нормальную систему дифференциальных уравнений 2-го порядка:

$$\begin{aligned}\frac{dx(t)}{dt} &= f_1(t, x(t), y(t)), \\ \frac{dy(t)}{dt} &= f_2(t, x(t), y(t)).\end{aligned}\tag{1}$$

Определение 1.

Первым интегралом нормальной системы (1) называется функция $F(t, x(t), y(t)) \equiv C$, сохраняющая постоянное значение вдоль каждой интегральной кривой системы (1).

Утверждение 1.

Функция $F(t, x(t), y(t)) \equiv C$ является первым интегралом системы (1) тогда и только тогда, когда ее производная в силу системы (1) равна нулю:

$$\left. \frac{dF}{dt}(t, x(t), y(t)) \right|_{(1)} = 0.$$

Система хищник-жертва. Размножение жертв ограничивается давлением на них со стороны хищников. Размножение хищников ограничивается количеством добытой ими пищи (количеством жертв):

$$\begin{aligned}\frac{dx}{dt} &= \alpha x - \beta xy, \\ \frac{dy}{dt} &= k\beta xy - my.\end{aligned}\tag{2}$$

- $x(t)$ - численность жертв;
- $y(t)$ - численность хищников;
- α - коэффициент естественного прироста жертв;
- m - коэффициент естественной смертности хищников;
- $V(x) = \beta x$ - количество (биомасса) жертв, потребляемых одним хищником за единицу времени, причем k -я часть полученной биомассы энергии расходуется хищником на воспроизводство, остальное тратится на поддержание основного обмена и охотничьей активности. Функцию $V(x)$ обычно называют трофической функцией или функциональным откликом хищника на плотность популяции жертвы.

$x^* = \frac{m}{k\beta}$, $y^* = \alpha/\beta$ - стационарное решение системы (2) (проверяется подстановкой). Докажем, что функция $(\frac{e^X}{X})^m (\frac{e^Y}{Y})^\alpha$, где $X = x/x^*$, $Y = y/y^*$ является первым интегралом системы (2). Продифференцируем эту функцию по t :

$$\begin{aligned}
 \frac{\partial}{\partial t}[(\frac{e^X}{X})^m (\frac{e^Y}{Y})^\alpha] &= m(\frac{e^X}{X})^{m-1} \cdot \frac{e^X \dot{X} X - e^X \dot{X}}{X^2} \cdot (\frac{e^Y}{Y})^\alpha + \alpha(\frac{e^Y}{Y})^{\alpha-1} \cdot \frac{e^Y \dot{Y} Y - e^Y \dot{Y}}{Y^2} \cdot (\frac{e^X}{X})^m = \\
 &= (\frac{e^X}{X})^m \cdot (\frac{e^Y}{Y})^\alpha \cdot [m\dot{X}(1 - \frac{1}{X}) + \alpha\dot{Y}(1 - \frac{1}{Y})] = \underbrace{(\frac{e^X}{X})^m \cdot (\frac{e^Y}{Y})^\alpha}_A \cdot [m\dot{X} + \alpha\dot{Y} - (m\frac{\dot{X}}{X} + \alpha\frac{\dot{Y}}{Y})] = \\
 &= \left\{ X = x/x^*, Y = y/y^*, x^* = \frac{m}{k\beta}, y^* = \alpha/\beta \right\} = A \cdot [\dot{x}k\beta - \dot{y}\beta - (m\frac{\dot{x}}{x} + \alpha\frac{\dot{y}}{y})] = \left\{ (2) \right\} = \\
 &= A \cdot [k\beta(\alpha x - \beta xy) + \beta(k\beta xy - my) - (\alpha m - \beta my + \alpha k\beta x - \alpha m)] = \\
 &= A \cdot [k\beta\alpha x - \beta my - (-\beta my + \alpha k\beta x)] = 0.
 \end{aligned}$$

Из утверждения 1 следует, что функция $(\frac{e^X}{X})^m (\frac{e^Y}{Y})^\alpha$ является первым интегралом системы (2). Уравнение $(\frac{e^X}{X})^m (\frac{e^Y}{Y})^\alpha = C$ описывает семейство вложенных друг в друга замкнутых кривых, соответствующих фазовым траекториям периодических решений системы (2). Заметим, что при увеличении C амплитуды колебаний x и y возрастают (рисунок 1).

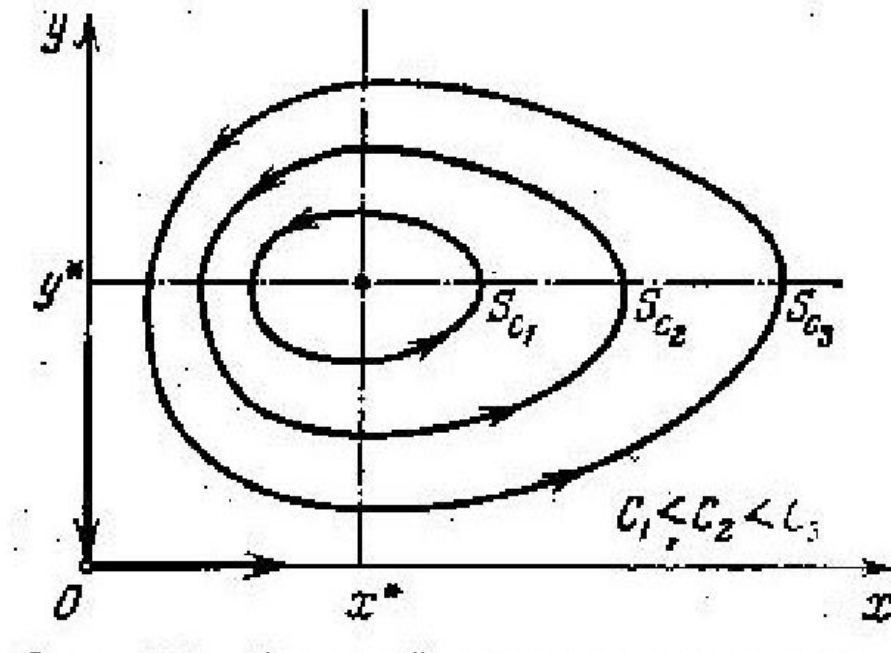


Рис. 1. Фазовый портрет

23. Сформулируйте необходимые и достаточные условия продуктивности неотрицательной, неразложимой матрицы в модели Леонтьева и докажите, что матрица является продуктивной или не является таковой.

Определение 1. Квадратная матрица $A (N \times N)$ называется *разложимой*, если одновременной перестановкой строк и столбцов её можно представить в виде

где $A_1 (m \times m)$, $A_3 ((N-m) \times (N-m))$, $m < N$.

$$\begin{bmatrix} A_1 & A_2 \\ 0 & A_3 \end{bmatrix},$$

Утверждение 1. Если матрица A разложима, то и матрица A^2 разложима, причём

$$A^2 = \begin{bmatrix} A_1^2 & \tilde{A}_2 \\ 0 & A_3^2 \end{bmatrix}$$

Теорема (Фробениуса-Перрона). Если неотрицательная матрица A неразложима, то у неё существует собственное число λ_A такое, что для любого другого собственного числа λ справедливо $|\lambda| < \lambda_A$, а соответствующий правый собственный вектор-столбец $\mathbf{x}_A$, ($A\mathbf{x}_A = \lambda_A \mathbf{x}_A$), и левый собственный вектор-строка $\mathbf{y}_A$, ($\mathbf{y}_A A = \lambda_A \mathbf{y}_A$), положительны.

Собственное число λ_A называют *фробениусовым числом*, а векторы $\mathbf{x}_A$, $\mathbf{y}_A$ — *правым и левым фробениусовыми векторами*.

Пусть $r = \min r_i$, $R = \max r_i$, где $r_i = \sum_{j=1}^N a_{ij}$.

Утверждение 3. Если A неотрицательная неразложимая матрица, то при $r < R$ ее фробениусово число $r < \lambda_A < R$. Если же $r = R$, то $\lambda_A = R$.

Доказательство Пусть вектор $\mathbf{y}_A$ таков, что $\sum_{i=1}^N y_A^i = 1$. Умножим равенство $\mathbf{y}_A A = \lambda_A \mathbf{y}_A$ справа на вектор-столбец $\mathbf{e} = (1, 1, \dots, 1)^T$. Получим $\sum_{j=1}^N \sum_{i=1}^N a_{ij} y_A^i = \lambda_A \sum_{i=1}^N y_A^i = \lambda_A$. Поменяв в левой части равенства порядок суммирования, имеем $\sum_{i=1}^N y_A^i \sum_{j=1}^N a_{ij} = \sum_{i=1}^N y_A^i r_i = \lambda_A$. Строгие неравенства следуют из того, что вектор $\mathbf{y}_A$ положителен и $r < R$. Если же $r = R$, то $\lambda_A = R$

Определение 3. Неотрицательную матрицу A назовём *продуктивной*, если существует неотрицательная матрица $(E-A)^{-1}$.

Утверждение 4. Неотрицательная неразложимая матрица A продуктивна тогда и только тогда, когда её фробениусово число $\lambda_A < 1$.

Доказательство: Достаточность. Пусть $\lambda_A < 1$. Если $\mathbf{x}_A$ правый фробениусов вектор, то $A\mathbf{x}_A = \lambda_A \mathbf{x}_A$. Отсюда и из положительности компонент вектора $\mathbf{x}_A$ следует, что $A^k \rightarrow 0$.

$$\begin{aligned} E - A^k &= (E - A)(E + A + A^2 + \dots + A^{k-1}) \\ E &= \lim_{k \rightarrow \infty} (E - A^k) = (E - A) \lim_{k \rightarrow \infty} (E + A + A^2 + \dots + A^{k-1}) \\ E &= (E - A)(E - A)^{-1} = (E - A) \cdot \text{const}, \\ &\text{где } \text{const} = E + A + A^2 + \dots + A^{k-1} \end{aligned}$$

Необходимость. Пусть матрица $(E-A)^{-1} \geq 0$. Для вектора $\mathbf{p} > 0$ решение $(E-A)\mathbf{x} = \mathbf{p}$ существует и $\mathbf{x} = (E-A)^{-1}\mathbf{p} > 0$. Очевидны соотношения $\lambda_A \mathbf{y}_A \mathbf{x} = \mathbf{y}_A A \mathbf{x} = \mathbf{y}_A (\mathbf{x} - \mathbf{p}) < \mathbf{y}_A \mathbf{x}$. Отсюда и из положительности числа $\mathbf{y}_A \mathbf{x}$ следует неравенство $\lambda_A < 1$ ■

Задача.

$$A = \begin{pmatrix} 0,5 & 0,1 & 0,1 \\ 0,4 & 0 & 0,6 \\ 0,3 & 0,3 & 0,3 \end{pmatrix}$$

Для того, чтобы определить продуктивна матрица A или нет мы должны найти её фробениусово число λ_A . В случае, если $\lambda_A < 1$, матрица A будет продуктивной, иначе – нет.

$r_1 = 0,7, r_2 = 1, r_3 = 0,9$, значит $r = 0,7, R = 1$. Поскольку $r < R$, то $\lambda_A < R = 1$. Ответ: Матрица A продуктивна.

24. Опишите математическую модель коллективного поведения П.С. Краснощёкова и рассмотрите случай, когда у всех членов коллектива коэффициенты индивидуализма равны нулю (стадо).

Пусть ЛПР ($i = 1, N$) может принять одно из двух $S = 1,2$ альтернативных решений.

Пусть μ_i есть коэффициент индивидуализма (самостоятельности) i -ого ЛПР. При $\mu = 1$ данное ЛПР абсолютно самостоятельно, а при $\mu = 0$ — это абсолютный конформист, меняющий решения в угоду любому чужому мнению.

α_j - априорная вероятность принятия индивидом j решения "1"

Апостериорная вероятность (вероятность после общения с коллективом) того, что i -ый индивид примет решение с номером "1" равно $p_i \in [0, 1]$, а с номером "2" равно $1 - p_i$.

Модель Краснощёкова описывает способ вычисления апостериорной вероятности принятия решения $s = 1$ с учетом коэффициента индивидуализма:

$$p_i = \alpha_i \mu_i + (1 - \mu_i) \sum_{j=1}^N \lambda_{ij} p_j,$$

В этой формуле $\lambda_{ij} \in [0, 1]$ — вероятность того, что i -е ЛПР примет решение $s = 1$ после общения с j -м ЛПР, при условии, что j -е ЛПР придерживается $s = 1$ -й альтернативы.

Модель Краснощёкова в матричной форме:

$$P = AM + (E - M)\Lambda P$$

В случае "стада" ($\mu_i = 0$ для всех i) $M = 0$. Таким образом, $(E - \Lambda)P = 0$. То есть определитель матрицы $E - \Lambda$ равен нулю, а у этого уравнения

существует бесконечное число решений. Однако, стадо движется не хаотично: все p_i равны между собой, то есть индивиды подражают друг другу:

$$p_i = \sum_{j=1}^N \lambda_{ij} p_j$$

При этом, если в коллективе появился индивид, у которого $\mu_k > 0$, то решения модели становится единственным: $p_i = \alpha_k$, то есть все копируют поведение k —того индивида.

25. Популяционные игры. Равновесие Нэша и строгое равновесие.

Популяционная игра - статическая модель взаимодействия в большой однородной группе индивидуумов.

$$G = \langle J, f_j(\pi, \omega), j \in J, \pi \in \Pi, \omega \in \Omega \rangle,$$

- J - множество стратегий участников игры;
- $\pi = (\pi_j)_{j \in J}$ - распределение игроков по стратегиям;
- $\Pi = \{\pi | \pi_j > 0 \sum_{j \in J} \pi_j = 1\}$ - стандартный симплекс;
- $f_j(\pi, \omega)$ - выигрыш игроков, использующих стратегию j в зависимости от распределения по стратегиям π и других параметров модели ω (например, общей численности популяции и состояния внешней среды);

Равновесием по Нэшу популяционной игры G называется такое распределение π^* , что всякая стратегия, используемая с положительной частотой, является оптимальным ответом на данное распределение при любом значении параметра ω , т.е

$$\forall \omega \in \Omega, \forall j \in J (\pi_j^* > 0) \Rightarrow j \in \text{Arg max}_{i \in J} f_i(\pi^*, \omega).$$

Если $f_j(\pi, \omega) = \alpha(\pi, \omega) \bar{f}_j(\pi) + b(\pi, \omega)$, т.е. функции выигрыша разложимы, то: $\forall j \in J : \pi_j^* > 0 \rightarrow j \in \text{Arg max}_{i \in J} \bar{f}_i(\pi^*)$.

Эволюционно устойчивой стратегией (ЭУС) для популяционной игры G называется такое распределение π^* , что

$$\forall \omega \in \Omega, \forall \pi \neq \pi^*, \exists \bar{\lambda}(\pi) \in (0, 1) : \forall \lambda \in (0, \bar{\lambda}(\pi)) f_{\pi^*}(\lambda \pi + (1 - \lambda) \pi^*, \omega) > f_{\pi}(\lambda \pi + (1 - \lambda) \pi^*, \omega), \text{ где } f_{\pi}(\pi', \omega) = \sum_{j \in J} \pi_j f_j(\pi', \omega) -$$

средний выигрыш смешанной стратегии, или распределения π , если индивидуумы в популяции распределены по чистым стратегиям.

Понятие ЭУС можно интерпретировать следующим образом. Пусть в некоторую популяцию, находящуюся в состоянии равновесия π^* , внедряется относительно небольшая группа "мутантов" с распределением

по стратегиям π . Тогда, если распределение π^* является эволюционно устойчивым, то внедрившаяся группа не сможет закрепиться в популяции, так как ее средняя приспособленность меньше, чем приспособленность исходной стратегии π^* .

Всякая ЭУС является равновесием Нэша. Действительно, если π не является равновесием, то мутанты с чистой стратегией лучшего ответа на π получают больший выигрыш, чем средний выигрыш в основной популяции. В силу непрерывности f относительно π это справедливо для любой достаточно малой группы таких мутантов. Данное утверждение справедливо, если доля отдельного индивида в популяции пренебрежимо мала в том смысле, что изменение его стратегии не влияет на значения функций выигрыша.

Уточним понятие ЭУС для взаимодействий в группах конечной численности, где изменение стратегии отдельного индивида влияет на значение функций выигрыша.

Для симметричной игры в нормальной форме с n игроками, множеством стратегий S и функцией выигрыша $f_j(s_j, s_{J \setminus j})$ ЭУС определяется как симметричная ситуация $s_j \equiv s$ такая, что при любом изменении стратегии отдельным игроком его выигрыш в новой ситуации будет не больше, чем выигрыш любого из остальных игроков, сохранивших прежнюю стратегию.

Строгим равновесием популяционной игры G называется такое распределение π^* , что все игроки используют одну и ту же стратегию, которая является единственно лучшим ответом на это распределение:

$$\exists \varepsilon > 0, \exists j \in J : \pi_j^* = 1, \text{ и } \forall i \neq j, \forall \omega \in \Omega : f_j(\pi^*, \omega) > f_i(\pi^*, \omega) + \varepsilon.$$

Всякое строгое равновесие является ЭУС.

26. Модель динамики репликаторов.

В МДР предполагается, что в однородной популяции новые индивидуумы (потомки) наследуют стратегии родителей и сохраняют их в течение всего времени жизни.

Пусть взаимодействие между индивидуумами популяции в каждый период времени описывается популяционной игрой $G = \langle J, f_j(\pi, N), j \in J, \pi \in \Pi, N \geq 0 \rangle$. Выигрыш $f_j(\pi, N)$ представляет собой сумму средней рождаемости $fer_j(\pi, N)$ и выживаемости $\nu(\pi, N)$ для индивидуумов, использующих стратегию j , если общая численность популяции равна N , а распределение по стратегиям - π . Обозначим $N_j(t)$ число особей, использующих в период t стратегию $j \in J$. Тогда динамика такой популяции описывается системой

$$N_j(t+1) = N_j(t)f_j(\pi(t), N(t)) \quad (1)$$

Таким образом, состояние системы в каждый период времени полностью описывается $\bar{N}(t) = (N_j(t))_{j \in J}$, по которому однозначно определяется распределение по стратегиям $\pi(t) = (\pi_j(t) = N_j(t)/N(t))_{j \in J}$. Отметим, что в случае, когда $f_j(\pi) = \alpha(\pi, N)\bar{f}_j(\pi)$ из системы (1) вытекает автономная модель динамики $\pi(t)$:

$$\pi_j(t+1) = \pi_j(t)\bar{f}_j(\pi(t)) / \sum_{i \in J} \pi_i(t)\bar{f}_i(\pi(t)), j \in J.$$

27. Теоремы о связи равновесий Нэша и строгих равновесий с устойчивыми точками модели динамики репликаторов.

$G = \langle J, f_j(\pi, N), j \in J, \pi \in \Pi, N \geq 0 \rangle$ - популяционная игра.

$f_j(\pi, \omega)$ - выигрыш игроков, использующих стратегию j , π - распределение по стратегиям, N - общая численность популяции.

$N_j(t+1) = N_j(t)f_j(\pi(t), N(t))$ - динамика такой популяции.

Состояние системы в каждый период полностью характеризуется вектором $\bar{N}(t) = (N_j(t))_{j \in J}$, по которому однозначно определяется распределение по стратегиям $\pi(t) = (\pi_j(t) = \frac{N_j(t)}{N(t)})_{j \in J}$.

$N_s = \pi_s N$ - численное использование стратегии s .

(*) $N_s(t+1) = N_s(t)f_s(\pi(t), N(t)), t = 1, 2, \dots$, где $f_s(\pi, N) = fer_s(\pi, N) + \nu_s(\pi, N)$ - функция приспособленности стратегии s .

Теорема 3.1 (о связи р. Нэша и уст. точек МДР). $\square f_s$ разложима:

$$f_s(\pi, \omega) = a(\pi, \omega)\bar{f}_s(\pi) + b(\pi, \omega), a(\pi, \omega) \geq 0 \Rightarrow$$

1) $\forall$ уст. по Ляпунову распределение π^* системы (*) является р. по Нэшу популяционной игры $\bar{G} = \langle S, \bar{f}_s(\pi), s \in S, \pi \in \Pi \rangle$

2) если начальное распределение $\bar{N}(0) > 0$ и для траектории $\{\bar{N}(t)\} \exists \lim_{t \rightarrow \infty} \pi(t, \bar{N}(0)) = \pi^*$, то π^* является р. Нэша указанной популяционной игры.

Теорема 3.2 (об асимптотически уст. ЭУС). Пусть в усл. т. 3.1 π^* - ЭУС для игры $\bar{G} \Rightarrow \pi^*$ - асимптотически уст. распределение системы (*).

Теорема 3.3 (о связи множеств доминирующих стратегий с динамикой поведения). Пусть S - множество строго доминирующих стратегий в игре $G' = \langle S, \ln \bar{f}_s(\pi), s \in S, \pi \in \Pi \rangle$. Тогда для $\forall s \in S$ и $\forall \bar{N}(0) > 0$ $\lim_{t \rightarrow \infty} \pi_s(t, \bar{N}(0)) = 0$ на соответствующей траектории системы (*).

28. Модель взаимодействия родственников.

Утверждение о доминирующей стратегии.

Распространение альтруизма и кооперации.

Пусть взаимодействие в популяции характеризуется множеством стратегий S и функциями приспособленности $f_s(\pi)$, $s \in S$. Главная особенность данной модели – индивидуум способен различать сибсов (родных братьев или сестер) среди других членов популяции и выбирать стратегию в зависимости от этого признака.

Полная стратегия (s, s') включает:

- Компоненту s , применяемую к сибсам
- Компоненту s' для прочих индивидуумов

В данный период времени индивидуум с некоторой интенсивностью $\lambda_r \in (0,1)$ взаимодействует с сибсами, а с интенсивностью $1 - \lambda_r$ – с остальными индивидуумами из популяции. Общая приспособленность аддитивно зависит от результатов взаимодействия с родственниками и с остальной частью популяции:

$$\bar{f}_{(s,s')}(\pi') = \lambda_r f_s^r(s) + (1 - \lambda_r) f_{s'}(\pi'),$$

- $f_s^r(s)$ – функция, описывающая результаты взаимодействия с сибсами
- π' – распределение по компоненте s'
- Предполагается, что все сибсы применяют одну (унаследованную) стратегию

Таким образом, взаимодействие характеризуется популяционной игрой:

$$\bar{G} = \langle \bar{S} = \{(s, s') \in S \times S\}, \bar{f}_{(s,s')}(\bar{\pi}) = \lambda_r f_s^r(s) + (1 - \lambda_r) f_{s'}(\pi') \rangle,$$

где $\bar{\pi}$ – распределение по полным стратегиям.

Теорема 1.

Всякая стратегия (s, s') , в которой $s \notin \underset{i}{\operatorname{Argmax}} f_i^r(i)$, строго доминируется стратегией (s^*, s') , где $s^* \in \underset{i}{\operatorname{Argmax}} f_i^r(i)$. Распределение $\bar{\pi}$ является равновесием Нэша в том и только том случае, если для всех указанных неоптимальных стратегий $\pi_{ss'} = 0$, а соответствующие распределение π' является равновесием Нэша для игры $\langle S, f_{s'}(\pi') \rangle$.

Таким образом, в отношениях типа “дилеммы заключенного” сибсы используют кооперативную стратегию применительно друг к другу.

Чтобы объяснить распространение альтруистического поведения, рассмотрим следующую модификацию модели.

Если отказаться от предположения об одинаковом состоянии сибсов и считать, что они могут оказаться в разных ролях (например, доминирующей и подчиненной), а стратегия и функция приспособленности могут зависеть от роли, то полная стратегия в отношении родственников задается парой $s = (s_\alpha, s_\beta) \in S \times S$.

В результате исключения строго доминируемых стратегий остаются лишь стратегии $s^* = (s_\alpha^*, s_\beta^*) \rightarrow \max_{(s_\alpha, s_\beta)} (f^\alpha(s) + f^\beta(s))$, обеспечивающие максимум суммарной приспособленности.

На основании теоремы 1 можно сделать вывод, что эволюция поведения в самовоспроизводящихся популяциях ведет к формированию поведения, максимизирующего суммарную приспособленность сибсов.

Кооперация и альтруизм.

Дилемма заключенного:

Два игрока, у каждого две стратегии: кооперативная (К) и эгоистичная (Э). При любом поведении партнера выгоднее эгоистичная стратегия. Суммарный выигрыш максимален, когда оба действуют кооперативно.

Пример:

	К	Э
К	(5;5)	(1;6)
Э	(6;1)	(2;2)

В данной игре существует единственная точка равновесия Нэша, которая соответствует эгоистичному поведению и является решением по доминированию. В реальности игроки часто ведут себя кооперативно.

Еще в больше степени отклоняется от максимизации индивидуальной приспособленности альтруистическое поведение (А – альтруистическое, Э – эгоистичное).

	А	Э
А	(5;5)	(1;10)
Э	(10;1)	(2;2)

Здесь альтруистическое поведение одного из партнеров в сочетании с эгоистичным поведением другого соответствует максимизации суммарной приспособленности. При этом альтруист получает меньше своего гарантированного выигрыша, который он мог бы получить в равновесии Нэша.

Пример кооперации:

Охота хищников, эгоисты экономят затраты в ущерб общей эффективности.

Пример альтруизма:

Альтруизм родителей по отношению к детям.

Билет 29

Многоуровневая модель налоговой инспекции.

В базовой модели налоговой инспекции с учетом коррупции предполагалось, что инспекторы и налогоплательщики максимизируют ожидаемые доходы, проверка требует фиксированных издержек и всегда верно определяет категорию проверяемого агента. Однако инспектор, обнаруживший факт уклонения, может быть подкуплен пойманным субъектом. Величина взятки лежит между максимальной приемлемой для плательщика и минимальной приемлемой для инспектора. Для устранения коррупции руководство проводит ревизии: перепроверяет некоторых инспекторов и увольняет скрывших уклонение от уплаты налогов. Стратегия организации инспекции заключается в определении вероятностей проверок и ревизий при фиксированных затратах на одну проверку. Организатор инспекции максимизирует чистый налоговый доход.

В модели предполагается способность центра контролировать фактическую частоту проверок без проведения ревизий. Также есть проблема возможной нехватки честных ревизоров.

Для устранения различных ограничений используется иерархическая контролирующая структура (она и описана ниже), подавляющая коррупцию на всех уровнях с привлечением малого числа честных инспекторов. Предполагается, что в распоряжении организатора инспекции есть доверенные лица, которые проводят проверки на верхнем уровне и всегда проверяют правильно, но стоимость их работы очень высока. Также организатор может привлекать для проверок неограниченное кол-во рациональных инспекторов, готовых брать взятки, если им это выгодно. Организатор определяет кол-во уровней инспекции, вероятность проведения проверки каждым уровнем и зарплаты рациональных инспекторов.

Базовая модель. Формализация задачи.

- Рассматривается фиксированное число N агентов уровня 0 (налогоплательщиков).
- Для каждого из них определен возможный набор действий (налоговых платежей) T_0 . Каждое действие t_0 характеризуется затратами агента. Оптимальное с точки зрения инспекции действие $t_0^*(I)$ зависит от значения некоторой случайной величины (дохода) $I \in [I_{\min}, I_{\max}]$ (например, $t_0^*(I)$ – заданное налоговое правило). $t_0 \in [t_{\min}, t_{\max}]$, где $t_{\min} = t_0^*(I_{\min})$, $t_{\max} = t_0^*(I_{\max})$.
- Независимые и одинаково распределенные случайные величины I имеют функцию распределения $F(I)$, известную всем участникам инспекции.
- Для проведения инспекции могут использоваться 2 типа сотрудников:
 - доверенные лица лидера, издержки на проверку которыми очень высоки.
 - любое необходимое число рациональных инспекторов, максимизирующих свой ожидаемый доход с учетом зарплат, взяток и штрафов.
- Проблема контроля возникает в связи с тем, что конкретное значение случайного фактора I наблюдается только действующим на нижнем уровне агентом.

Базовая модель. Задачи с фиксированными затратами и штрафами

Контролирующая иерархическая структура строится следующим образом:

- Инспекторы первого уровня проверяют агентов уровня 0 с вероятностью $p_1(t_0)$.
- Если проверка выявляет $t_0 < t_0^*$, то агент нулевого уровня выплачивает штраф $f_0(t_0^*(I) - t_0(I))$, $f_0 > 1$.
Стоимость одной проверки на этом уровне составляет c_1 .
- Инспектор первого уровня может вступить в сговор с проверяемым агентом. Для предотвращения коррупции организуется проверка 2-го уровня.
- Вероятность проверки $p_2(t_0, t_1)$ зависит от сообщений агентов уровней 0 и 1.
- ...
- На верхнем уровне k честными инспекторами осуществляется проверка с вероятностью $p_k(t_0, t_1, \dots, t_{k-1})$.
- Если проверка уровня l выявляет $t_l > t_{l-1}$, то все агенты нижестоящих уровней r ($r = 0, 1, \dots, l-1$) в этой цепочке проверок платят штраф $f_r(t_l - t_{l-1})$.

Замечание 1

В соответствии с нашим подходом, целью инспекции не является выявление коррупции (поскольку это сложно реализуемо и затратно). Вместо этого предлагается предотвратить отклонение от честного поведения на каждом уровне.

Стратегия инспекции P включает:

- Количество уровней k
- Вероятности проверок $p_1(t_0), \dots, p_k(t_0, \dots, t_{k-1})$.

Следующие параметры заданы экзогенно в этой модели:

- Штрафные коэффициенты $f_0, \dots, f_{k-1}$
- Расходы на проверки $c_1, \dots, c_k$.

Задача

Задача состоит в нахождении стратегии инспекции, подавляющей коррупцию и обеспечивающей правильные действия агентов нулевого уровня с минимальными издержками на проверки.

Билет 30. Оптимальная стратегия проверок при фиксированных затратах на проверки и штрафах.

Рассматривается фиксированное число N агентов уровня 0 (налогоплательщиков). Для каждого из них определен возможный набор действий (налоговых платежей) T_0 . Каждое действие t_0 характеризуется затратами агента. Оптимальное с точки зрения инспекции действие $t_0^*(I)$ зависит от значения некоторой случайной величины (дохода) $I \in [I_{min}, I_{max}]$ (например, $t_0^*(I)$ - заданное налоговое правило). $t_0 \in [t_{min}, t_{max}]$, где $t_{min} = t_0^*(I_{min})$, $t_{max} = t_0^*(I_{max})$.

Независимые и одинаково распределенные случайные величины I имеют функцию распределения $F(I)$, известную всем участникам инспекции. Для проведения инспекции могут использоваться 2 типа сотрудников:

доверенные лица лидера, издержки на проверку которыми очень высоки, любое необходимое число рациональных инспекторов, максимизирующих свой ожидаемый доход с учетом зарплат, взяток и штрафов.

Проблема контроля возникает в связи с тем, что конкретное значение случайного фактора I наблюдается только действующим на нижнем уровне агентом.

Контролирующая иерархическая структура строится следующим образом:

- Инспекторы первого уровня проверяют агентов уровня 0 с вероятностью $p_1(t_0)$;
- Если проверка выявляет $t_0 < t_0^*$, то агент нулевого уровня выплачивает штраф $f_0(t_0^*(I) - t_0(I))$, $f_0 > 1$. Стоимость одной проверки на этом уровне составляет c_1 ;
- Инспектор первого уровня может вступить в сговор с проверяемым агентом. Для предотвращения коррупции организуется проверка 2 уровня;
- Вероятность проверки $p_2(t_0, t_1)$ зависит от сообщений агентов уровней 0 и 1;
- ...;
- На верхнем уровне k честными инспекторами осуществляется проверка с вероятностью $p_k(t_0, t_1, \dots, t_{k-1})$;
- Если проверка уровня l выявляет $t_l > t_{l-1}$, то все агенты нижестоящих уровней l в этой цепочке платят штраф $f_r(t_l - t_{l-1})$.

В связи с нашим подходом, целью инспекции не является выявление коррупции (это сложно реализуемо и затратно). Вместо этого предлагается предотвратить отклонение от честного поведения на каждом уровне.

Стратегия инспекции P включает:

количество уровней k , вероятности проверок $p_1(t_0), \dots, p_k(t_0, t_1, \dots, t_{k-1})$.

Следующие параметры заданы экзогенно в этой модели:

штрафные коэффициенты $f_0, \dots, f_{k-1}$, расходы на проверки $c_1, \dots, c_k$.

Задача состоит в нахождении стратегии инспекции, подавляющей коррупцию и обеспечивающей правильные действия агентов нулевого уровня с минимальными издержками на проверки.

Рассмотрим коалицию C_l , включающую некоторое количество агентов уровня 0 и инспекторов уровней $1, \dots, l, l < k$, проверяющих работу агентов из этой коалиции. Стратегия C_l

задается функциями $t_0(I), \dots, t_l(I)$, определяющими сообщения уровней $i=0, \dots, l$ в случае проверки какого-либо агента уровня 0 из этой коалиции.

Определение. Назовем стратегию Р устойчивой к отклонению коалиции C_l , если суммарный выигрыш ее членов достигает максимума при честном поведении, т.е.:

$$t_0(I) = t_0^*(I), t_r(I) = t_0^*(I), r = 1, \dots, l,$$

при условии честного поведения агентов верхних уровней $l+1, \dots, k-1$.

Назовем стратегию Р устойчивой к коалиционным отклонениям, если это условие выполнено для всех $l=1, \dots, k-1$.

При честном поведении ожидаемые затраты на одного агента составят:

$$\int_{I_{min}}^{I_{max}} (p_1(P, I)(c_1 + p_2(P, I)(c_2 + \dots + p_{k-1}(P, I)(c_{k-1} + p_k(P, I)c_k) \dots)) dF(I)$$

где $p_i(P, I) = p_i(t_0^*(I), \dots, t_0^*(I))$.

Утверждение 1. Оптимальные вероятности проверок в стратегии, устойчивой к коалиционным отклонениям, удовлетворяют условиям

$$p_1(t_0) = \hat{p}_1 = \frac{1}{f_0}, p_s(t_0, t_1, \dots, t_{s-1}) = \hat{p}_s = \frac{\sum_{i=0}^{s-2} f_i}{\sum_{i=0}^{s-2} f_i}$$

для любых $t_0 \leq t_1 \leq \dots \leq t_{s-1} < t_{max}, s = 2, \dots, k$.

Теперь для заданной стратегии Р рассмотрим некооперативное СПР и определим условия существования СПР, соответствующего честному поведению на всех уровнях $0, 1, \dots, k-1$.

- Рассмотрим случай, когда на каждом уровне $s \leq l-1$ отклонение уровня 0 не было полностью выявлено ($t_{l-1} < t_0^*$).
- При каких условиях возможен взаимовыгодный сговор агентов уровней $0, 1, \dots, l$, если агенты верхних уровней действуют честно?

Сговор, выгодный для всех агентов $0, 1, \dots, l$, возможен, если для некоторых $t_l \in [t_{l-1}, t_0^*(I)), b_{il} \geq 0, i = 1, \dots, l-1$ разрешима следующая система:

$$\begin{cases} p_{l+1}(t_0, \dots, t_l) * f_i * (t_0^*(I) - t_l) + b_{il} < f_i * (t_0^*(I) - t_l), i = 0, \dots, l-1, \\ \sum_{i < l} b_{il} - p_{l+1}(t_0, \dots, t_l) * f_l * (t_0^*(I) - t_l) > 0. \end{cases}$$

Здесь b_{il} - взятка, выплачиваемая агентом уровня i последнему проверяющему, t_l - его сообщение.

Определение. Если для любых $I \in (I_{min}, I_{max}], l = 1, \dots, k-1; t_0 \leq \dots \leq t_l < t_0^*(I)$ система несовместна, будем говорить, что стратегия Р определяет СПР с честным поведением.

Утверждение 2. Стратегия Р определяет СПР с честным поведением тогда и только тогда, когда для любых $t_0, \dots, t_{k-1} < t_{max}$ вероятности проверок удовлетворяют условию:

$$p_1(t_0) \geq \frac{1}{f_0}, p_2(t_0, t_1) \geq \frac{f_0}{f_0 + f_1}, p_s(t_0, t_1, \dots, t_{s-1}) \geq \frac{\sum_{i=0}^{s-2} f_i}{\sum_{i=0}^{s-1} f_i}, s = 2, \dots, k.$$

Утверждение 3. Оптимальная стратегия в классе СПР с честным поведением и оптимальная стратегия, устойчивая к коалиционным отклонениям, совпадают и удовлетворяют условию

$$p_1(t_0) = \hat{p}_1 = \frac{1}{f_0}, p_s(t_0, t_1, \dots, t_{s-1}) = \hat{p}_s = \frac{\sum_{i=0}^{s-2} f_i}{\sum_{i=0}^{s-1} f_i}$$

для любых $t_0 \leq t_1 \leq \dots \leq t_{s-1} < t_{max}, s = 2, \dots, k.$

31. Поколения архитектур компьютеров и парадигмы программирования.

Середина 70-х годов: векторно-конвейерные компьютеры.

Особенности архитектуры: векторные функциональные устройства, зацепление функциональных устройств, векторные команды в системе команд, векторные регистры.

Программирование: векторизация самых внутренних циклов

Суперкомпьютер Cray-1.

Начало 80-х годов: векторно-параллельные компьютеры

Особенности архитектуры: векторные функциональные устройства, зацепление

функциональных устройств, векторные команды в системе команд, векторные регистры.

Небольшое число процессоров объединяются над общей памятью.

Программирование: векторизация самых внутренних циклов и распараллеливание на

внешнем уровне, единое адресное пространство, локальные и глобальные переменные.

Суперкомпьютер Cray X-MP

Начало 90-х годов: массивно-параллельные компьютеры

Особенности архитектуры: тысячи процессоров объединяются с помощью коммуникационной сети по некоторой топологии, распределенная память.

Программирование: обмен сообщениями, отсутствие единого адресного пространства,

PVM, Message Passing Interface. Необходимость выделения массового параллелизма, явного

распределения данных и согласования параллелизма с распределением.

Суперкомпьютер Cray T3D, Суперкомпьютер Intel Paragon XPS140

Середина 90-х годов: параллельные компьютеры с общей памятью

Особенности архитектуры: сотни процессоров объединяются над общей памятью.

Программирование: единое адресное пространство, локальные и глобальные

переменные, Linda, OpenMP

DEC AlphaServer, суперкомпьютер Sun StarFire

Начало 2000-х: кластеры из узлов с общей памятью

Особенности архитектуры: большое число многопроцессорных узлов объединяются вместе

с помощью коммуникационной сети по некоторой топологии, распределенная память; в рамках каждого узла несколько (многоядерных) процессоров объединяются над общей памятью.

Программирование: неоднородная схема MPI+OpenMP; необходимость выделения

массового параллелизма, явное распределение данных, обмен сообщениями на внешнем

уровне; распараллеливание в едином адресном пространстве, локальные и глобальные

переменные на уровне узла с общей памятью.

Суперкомпьютер МГУ "**Чебышев**", "**К**" суперкомпьютер

Середина 2000-х: кластеры из узлов с общей памятью с ускорителями

Особенности архитектуры: большое число многопроцессорных узлов объединяются вместе

с помощью коммуникационной сети по некоторой топологии, распределенная память; в

рамках каждого узла несколько (многоядерных) процессоров объединяются над общей

памятью; на каждом узле несколько ускорителей (GPU, Phi).

Программирование: MPI+OpenMP+OpenCL/CUDA;

Суперкомпьютер МГУ "**Ломоносов**", суперкомпьютер **Tianhe-2**

70-е – Векторизация циклов

80-е – Распараллеливание циклов (внешних) + Векторизация (внутренних)

90-е - MPI

середина 90-х - OpenMP

середина 2000-х - MPI+OpenMP

2010-е - CUDA, OpenCL, MPI+OpenMP+ускорители (GPU, Xeon Phi) ...

Для каждого поколения нужно анализировать алгоритмы (чтобы приспособить их под новую платформу) и описывать найденные свойства (для эффективной реализации).

Описание алгоритмов: математическое описание, сложность, информационный граф, свойства и особенности, ресурс параллелизма,

масштабируемость, локальность данных, детерминированность, вычислительная мощность, входные/выходные данные, вычислительное ядро, локальность вычислений, производительность, эффективность, коммуникационный профиль.

32. Архитектурные особенности современных микропроцессоров.

Микропроцессором (МП) называется программно-управляемое устройство, осуществляющее процесс цифровой обработки информации и управления им и построенное, как правило, на одной БИС (большая интегральная схема).

Архитектурой процессора называется комплекс его аппаратных и программных средств, предоставляемых пользователю.

Особенность	В чем заключается
Многозадачность	Возможность работы в одном из двух режимов: реальном (real) и защищенном (protected). В <i>реальном</i> режиме возможно выполнение только одной программы. В <i>защищенном</i> (protected) режиме обеспечивается выполнение сразу нескольких программ за счет переключения между задачами
Поддержка системы виртуальных машин	возможность моделирования в одном МП работу нескольких компьютеров, управляемых <i>разными</i> ОС.
Конвейерная обработка команд	Одновременное выполнение разных тактов последовательных команд в разных частях МП с непосредственной передачей результатов выполнения из одной части МП в другую.
Кэширование данных	Использование высокоскоростного буфера для обмена данными между микропроцессорной памятью (регистрами МП) и основной памятью ЭВМ.
Суперскалярная архитектура	Наличие в микропроцессоре более 1 конвейера для выполнения команд (параллелизм на уровне инструкций)
Гарвардская архитектура процессора	В кэш-памяти 1 уровня предусмотрено разделение команд и данных, которые хранятся отдельно друг от друга для повышения эффективности обработки
Расширенный набор инструкций	Новые команды, расширяющие базовый набор инструкций МП, для работы с мультимедийной информацией и одновременной однопоточной обработки множественных данных.
Многоядерные процессоры	Объединение двух или более исполняющих устройств на одной ИС, действующих как единое устройство.
Технология автоматического увеличения тактовой частоты процессора	Для обеспечения дополнительной производительности и при условии соблюдения ограничений по мощности, температуре и току, процессор может автоматически «разгоняться», то есть увеличивать рабочую тактовую частоту всех своих ядер.

33. Программно-аппаратная архитектура суперкомпьютеров Ломоносов и Blue Gene/P.

IBM Blue Gene – название архитектуры семейства суперкомпьютеров. Проект начался в 1999 году, изначально использовался для фолдинга белков и расшифровки генома. Всего было 3 поколения: BG/L, BG/P и BG/Q. Это массивно-параллельная (а не кластерная) система, то есть все конструктивные элементы (процессоры, коммуникация) объединены одним аппаратным решением энергоэффективность, быстрая коммуникация. Долго занимал первые места в топ-500.

Одна стойка BG/P состоит из 1024 4-х ядерных вычислительных узлов с производительностью 13.6 GF/s и 2 Gb оперативной памяти. Таким образом, стойка имеет производительность 13.6 TF/s память 1 Tb.

Конфигурация на ВМК включает в себя 2 стойки: 2048 4-х ядерных узлов с 4 Tb памяти и пиковой производительностью 27.2 TF/s. Реальная производительность по тесту Linpack – 23.5 TF/s или 85% от пиковой.

Основной единицей является 4-х ядерный 32-х битный процессор PowerPC 450 850 Mhz. Узлы объединяются в узловую карту, которая содержит 32 вычислительных узла и до 2-х узлов ввода-вывода для обращения к внешней памяти. 32 узловые карты составляют одну стойку.

PowerPC 450 850 Mhz производит операции над числами с плавающей точкой, имеет 2 Gb оперативной памяти, причем физическая память равна виртуальной, и управляется облегченной ОС. ОС умеет создавать, управлять и отлаживать процессы, управлять памятью и производить операции ввода-вывода. Память имеет трехуровневый кэш. Первый и второй уровни – свои для каждого ядра объемом 32 kb и 16 b соответственно. Второй используется для выборки информации из первого. Третий уровень общий для всех ядер, разделен на две части по 4 Mb.

Кроме того, в процессорах используется Double Hammer FPU – сопроцессор, ускоряющий производство операций с вещественными числами.

Процессы могут выполняться в 3-х режимах:

1. SMP-режим: 1 MPI процесс на узел, разделенный на 4 потока, используется 2 Гб памяти на процесс
2. Dual-режим: 2 MPI процесса на узел, разделенный по 2 потока, используется 1 Гб памяти на процесс
3. Quad-режим: 4 MPI процесса на узел, по 1 потоку на каждый, используется 512 Кб памяти на процесс

Для коммуникации между узлами используется несколько коммуникационных сетей:

1. 3-х мерный тор – основная сеть, связывает процессорные элементы и реализует все передачи для MPI. Каждый процессор имеет связь с ближайшими соседями в 3-х мерной решетке. У каждого узла есть 6 соседей. Получается, процессоры образуют “куб”, боковые грани которого замыкаются между собой, то есть тор.
2. Коллективная сеть-дерево – соединяет все вычислительные узлы с узлами ввода-вывода. Имеет большую пропускную способность, но поддерживаются только для прямоугольных решеток.
3. Высокоскоростная сеть глобальных прерываний – для быстрой барьерной синхронизации.

В общем, архитектура BlueGene/P состоит из: вычислительных узлов, многопроцессорный front-end узел (с полноценной ОС, для подготовки программ к выполнению на вычислительных узлах) и служебный узел (мониторинг состояния системы).

Состав ПО: Linux (узлы в-в), компиляторы IBM XL (C++, Fortran) и GNU, система управления заданиями LoadLeveler (нужна для запуска заданий), файловая система GPFS и библиотеки подпрограмм (ESSL и MASS).

ОС узла - Compute Node Kernel. Это урезанная Linux, без некоторых системных вызовов (в основном, динамического создания процессов).

Суперкомпьютер Ломоносов – гибридный суперкомпьютер (помимо ЦП традиционной архитектуры используют для вычислений специализированные процессоры е.г. графические). В нем используется 3 вида вычислительных узлов и процессоры с различной архитектурой. Основные узлы (90% производительности) – blade-платформа T-blade.

Пиковая производительность – 1.7 Пфлопс.

Типы вычислительных узлов:

1. Тип T-Blade2(link is external)(УВ1). 12 Гб оперативной памяти, 2 процессора Intel Xeon 5570 по 4 ядра, всего процессоров 8320, всего узлов 4160.
2. Тип T-Blade1(УВ2). 24 Гб оперативной памяти, 2 процессора Intel Xeon 5570 по 4 ядра, всего процессоров 520, всего узлов 260.
3. Тип T-Blade2(link is external)(УВ1). 24 Гб оперативной памяти, 2 процессора Intel Xeon 5670 по 6 ядер, всего процессоров 1280, всего узлов 640.
4. Тип T-Blade1(УВ2). 48 Гб оперативной памяти, 2 процессора Intel Xeon 5670 по 6 ядер, всего процессоров 80, всего узлов 40.
5. Тип Узлы на базе IBM® Cell (УВ3). 16 Гб оперативной памяти, процессор PowerXCell 8i с 8 ядрами, всего процессоров 60, всего узлов 30.

Состав ПО: Языки программирования: C/C++, Fortran 77/90/95; Наборы компиляторов: Intel 12, GNU 4.4, Pathscale, PGI; Среды исполнения: OpenMPI 1.4, MVAPICH 1.1, IntelMPI 4;

Все узлы связаны тремя независимыми сетями:

1. Системная сеть - QDR InfiniBand, 40 Гбит/сек
2. Сервисная сеть - Ethernet, 10 Гбит/сек, 1 Гбит/сек и 100 Мбит/сек
3. Управляющая сеть - Ethernet, 10 Гбит/сек и 1 Гбит/сек
4. Сеть барьерной синхронизации и сеть глобальных прерываний, T-Платформы

34. Последовательная и параллельная сложность алгоритмов.

Последовательная сложность алгоритма – число операций, которые нужно выполнить при его последовательном выполнении

Параллельная сложность алгоритма – число шагов, за которое можно выполнить алгоритм в предположении неограниченного числа необходимых процессоров (ядер, вычислительных узлов и т.д.)

Граф алгоритма — это ориентированный ациклический мультиграф, вершины которого соответствуют операциям алгоритма, а дуги — передаче данных между ними. Вершины графа алгоритма могут соединяться несколькими дугами, в частности, когда в качестве разных аргументов одной и той же операции используется одна и та же величина. Граф алгоритма почти всегда является параметризованным графом. Он используется как удобное представление алгоритма при исследовании его структуры, ресурса параллелизма, а также других свойств. Его можно рассматривать как параметризованную информационную историю. Он сохраняет ее информативность, при этом обладая компактностью за счет параметризации. Граф можно построить по исходному тексту программы.

Ярусно параллельная форма (ЯПФ) – представление графа алгоритма, в котором:

1. Все вершины разбиты на пронумерованные подмножества ярусов
2. Начальная вершина каждой дуги расположена на ярусе с номером меньшим, чем номер яруса конечной вершины
3. Между вершинами, расположенными на одном ярусе не может быть дуг

Высота ЯПФ – число ярусов. **Ширина яруса** – число вершин, содержащихся в ярусе. **Ширина ЯПФ** – максимальная ширина ярусов в ЯПФ.

Каноническая ЯПФ – ЯПФ, высота которой на единицу больше длины критического пути, а входные вершины расположены в первом ярусе. Для заданного графа его каноническая ЯПФ единственна. Высота канонической ЯПФ соответствует параллельной сложности алгоритма.

Критический путь графа – путь максимальной длины в ориентированном ациклическом графе.

Ресурс параллелизма – это где и насколько мы можем сделать программу более эффективной за счет распараллеливания.

35. Информационный граф и ресурс параллелизма алгоритмов.

Информационный граф (или граф алгоритма) - ориентированный граф, состоящий из вершин, соответствующих операциям, и направленных дуг, соответствующих передаче данных между ними.

Вершины графа алгоритма могут соединиться несколькими дугами, в частности когда в качестве разных аргументов одной операции используется одна и та же величина. Граф алгоритма почти всегда является параметризованным графом.

Граф алгоритма используется как удобное представление алгоритма при исследовании его структуры, ресурса параллелизма, а также других свойств. Его можно рассматривать как параметризованную информационную структуру. Он сохраняет информативность структуры, при этом обладает компактностью за счет параметризации.

Особенностями графа алгоритма являются:

- его ацикличность;
- невозможность, в общем случае, его описания простым перечислением, в силу того, что его составляющие могут зависеть от внешних параметров решаемой им задачи (например, алгоритм, реализующий метод Гаусса — от размера матрицы).

Ресурс параллелизма - это где и насколько мы можем сделать программу более эффективной за счет распараллеливания.

Ресурс параллелизма алгоритма включает:

- Оценку параллельной сложности алгоритма
- Структуру параллелизма (иерархичность)
- Ключевые параллельные ветви в терминах конечного и массового параллелизма

Конечный параллелизм - параллелизм, определяемый информационной независимостью некоторых фрагментов в тексте программы.

Массовый параллелизм - параллелизм, определяемый информационной независимостью итераций циклов программы.

Билет 36. Методы организация параллельных вычислений при суперкомпьютерном решении сеточных задач

Топология - это механизм сопоставления процессами некоторого коммуникатора альтернативной схемы адресации. В *MPI* топологии виртуальны. Может использоваться системой для оптимизации распределения процессов по физическим процессорам.

Декартова топология - это:

- логическая топология определена многомерной решеткой;
- обобщение линейной и матричной топологий на произвольное число измерений.

Пример. Рассмотрим конкретную задачу: уравнение Лапласа на прямоугольнике.

$$\begin{cases} \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \\ x \in [0, 1], \\ u(x, 0) = \sin(\Pi x), \\ u(x, 1) = \sin(\Pi x)e^{-x}, \\ u(0, y) = u(1, y) = 0. \end{cases}$$

Рассмотрим 5-точечный шаблон:

$$u_{i,j}^{n+1} \cong \frac{u_{i+1,j}^n + u_{i-1,j}^n + u_{i,j+1}^n + u_{i,j-1}^n}{4}.$$

Создается декартова топология (механизм сопоставления процессами некоторого коммуникатора альтернативной схемы адресации; виртуальный). Необходим обмен данными на границах, будем использовать следующие функции *MPI*.

Основные функции для реализации:

1. Создание топологии: *int MPI_Cart_create*;
2. Определение декартовых координат процесса по его рангу *int MPI_Cart_coords*;
3. Определение ранга процесса по его декартовым координатам: *int MPI_Cart_rank*;

4. Разбиение решетки на подрешетки меньшей размерности: *int MPI_Cart_sub*;
5. Поддержку процедуры последовательной передачи данных по одному из измерений решетки: *int MPI_Cart_shift*;
6. Получение информации о топологии по коммуникатору и числу измерений: *MPI_Cart_get*.

Блокирующие передачи:

1. *int MPI_Send*;
2. *int MPI_Recv*.

Коллективные операции:

1. *int MPI_Bcast* - рассылка сообщения от процесса source всем процессам, включая рассылающий процесс;
2. *int MPI_Gather* - сборка данных со всех процессов группы в один массив;
3. *int MPI_AllReduce* - выполнение глобальных операций с возвратом результатов во всех процессах в одном буфере;
4. *int MPI_Reduce* - Функция аналогична предыдущей, но результат будет записан в буфер только у одного процесса.

Суперкомпьютерное моделирование турбулентных течений Рассмотрим уравнение Навье-Стокса для несжимаемой жидкости

$$\begin{cases} \operatorname{div} \vec{v} = 0 \\ \frac{\partial \vec{v}}{\partial t} + (\vec{v}, \nabla) \vec{v} + \frac{1}{\rho} \operatorname{grad} p = \nu \Delta \vec{v} \end{cases}$$

Тут над $\vec{v}$ и p пишется $\sim$, но в теке не получается сделать

Неизвестные: $\vec{v}$ - скорость течения, p - давление. После обнормировки (???) получаем систему:

$$\begin{cases} \operatorname{div} \vec{v} = 0 \\ \frac{\partial \vec{v}}{\partial t} + (\vec{v}, \nabla) \vec{v} + \operatorname{grad} p = \frac{1}{Re} \Delta \vec{v} \end{cases} \quad (1)$$

Турбулентное течение состоит из вихрей различных размеров. Крупные вихри переходят в более мелкие, те в еще более мелкие, в еще-еще более мелкие, в еще совсем более мелкие, и так масштабы течения уменьшаются, пока не дойдут до Колмогоровского масштаба, на котором кинетическая энергия движения жидкости или газа переходит в тепло, во внутреннюю энергию, за счет вязкости среды. Происходит переход кинетической энергии по каскаду от крупных вихревых структур течения к более и более мелким.

То есть, чем больше в изучаемом течении перепад масштабов, тем больше нужно сеточных ячеек. У течения есть характеристика – число Рейнольдса (Re). Оно, по сути, определяет перепад масштабов от крупных структур до самых мелких, после которых происходит диссипация в тепло. Чем больше размер объекта или чем больше скорость течения или чем меньше вязкость среды – тем больше это число. Соответственно, тем более высокое нужно пространственное и временное разрешение.

Известно, что при достаточно больших Re ($Re > Re_{crit}$, для разных областей этот Re_{crit} разный) возникают турбулентные течения.

Есть 3 основных моделей турбулентностей:

1) Прямое численное моделирование – DNS (Direct Numerical Simulation). Аппроксимируем ур-е (1) на мелкой сетке, ловящие мелкие вихри. Если использовать явные схемы, то из-за мелкой сетки - очень много времени, след-но нужен суперкомп. неявные схемы - ограничение по времени меньше, но получаем сложную и объемную СЛАУ, к-рая может не поместиться в ОЗУ, след-но нужен суперкомп.

2) Метод крупных вихрей – LES (Large Eddy Simulation).

Моделируем сетку крупнее колмогоровского масштаба и решаем на ней (1), след-но теряем мелкие вихри. После возможно восстановить мелкие вихри,

напрмиер каким-либо импирическими моделями из физ экспериментов. Тоже нужен суперкомп.

3)Осредненные уравнения – RANS (Reynolds Averaged Navier – Stokes).

Если вообще все масштабы течения загнать в модель, то получится стационарный расчет RANS. Все масштабы течения мы отдали на откуп модели и получили стационарное осредненное решение – как если бы в каждой ячейке был датчик измерений физических величин, и мы бы усредняли его показания на длительном интервале времени.

38. Использование суперкомпьютеров для решения задач молекулярного моделирования.

Специфические свойства наносистем, определяющие задачи из области:

1. Большое количество молекул в нанокластере. Моделирование **Ab initio** - без доп. предположений.
2. Нестабильность наносистем. Необходимость устойчивого расчета систем близких к границе устойчивости
3. Расчет фазовых переходов с многими границами, узких слоев, окружающих движущиеся границы фазовых переходов
4. Необходимость решения задач многомерной оптимизации потенц. энергии в условиях, когда надо искать близкие минимумы и седловые точки
5. Структура молек. соединений неоднозначна $\Rightarrow$ решения не единственны

1. Квантовомеханический уровень

1.1 Модель молекулярного переключения

Квантовомеханические вычисления *из первых принципов* - решение уравнений кв. механики, расчет межмолекулярных связей, электронной конфигурации, поверхности потенц. энергии системы.

1. Потенциал взаимодействия заряж. частиц - **Закон Кулона**. Нет подгоночных параметров.
2. **Наномасштаб**. Характерный масштаб пространства $\sim 10^{-10} m$, времени $\sim 10^{-15} c$.
3. **Метод - CPMD (Car-Parrinello Molecular Dynamics)** - мат. модель молек. динамики Кар-Перринелло на основе приближения Борна-Оппенгеймера.
Силы, действ. на ядра со стороны электронов - из уравнения Шредингера для электронов в основном состоянии. Уравнения модели **CPMD** на основе функционала плотности **DFT**:

$$M_I \frac{d^2 R_I}{dt^2}(t) = -\frac{\partial}{\partial R_I} E_{KS}[\{\psi_i\}, \{R_I\}],$$
$$\mu_i \frac{d^2 \psi_i}{dt^2}(t) = -\frac{\partial}{\partial \psi_i^*} E_{KS}[\{\psi_i\}, \{R_I\}] + \sum \Lambda_{ij} \psi_j, \quad (1)$$

где M_I, R_I - масса и координата ядра I ,

μ_i - фиктивная масса электрона,

ψ_i - орбитали Кона-Шэма, E_{KS} - энергия Кона-Шэма, $\sum \Lambda_{ij} \psi_j$ - голономные связи.

На каждом шагу считаются:

$-\frac{\partial}{\partial R_I} E_{KS}$ - силы, действ. на **ионы**,

$-\frac{\partial}{\partial \psi_i^*} E_{KS}$ - силы, действ. на **электроны**,

Λ_{ij} - множ.-ли Лагранжа

Энергия Кона-Шэма:

$$E_{KS}[\{\psi_i\}, \{R_I\}] = \sum_i^{OCC} f_i \int \psi_i^*(r) \left[-\frac{1}{2} \nabla^2 \psi_i(r) \right] dr +$$
$$+ \int U_{ext}(r) \rho(r) dr + \frac{1}{2} \int \int \frac{\rho(r) \rho(r')}{|r - r'|} dr dr' + E_{XC}[p] + E_I[\{R_I\}],$$

где U_{ext} - ионный потенциал, $\rho(r) = \sum_i f_i |\psi_i|^2$ - плотность заряда электронов, f_i - числа заселенности, E_{XC} - обменно-корреляционная энергия, E_I - электростатическое ионное взаимодействие.

Орбитали Кона-Шэма :

$$\psi_i(r) = \frac{1}{\sqrt{\Omega}} \sum_k c_i(k) e^{ikr}$$

где Ω - объем ячейки, k - вектор обратной решетки, $c_i(k)$ - Фурье-коэффициент орбитали i .

Идея - разделение масштабов движ-я быстрых электронов и медленных ядер. Реализовано в CPDM. Через движение всех ядер ищут путь реакции. Основное - вычисление барьера на поверхности свободной энергии FES. Его нужно преодолеть для реакции переключения.

Моделирование молекулярного переключения: Оптимизация геометрии молекулы (атомный состав, равновесные атомные связи минимальной энергии), затем термализация молекулы термостатом Нозе-Хувера (получ. термодинамическое равновесие).

4. Система большая ~ 25000 атомов. Нужны суперкомпьютеры и параллельные вычисления
5. Масштабируемость CPMD кодов

1.2 Вычисление профиля поверхности свободной энергии методом метадинамики

В молекулярной динамике позволяет строить поверхность свободной энергии вдоль координационной переменной, определяющее главное направление реакции. Цель - эффективный анализ поверхности свободной энергии $F(s)$ как функции s (длины связей и торсионные углы). Свободная энергия:

$$F(s) = -k_B T \log \left(\int \exp \left(-\frac{U(\{R_J\})}{k_B T} \right) d\{R_I\} \right),$$

где k_B - постоянная Больцмана, T - абсолютная температура.

Уравнение метадинамики

$$\mu s'' = -k(s - S(\{R_J\})) - \frac{\partial}{\partial s} V(s, t)$$
$$M_I R_I'' = -\nabla_{R_I} U(R_I) - \frac{\partial S(\{R_J\})}{\partial R_I} k(S(\{R_J\}) - s).$$

где $V(s, t)$ - искусственный потенциал, зависит от времени, добавляется вдоль траектории в моменты t_i :

$$V(s, t) = - \sum_{t_i < t} H_i \cdot \exp \left[-\frac{(s(t) - s^i)^2}{2W_i^2} \right]$$

. Этот потенциал - сумма отталкивающих холмов в форме гауссианов, каждый высотой H . Метадинамика s - эффективность, с которой исследуется поверхность свободной энергии.

1.3 Численная параллельная реализация CPMD на супер-ЭВМ

Наиболее затратно - параллельный расчет эл. структуры молекулы. Уравнения Кона-Шэма - необходимо решать задачу на собственные значения для волновых функций. Если N_e - число электронов, то сложность - $O(N_e^3)$. Для решения - прямое и обратное преобразование Фурье. Решение - трехмерный ряд Фурье для n -го электрона. Коэффициенты разложения - через трехмерный метод быстрого преобразования Фурье (3dFFT).

Уровни распараллеливания.

- **Крупнооблачное** - на распределенной памяти на основе MPI.
- **3dFFT-метод** - из реального в пространство Фурье
- **OpenMP-распараллеливание** - для длинных циклов, общая память на узле.
- **Taskgroups-метод** - организация процессоров в группы. Процессоры - двумерная сетка. Требуется в два раза меньше коммуникаций. Проц. группы могут быть оптимально распределены по тору Blue Gene.
- **Методы копирования** - для вычисления интегралов по траекториям. Многомерная метадинамика.

2. Методы классической молекулярной динамики со связями и методы Монте-Карло

Используется в нанобиомедицине, дизайне лекарств, протеомике (синтез/исследование белков).

В этих моделях используются вычисленные силы связи между атомами, рассчитанными на предыдущем шаге моделирования. Движения ядер - классич. законы механики, связи и силы - приближение Борна-Оппенгеймера. Записываются уравнения Ньютона для всех ядер. Силы определяются градиентами поверхности потенциальной энергии.

$$M_i \frac{d^2 R_i}{dt^2} = - \sum_j \frac{\partial U(|R_i - R_j|)}{\partial R_i}$$

Требуется найти оптимальное расположение ядер, при которой молекулярная структура будет иметь минимум потенц. энергии. Ядра движутся - рельеф поверхности потенц. энергии меняется. Поиск - с использованием собственной динамики частиц, которые представляют связанные структуры. Термическая температура - неопределенность в ответе. Это средняя кинет. энергия движения молекул, вычисляется на каждом временном шаге. Нормировкой скоростей или термостатом получают сходимость процесса к системе с целевой температурой. После достижения термодинамич. равновесия вычисляются термодинамические средние. Таким образом решается задача дискретной оптимизации энергии с использованием уравнений динамики Ньютона.

В методе М-К нет движения молекул, анализируются расположения - статистический ансамбль с заданным распределением. Температура его параметр. Для равновесия и оптимальной структуры - сотни миллионов шагов процесса М-К. Принципиальным является точность расчета равновесия системы. Для протеинов, растворов полимеров рифленая потенциальная поверхность имеет много слабо отличающихся минимумов. Можно получить неустойчивую структуру. Для систем с жесткими связями потенциальная функция хорошо определена, химические связи могут тоже быть включены в модель. Если ошибки модели порядка величин локальных энергетических барьеров, то ее использование может привести к неправильным выводам о функции проектируемого устройства. Модели оптимизации геометрии систем молекул широко используются в молекулярных конструкторах.

3. Нелинейные модели сплошной среды

В этом случае система - сплошная среда, эволюция - непрерывные макро функции. Уравнения - из общих законов сохранения. Материальные уравнения, уравнения состояния - опред. из эксперимента или моделирования. Нелинейность зависимости коэффициента диффузии от плотности приводит к нелинейным уравнениям диффузии. Эти модели глобальны (в отличие от кинетики молекул) и учитывают граничные условия. Широко используются при описании диффузии, движения границ фазовых переходов, стенок магнитных доменов в наносистемах.

Билет 39 . Архитектурные особенности графических процессоров, направленные на массивно-параллельные вычисления.

CPU - небольшое число мощных, независимых ядер, 3-х уровневый кэш

GPU - вместо ядер потоковый мультипроцессор (streaming multiprocessor, SM):

1. 32 скалярных ядра CUDA cores, ~1.5ГГц
2. 2 Warp Scheduler (warp - 32 ядра)
3. Файл регистров, 128 КБ
4. 2-х уровневый кэш
5. Текстовые юниты
6. 16 Special Function Unit - интерполяция ~ трансцендентная математика один. Точности

Много простых ядер, работающих на небольшой частоте. Небольшие кэши на GPU: 32 ядра делят 64 КБ L1 кэша, общий L2 кэш, L3 нет.

Оперативная память с высокой пропускной способностью и **высокой латентностью**, оптимизированная для коллективного доступа. Так как память с высокой латентностью, то чтобы эффективнее загружать ядра на много нитей и за счет быстрого переключения контекста перекрывать обращения одних нитей вычислениями в других имеется следующая иерархия:

Нить -> warp (32 нити) -> Block (до 32 warp) -> Grid (решетка из блоков)

Block и Grid - 3d, каждая нить, блок получают свои координаты.

На уровне нитей - SIMD - все нити выполняют одно и то же действие с разными данными.

Warp-ы выполняются независимо, называется все - SIMT.

Глобальная память

Расположение в DRAM GPU, до 4ГБ, можно выделять как с хоста, так и с устройства.

Работа оптимизирована так, чтобы отдать максимальное число данных за одно обращение.

Транзакция - загрузка из глобальной памяти пула в 128 байт, начиная с адреса, кратного 128. Взаимодействие идет через кэш. Размер кэш линии L1 - 128 байт, L2 - 32 байта, можно работать не через L1 кэш, тогда транзакции по 32 байта.

Данные хранятся по строкам, чтобы добиться выравнивания по 128 байт, можно воспользоваться CudaMallocPitch - передаем, сколько у нас будет входных данных, возвращаем с учетом padding.

Билет 40. Методы эффективной организации параллельных вычислений на графических процессорах

Коротко:

1. Утилизация латентности памяти

Цель: эффективно загружать ядра

Решение CPU: сложная иерархия кэшей

GPU: За счет наличия сотен ядер и миллионов нитей на GPU легче утилизировать всю полосу пропускания. Быстрое переключения контекста – эффективное перераспределение вычислений

2. SIMT(Single instruction, multiple thread) и масштабирование

Виртуальное: Поддержка миллионов нитей в GPU; независимость виртуальных блоков

Аппаратное: Мультипроцессоры независимы -> Можно “нарезать” GPU с разным кол-вом SM (streaming multiprocessor)

3. Асинхронность в CUDA

Чтобы GPU больше времени работало в фоновом режиме, параллельно с CPU, некоторые вызовы являются асинхронными. Отправляют команду на устройство и сразу возвращают управление хосту

4. Работа с глобальной памятью

a. Расположена в DRAM GPI

b. Объем до 4 Гб

c. Использование кэшей L1, L2

Общие принципы эффективной работы:

- Обращения нитей варпа* в память должны быть пространственно-локальными
- Начала строк матрицы должны быть выровнены
- Использование массивов структур
- Выбор режима кэша L1
- Избегать косвенной адресации
- Избегать обращений нитей варпа* к столбцу матрицы

*Варп – группа из 32 потоков и является минимальным объемом данных, обрабатываемых SIMD-способом в мультипроцессорах CUDA.

Полную версию билета см в gosu.pdf, all.pdf (билет 40) или билеты_скиподы.pdf (билет 16)