

1. Основные понятия теории динамических систем.

Опр. 1 (Динамическая система). **динамической системой** называется пара из фазового пространства P (метрическое пространство или многообразие) и однопараметрической непрерывной или дискретной группы преобразований $P \times \mathbb{R} \rightarrow P$ или $P \times \mathbb{Z} \rightarrow P$, обозначаемой как $\phi^t(x)$, $x \in P$. Для данной группы отображений должны выполняться следующие свойства:

- $\phi^0(x) = x$;
- $\phi^{t_1} \circ \phi^{t_2} = \phi^{t_1+t_2}$;
- ϕ^t дифференцируема по времени, определена и обратима для любых корректных значений t .

Опр. 2 (Траектория). **Траекторией**, проходящей через точку $x \in P$ называется множество $\{\phi^t(x) | t \in T\}$

В случае непрерывного времени динамическую систему можно задать уравнением следующего вида:

$$\dot{x} = F(x), \quad F(x) = \left. \frac{d\phi^t(x)}{dt} \right|_{t=0}.$$

2. Ляпуновский показатель. Вычисление ляпуновского показателя в случае анализа систем, заданных аналитически.

В соответствии со **2 теоремой Ляпунова** погрешность предсказания любого временного ряда меняется по следующему закону:

$$\varepsilon_t = \varepsilon_0 e^{\lambda t},$$

где ε_0 – погрешность последнего известного значения ряда, ε_t – погрешность предсказания t шагов спустя, $\lambda = \text{const}$ – **старший показатель Ляпунова**. В случае регулярных рядов $\lambda < 0$, в случае хаотических – $\lambda > 0$.

Отсюда следует, что для хаотических рядов существует **горизонт прогнозируемости** – количество шагов T , такое, что после него погрешность любого предсказания превышает допустимый порог ε_{\max} : $T \sim \frac{1}{\lambda} \ln \frac{\varepsilon_{\max}}{\varepsilon_0}$.

Пусть рассматриваемый временной ряд соответствует динамической системе $\dot{x} = f(x)$. Зафиксируем точку $x(t)$, принадлежащую какой-либо траектории. Рассмотрим малое возмущение этой траектории $x(t) + \varepsilon u(t)$, также являющееся решением уравнения $\dot{x} = f(x)$. Подставляя разность двух рассматриваемых решений в уравнение, получаем следующее соотношение (считая, что f дифференцируема):

$$\begin{aligned} \varepsilon \dot{u} &= (x(t) + \varepsilon u(t)) - f(x(t)) = \left. \frac{\partial f}{\partial x_i} \right|_{x=x(t)} \cdot \varepsilon u(t) + O(\varepsilon), \\ \dot{u}(t) &= \nabla f|_{x(t)} \cdot u(t) = A(x(t)) \cdot u(t). \end{aligned} \quad (1)$$

Уравнение (1) называется линеаризацией динамической системы на траектории $x(t)$. Очевидно, что тождественный ноль является решением линеаризованной системы.

При этом важно отметить, что, вообще говоря, x и u не лежат в одном пространстве: x является элементом фазового пространства, u лежит в касательном пространстве, построенном в точке x . Эволюция возмущения траектории позволяет определить характеристический показатель Ляпунова $\lambda = \lim_{T \rightarrow \infty} \frac{1}{T} \ln \frac{\|u(T)\|}{\|u(0)\|}$. В зависимости от начальных условий характеристический показатель может принимать только значения из Ляпуновского спектра, то есть максимум n различных условий. Если хотя бы один показатель положителен, то система является хаотической, в случае, когда положительные несколько показателей, говорят о гиперхаосе.

Если явный вид системы известен, что бывает чуть менее, чем никогда, то для нахождения спектра показателей Ляпунова можно применить алгоритм Бенеттина:

1. Рассмотрим совместно уравнение исходной динамической системы и n линеаризованных систем с ортонормированными начальными условиями. Удобно выбрать в качестве начальных условий для i -й системы вектор $e_i = (\delta_{ji})_{j=1}^n$. Полученная система в сумме состоит из $n \cdot (n + 1)$ скалярных уравнений.
2. Будем одновременно проводить численное интегрирование всех систем, используя метод Рунге-Кутты 4 порядка.

3. Также на каждом шаге вычисляем весь набор **объемных показателей Ляпунова**, пользуясь любым определением скалярного произведения:

$$\hat{\kappa}_j(t) = \ln \sqrt{\det \begin{pmatrix} (u_1, u_1) & \dots & (u_1, u_j) \\ \vdots & \ddots & \vdots \\ (u_j, u_1) & \dots & (u_j, u_j) \end{pmatrix}}$$

4. После каждого шага интегрирования будем производить ОГШ для решений линеаризованных систем, так, чтобы их решения оставались ортонормированными. При этом обязательно нужно учитывать, что на каждом шаге сначала вектора возмущений u эволюционируют во времени в соответствии с линеаризованной матрицей системы в точке $x(t)$, затем вычисляются $\hat{\kappa}_j(t)$ и только затем производится ортогонализация.
5. В пределе при $t \rightarrow \infty$ находим значения объемных показателей Ляпунова, усредняя по времени:

$$\kappa_j = \frac{1}{t} \sum_{\tau=1}^t \hat{\kappa}_j(\tau).$$

Так как $\kappa_j = \sum_{i=1}^j \lambda_i$, то разности соседних κ_j дают показатели Ляпунова.

3. Ляпуновский показатель. Вычисление ляпуновского показателя по временному ряду. Метод аналога. Фрейм-разложение.

Рассмотрим временной ряд (возможно, многомерный) x_i . Тогда для данного ряда можно построить **ряд z -векторов**, каждый элемент которого определяется следующим образом:

$$\vec{z}_i = (x_i \quad x_{i+1} \quad \dots \quad x_{i+d-1})^T,$$

$d = \text{const}$ – размерность вложения.

Построенный ряд можно использовать для оценки показателей Ляпунова. Для этого находятся пары близких z -векторов, после чего один из них рассматривается как возмущенная версия другого. Рассмотрим конкретные алгоритмы.

Метод аналога:

1. Для вектора «основной траектории» z_i находим его ближайшего соседа z_{j_1} ;
2. Обозначим разность между ними как $u_1(i+t) = z_{i+t} - z_{j_1+t}$. Для каждой такой пары (i, j_1) траектории расходятся, то есть норма вектора $u_1(t)$ экспоненциально возрастает. В момент t_1 , когда норма достигнет некоторого порога $\varepsilon_2 = \text{const}$, вектор $z_{j_1+t_1}$ заменяется на другой вектор z_{j_2} , ближайший к исходной траектории z_{i+t_1} в момент времени смены. Этот вектор также в какой-то момент t_2 заменяется на z_{j_3} и так далее.
3. Оценка старшего показателя Ляпунова вычисляется по следующей формуле:

$$\hat{\lambda}_1 = \frac{1}{t_p - t_0} \left(\ln \frac{\|u_1(t_1)\|}{\|u_1(t_0)\|} + \ln \frac{\|u_2(t_2)\|}{\|u_2(t_1)\|} + \dots + \ln \frac{\|u_p(t_p)\|}{\|u_p(t_{p-1})\|} \right), \quad (2)$$

где p – количество различных «ближайших соседей», использованных в процессе. Заметим, что если $p = 1$, то есть расхождение на расстояние ε_2 так и не произошло, то формула (2) соответствует формуле погрешности вычисления временного ряда:

$$\lambda = \frac{1}{t} \ln \frac{\varepsilon_{\max}}{\varepsilon_0} = \frac{1}{t} \ln \frac{\|u(t)\|}{\|u(0)\|}$$

Метод фрейм-разложения можно рассматривать как способ применить алгоритм Бенеттина, не зная явного вида уравнения динамической системы. Формально, для использования алгоритма Бенеттина достаточно уметь вычислять две вещи: значения z в точках $t = t_0, t_0 + \tau, t_0 + 2\tau, \dots$ и применять к любому вектору возмущения $u(t) = \phi^t(x(0) + \varepsilon u(0)) - \phi^t(x(0))$ оператор временной эволюции $Df : u(t) \mapsto u(t + \tau)$. Значения вектора z по определению представляют собой известный временной ряд, поэтому достаточно рассмотреть аппроксимацию Df на основе ряда.

Рассмотрим ближайших соседей точки z_i в пространстве z -векторов z_{k_j} и произвольный вектор возмущения u_i . Введем соответствующие разности как $y_j = z_{k_j} - z_i$. Тогда очевидно, что $\hat{y}_j = Df[y_j] = z_{k_{j+1}} - z_{i+1}$. Найдем разложение $u_i = \sum_j c_j y_j$ и будем считать, что $\widehat{Df[u]} = \hat{g} = \sum_j c_j \hat{y}_j$. Далее этот вектор можно использовать для нахождения показателей Ляпунова при помощи алгоритма Бенеттина.

В реальных расчетах может оказаться, что вектор u_i не лежит в линейной оболочке векторов y_{k_j} и точного разложения не существует, или, наоборот, разложение не единственно. Поэтому для отыскания коэффициентов c_i ищут такой вектор $g = \sum_j c_j y_j$, что $(u_i, g) = 1$, $g^2 \rightarrow \min_c$, после чего найденный вектор нормируется. В свою очередь, нахождение этого вектора сводится к решению определенной системы линейных уравнений.

4. Прогнозирование на основе кластеризации. Метод Уишарта.

Так как траектория системы прилегает к аттрактору и периодически проходит близко к самой себе в прошлом. Разумно предположить, что если начало какого-то участка траектории близко к уже известному куску, называемому мотивом, то и продолжение этого участка совпадает с мотивом.

Как правило, в качестве мотивов выбираются центроиды кластеров в пространстве z -векторов, для чего необходимо уметь их кластеризовать. Однако рассматриваемая постановка задачи омрачается тем, что неизвестно ни количество кластеров, ни даже его порядок – десятки или миллионы, что не позволяет использовать многие из существующих алгоритмов кластеризации.

Одним из алгоритмов, которые можно использовать в данной задаче, является алгоритм Уишарта (Wishart), принимающий набор точек $V_n = \{x_i\}_{i=1}^n$ и два параметра k, h . Параметр k представляет собой количество рассматриваемых ближайших соседей. Параметр h является пороговым значением, участвующим в определении того, считается ли некоторый набор точек «значимым». Более формально, класс $C \subseteq V_n$ называется значимым по высоте h , если

$$\max_{x_i, x_j \in C} \left| \frac{k}{nW(d_k(x_i))} - \frac{k}{nW(d_k(x_j))} \right| \geq h,$$

где $W(r)$ – объем гипершара радиусом r , $d_k(x_i)$ – расстояние от точки x_i до ее k -го ближайшего соседа.

Результатом работы алгоритма является набор меток кластеров w_i , где метка $w_j = 0$ обозначает межкластерный шум. Алгоритм состоит из следующих шагов:

1. Сортируем все точки по возрастанию $d_k(x_i)$. После этой операции в начале списка находятся так называемые «области сгущения» – точки, расположенные сравнительно близко к по крайней мере k другим точкам;

2. Рассматриваем список кластеров, в начале работы алгоритма не содержащий ни одного кластера. Помимо информации о том, какие точки содержит данный кластер, будем хранить метку, указывающую, является ли кластер «завершенным». В завершённые кластера новые точки не добавляются, хотя размер и номер завершённого кластера все еще может измениться за счет объединения его с другими кластерами. Кроме того, в соответствии с логикой работы алгоритма каждый завершённый кластер также является значимым по высоте h , что позволяет опустить часть проверок на значимость.
3. Рассматриваем граф G , вершины которого будут соответствовать объектам x_i . Изначально этот граф не содержит ни одной вершины; вершины добавляются по одной в соответствии с порядком, в котором отсортированы точки. Если в какой-то момент времени граф содержит n вершин, соответствующих n первым точкам, то точка x_{n+1} обрабатывается по следующим правилам:
 - (a) В граф добавляется новая вершина x_{n+1} (здесь и далее для простоты вершины графа обозначаются так же, как и исходные точки);
 - (b) Добавленная вершина соединяется со всеми вершинами $x_j, j \leq n$, для которых выполняется условие $d(x_{n+1}, x_j) \leq d_k(x_j)$, то есть таких, что x_{n+1} входит в список k ближайших соседей точки x_j (не наоборот);
 - (c) Если вершина x_{n+1} оказывается изолированной, то относим эту точку к новому кластеру;
 - (d) Если вершина x_{n+1} не изолирована и все ее соседи принадлежат одному и тому же кластеру, то относим эту точку либо к этому же кластеру, либо в шум, в зависимости от того, является ли кластер завершённым;
 - (e) Если соседи x_{n+1} принадлежат к различным кластерам, среди которых нет нулевого кластера (обозначающего межкластерный шум) и не более одного кластера является значимым, то все кластера, включающие каких-либо соседей точки x_{n+1} объединяются в один. Сама точка также относится к этому объединенному кластеру.
 - (f) В противном случае (то есть если соседи принадлежат к различным кластерам, причем или среди них есть нулевой, или количество значимых кластеров больше единицы), точка x_{n+1} добавляется к шуму, а кластера, содержащие соседние точки, либо помечаются как

завершенные (если они значимы по высоте h), либо также присоединяются к шуму.

При прогнозировании конец временного ряда сравнивается с началом каждого из обнаруженных мотивов. Если Евклидово расстояние между ними не превышает некоторого порога, то оставшаяся часть мотива, не участвовавшая в определении расстояния по очевидным причинам, дает предсказание на соответствующее число шагов вперед. Так как количество полученных предсказаний может быть большим, то их необходимо агрегировать. Как правило, рассматривается кластеризация множества предсказаний. Если предсказания образуют один плотный кластер с небольшим количеством выбросов, то в качестве окончательного предсказания выбирается центроид этого кластера. Если же обнаруживается несколько разнесенных кластеров сравнимой мощности, равномерное распределение прогнозов или какая-либо другая ситуация, отличная от целевой, то точка считается непрогнозируемой.

5. Плоскость энтропия-сложность.

Рассмотрим временной ряд x_t и соответствующий ему ряд из z -векторов размерности d :

$$\vec{z}_i = (x_i \ x_{i+1} \ \dots \ x_{i+d-1})^T.$$

Каждому такому вектору сопоставляется перестановка $\sigma \in S_d$, сортирующая этот вектор, то есть такая, что

$$z_{\sigma_0^{-1}} \leq z_{\sigma_1^{-1}} \leq \dots \leq z_{\sigma_{d-1}^{-1}}.$$

Заметим, что в случае, когда некоторые элементы вектора одинаковы, такая перестановка не единственна, но в случае непрерывного фазового пространства этой возможностью можно пренебречь. Также можно рассматривать не саму перестановку σ , а обратную к ней σ^{-1} . Такая замена приведет только к изменению порядка чисел, описывающих эмпирическое распределение, но не повлияет на конечный результат. В описанном случае перестановка σ может быть получена программно при помощи функции **argsort**, определенной во многих библиотеках.

Поскольку элементы временного ряда x_i можно рассматривать как наблюдения некоторого временного процесса, $\sigma(z)$ является дискретной случайной величиной, обладающей каким-то распределением P .

Если рассматриваемый процесс является шумом, то P – это равномерное распределение U на множестве S_d . А хаотической динамической системе отвечает некое распределение $P \neq U$.¹

Случайное распределение можно описать в терминах энтропии и неравновесности. Дадим соответствующие определения:

Опр. 3 (Информация по Шеннону). Рассмотрим произвольное распределение P и событие A . Тогда **информацией Шеннона**, соответствующей этому событию, называется значение $I_A = f\left(\frac{1}{P(A)}\right)$, где f – некоторая функция, удовлетворяющая следующим свойствам:

- Функция f возрастает;
- Если события A, B независимы, то $I_{AB} = I_A + I_B$.

Можно показать, что этим условиям удовлетворяет только логарифмическая функция с произвольным основанием. В целях нормировки общепринято использование двоичного логарифма; единицей измерения информации в таком случае является бит. Таким образом, получаем:

$$I_A = \log_2 \frac{1}{P(A)} = -\log_2 P(A).$$

Опр. 4 (Энтропия по Шеннону). **Энтропией по Шеннону** называется среднее количество информации, получаемое в результате одного наблюдения:

$$H = \mathbb{E}_P I_\omega = \sum_i I_{\omega_i} p_i = - \sum_i p_i \log p_i.$$

Известно, что энтропия принимает значения от 0 до $\log N$, где N – количество элементарных событий. Минимальное значение принимается в случае вырожденного распределения ($p_i = \delta_{ij}$), максимальное – в случае равномерного распределения. Также часто рассматривается **нормализованная энтропия**:

$$\bar{H}(P) = \frac{H(P)}{H_{\max}} = \frac{H(P)}{H(U)} = \frac{H(P)}{\log N} \in [0, 1].$$

¹Регулярной динамической системе также отвечает распределение, не являющееся равномерным, но к таким системам этот метод исследования обычно не применяется.

Опр. 5 (Сложность). Сложностью называется следующая величина:

$$S = Q[P, U] \cdot \bar{H}(P),$$

где неравновесность Q – некоторая метрика расстояния между распределениями, нормированная в диапазон от 0 до 1.

Будем использовать в качестве метрики расстояния **дивергенцию Йенсена-Шеннона**:

$$Q[P_1, P_2] = Q_0 \cdot \left(H\left(\frac{P_1 + P_2}{2}\right) - \frac{H(P_1) + H(P_2)}{2} \right),$$

$Q_0 = -2 \left(\left(\frac{N+1}{N} \right) \ln(N+1) - 2 \ln(2N) + \ln N \right)^{-1}$ – нормировочный коэффициент.

Возможные положения точки (\bar{H}, S) на плоскости «энтропия-сложность» заключены между двумя кривыми, $0 \leq H \leq 1, 0 \leq S \leq \frac{1}{2}$. При этом верхней части области соответствуют хаотические временные ряды, нижним областям – шум.

6. Инвариантная мера динамической системы.

Поскольку для хаотических систем сколь угодно малые отклонения в начальных условиях приводят к экспоненциально растущей ошибке предсказания, математический аппарат, описывающий эволюцию системы набором траекторий, оказывается не слишком полезным. Вместо этого используются понятия, взятые из теории вероятности.

Разобьем фазовое пространство X на ячейки, диаметры которых не превосходят $\varepsilon = \text{const}$. Для каждой ячейки вычислим долю времени, которое траектория системы проводит в этой ячейке за время T . Оказывается, что предельный результат при $T \rightarrow \infty$ не зависит от начальных условий и описывается некоторой вероятностной мерой на X . Эта мера называется **инвариантной мерой**, поскольку ее вид не меняется с течением времени.

Опр. 6 (Мера). **Мерой** на множестве X называется функция $\mu(A)$, определенная для некоторых подмножеств $A \subseteq X$ таким образом, что:

$$1. \forall A : \mu(A) \geq 0;$$

$$2. \forall A, B : A \cap B = \emptyset \implies \mu(A \cup B) = \mu(A) + \mu(B);$$

$$3. \mu(\emptyset) = 0. \text{ }^2$$

Множества, на которых определено значение меры, называются **измеримыми**.

Если мера всего пространства $\mu(X) = 1$, то мера называется **вероятностной**.

Наиболее известна мера Лебега, соответствующая длине, площади, объему или их многомерному аналогу.

Рассмотрим некоторое вероятностное распределение на фазовом пространстве, заданное плотностью $p(x)$. Найдем вид этого распределения $q(y)$ при смене координат, заданной отображением $y = f(x)$. Отметим, что функция $q(y)$ также описывает вероятностное распределение на этом же фазовом пространстве после перемещения каждой точки x в $f(x)$. Из соображений сохранения вероятности получаем:

$$\begin{aligned} q(y) &= \sum_{x_i: f(x_i)=y} \frac{p(x_i)}{\left| \det \left(\frac{\partial f}{\partial x} \right) \right|} = \int_X \delta(f(x) - y) p(x_i) \frac{dy}{\left| \det \left(\frac{\partial f}{\partial x} \right) \right|} = \\ &= \int_X \delta(f(x) - y) p(x_i) dx. \end{aligned} \quad (3)$$

Теперь рассмотрим хаотическую динамическую систему с дискретным временем, заданную уравнением $x_{n+1} = f(x_n)$. Если в уравнение (3) в качестве начального распределения $p(x)$ подставить инвариантную меру, то после действия отображения вид распределения не должен измениться. Таким образом, получаем **уравнение Перрона-Фробениуса** для дискретного времени:

$$p(y) = \int \delta(f(x) - y) p(x) dx. \quad (4)$$

Аналогичное уравнение для системы $\dot{x} = F(x)$ с непрерывным временем может быть получено из (4). Рассмотрим отображение $\phi^\tau(x)$, ставящее в соответствие точке x точку $\hat{x}(\tau)$, где функция $\hat{x}(t)$ – решение задачи Коши $\dot{x} = F(x)$ с начальными условиями $x(0) = x$. Очевидно, что это преобразование также не должно менять вид плотности инвариантной меры. Следовательно, подставляя в (4), получаем:

$$p(y) = \int \delta(\phi^\tau(x) - y) p(x) dx.$$

²Впрочем, это равенство следует из предыдущего.

Продифференцируем полученное уравнение по τ :³

$$0 = \int p(x) \nabla \delta(\phi^0(x) - y) \cdot d_\tau \phi^\tau(x) \Big|_{\tau=0} dx.$$

Так как $\phi^\tau(x) = \hat{x}(\tau)$, то $d_\tau \phi^\tau(x) \Big|_{\tau=0} = d_\tau \hat{x}(\tau) \Big|_{\tau=0} = \dot{x}(0) = F(x)$, $\phi^0(x) = x$. В таком случае, по свойствам дельта-функции:

$$0 = \dots = \int p(x) \nabla \delta(x - y) \cdot F(x) dx = -\nabla(p(y)F(y)).$$

Таким образом, уравнение Перрона-Фробениуса свелось к хорошо изученному уравнению движения сжимаемой жидкости.

Теор. 1 (Крылова-Боголюбова). *Если существует хотя бы одно компактное множество A , инвариантное относительно $\phi^\tau(x)$, то для системы $\dot{x} = F(x)$ существует хотя бы одна вероятностная инвариантная мера μ .*

Теор. 2 (Эргодическая теорема (не мультипликативная)). *Пусть μ – инвариантная мера динамической системы, $g : \mathbb{R}^n \rightarrow \mathbb{R}$ – непрерывно измеримая функция на фазовом пространстве. Тогда для почти всех x по мере μ предельное среднее значение $g(x)$ равно теоретическому матожиданию:*

$$\forall t : \lim_{T \rightarrow +\infty} \frac{1}{T} \int_t^{t+T} g(x(\tau)) d\tau = \int g(x) \mu(dx) = \text{const}$$

Однако необходимо учесть, что понятие «почти всюду» для меры μ может значительно отличаться от такового для меры Лебега. Например, для системы $\dot{x} = -x$ («черная дыра») носитель меры состоит из одной точки $x = 0$; для всех остальных точек равенство может не соблюдаться.

Теор. 3 (Теорема Пуанкаре о возвращении множеств). *Пусть A – измеримое множество, инвариантная мера которого больше нуля. Тогда $\exists t > 1 : \mu(A \cap \phi^t(A)) > 0$. Иными словами, траектории, начинающиеся в данном множестве, будут бесконечно много раз возвращаться в это же множество.*

Теор. 4 (Теорема Пуанкаре о возвращении траекторий). *Почти все точки x по мере μ устойчивы по Пуассону. Под устойчивостью по Пуассону понимается следующее свойство: для любой окрестности $U(x) \forall T : \exists t > T : \phi^t(x) \in U$, т.е. любая траектория бесконечно много раз возвращается в окрестность своей начальной точки.*

³В общем случае $p(y)$ может зависеть от времени, в таком случае в левой части получим $\partial_\tau p(y, t + \tau)$

7. Мультипликативная эргодическая теорема.

Рассмотрим динамическую систему с дискретным временем $x_{n+1} = f(x_n)$, в которой каждую траекторию можно сколь угодно долго продлевать по времени в обе стороны.

Теор. 5 (Мультипликативная эргодическая теорема, теорема Оселедца). Пусть для всех k , x_k фундаментальная матрица решений линеаризованной системы Φ_k , определяемая следующим образом:

$$\Phi_{k+1} = B_k \Phi_k, \quad \Phi_0 = I$$

определена и невырождена, матрица $B_k = \left(\frac{\partial f_i}{\partial x_j} \right) \Big|_{x_k}$ ограничена.

Тогда система обладает следующими свойствами:

1. Почти для всех точек (по инвариантной мере μ) выражение $\lim_{k \rightarrow \infty} \frac{1}{k} \ln \|B_k\|$ может принимать не более чем n различных значений. Эти значения называются **показателями Ляпунова** и обозначаются λ_i .
2. В каждой точке x_k касательное пространство распадается в прямую сумму подпространств $R_i(x_k)$ так, что если начальный вектор возмущения $u_0 \in R_i(x_k)$, то $\lim_{k \rightarrow \pm \infty} \frac{1}{k} \ln \|u_k\| = \pm \lambda_i$.
3. Подпространства, определенные в предыдущем пункте, инвариантны в том смысле, что $R_i(x_{k+1}) = B_k R_i(x_k)$.

8. Энтропия Колмогорова-Синая.

Опр. 7 (Энтропия Колмогорова-Синая). Рассмотрим динамическую систему с дискретным временем $x_{n+1} = f(x_n)$ или с дискретизованным непрерывным временем $x(t + \tau) = \phi^\tau(x(t))$. Разобьем фазовое пространство на непересекающиеся множества A_i , такие, что $\text{diam } A_i < \varepsilon$ по любой метрике, например, Евклидовой.

Тогда введем следующую последовательность разбиений:

$$A_{i_1 i_2 \dots i_k} = \bigcap_{j=1}^k f^{1-j} (A_{i_j}),$$

то есть отнесем точку x к множеству $A_{i_1 i_2 \dots i_k}$, если изначально она находится в множестве A_{i_1} , после одного временного шага в множестве A_{i_2} и так далее до A_{i_k} . Заметим, что некоторые из построенных множеств (или даже подавляющее большинство) могут быть пустыми.

Для каждого разбиения глубины k подсчитаем энтропию Шеннона, соответствующую инвариантной мере системы μ :

$$H(k) = - \sum_{i \in \{1, \dots, N\}^k} \mu(A_{*i}) \log \mu(A_{*i})$$

Тогда **метрической энтропией** или **энтропией Колмогорова-Синая** называется предельное приращение $H(k)$ при росте k :

$$K = \lim_{\varepsilon \rightarrow 0} \lim_{k \rightarrow \infty} (H(k+1) - H(k)) = \lim_{\varepsilon \rightarrow 0} \lim_{k \rightarrow \infty} \frac{H(k)}{k}$$

Для хаотических систем $K > 0$, что означает, что любое конечное количество информации о системе в начальный момент времени постепенно перестает быть хоть сколько-нибудь полезным в силу нарастания неопределенности. О таких системах говорят, что они производят информацию. Для регулярных динамических систем $K = 0$.

9. Теорема Такенса. Выбор параметров реконструкции.

Теор. 6 (Теорема Такенса). *Рассмотрим многообразие M^k размерности k , являющееся подмножеством фазового пространства некоторой динамической системы $y(t) = \phi^t(y(0))$. Будем считать, что известный временной ряд x_i образован значениями наблюдаемой функции состояния системы $x_i = h(y(i\tau))$, $y(t) \in M^k$. Введем еще одну функцию состояния системы $\Lambda(y) : M^k \rightarrow \mathbb{R}^m$. Тогда, если M^k , h , ϕ^t дважды дифференцируемы, $m \geq 2k + 1$ и все неподвижные точки и циклы порядка $i\tau$, $i < m$ обладают простыми отличными от 1 собственными значениями матрицы линеаризации, значения $h(y)$ для них различны, то в общем случае функция Λ задает вложение многообразия в \mathbb{R}^m . Это означает выполнение следующих свойств:*

1. Функция Λ взаимно-однозначна, обратима и обе функции Λ , Λ^{-1} дифференцируемы;

2. Для образов траекторий в \mathbb{R}^m выполнены те же свойства, что для исходных траекторий, в частности, через каждую точку поверхности $S^k = \Lambda(M^k)$ проходит только одна траектория;
3. На поверхности S_k можно ввести динамическую систему с дискретным временем, определяющую значения $z_i = \Lambda(y(i\tau)) = \Lambda \circ \phi^{i\tau}(y(0)) = \Lambda \circ \phi^t \circ \Lambda^{-1}(z_0) = \Psi(z_0)$.

Заметим, что обе динамические системы можно рассматривать как одну динамическую систему с точностью до невырожденной замены координат. Следовательно, все характеристики системы, не меняющиеся при таких заменах, в частности, весь спектр показателей Ляпунова, можно вычислять по экспериментальным данным.

В качестве стандартного выбора функции $\Lambda(y)$ рассматривается преобразование наблюдаемых значений в ряд z -векторов:

$$z_i = \begin{pmatrix} x_i \\ x_{i+1} \\ \dots \\ x_{i+m-1} \end{pmatrix} = \begin{pmatrix} h(y(i\tau)) \\ h(y((i+1)\tau)) \\ \dots \\ h(y((i+m-1)\tau)) \end{pmatrix} = \begin{pmatrix} h \\ h \circ \phi^\tau \\ \dots \\ h \circ \phi^{(m-1)\tau} \end{pmatrix} (y(i\tau)).$$

Такой выбор отображения позволяет варьировать m в достаточно широких пределах без каких-либо специальных методов, а также сводит задачу прогнозирования ряда к задаче нелинейной регрессии:

$$x_{i+1} = F(x_i, x_{i-1}, \dots, x_{i-m+1}).$$

Свойства динамической системы на поверхности S^k зависят от свойств исходной динамической системы, от наблюдаемой функции h , а также от задержки τ и размерности вложения m . Так как два последних параметра поддаются контролю, то естественно сформулировать задачу оптимального выбора этих параметров.

Для определения размерности вложения используется метод **ложных ближайших соседей**. Рассмотрим пару «близких» векторов $z_i^{(m)}, z_j^{(m)}$ в реконструкции с размерностью вложения m и соответствующие им вектора $z_i^{(m+1)}, z_j^{(m+1)}$. Если эти вектора соответствуют близким точкам многообразия M^k , то при увеличении размерности на единицу они также будут близки. В противном случае, как правило, рассматриваемая пара векторов является ложными ближайшими соседями, то есть $\|z_i^{(m)} - z_j^{(m)}\|$ мало, а $\|z_i^{(m+1)} - z_j^{(m+1)}\|$ напротив,

велико. При достижении размерности, при которой получается правильная реконструкция, количество ложных ближайших соседей резко уменьшается.

Что же касается выбора τ , то считается, что параметры m, τ являются не вполне независимыми. Определяющим фактором, судя по всему, является величина окна реконструкции $w = (m - 1)\tau$, равная временному интервалу между двумя крайними точками z -вектора.

10. Предсказательная сложность. Обучаемость и предсказуемость. Обучение в случае распределения, допускающего параметризацию.

Существует несколько подходов к формальному определению понятия сложности системы. В контексте временных рядов рассматриваются следующие подходы:

1. Колмогоровская сложность – длина наиболее короткого описания ряда. Под описанием ряда здесь понимается алгоритм, генерирующий данный ряд;
2. Предсказательная сложность – некая мера, показывающая, насколько сложно прогнозировать данный временной ряд.

Сконцентрируемся далее на предсказательной сложности. Рассмотрим временной ряд $x(t)$. Будем считать, что нам известны значения ряда x_p на промежутке $t \in (-T, 0)$ и мы пытаемся предсказать значения x_f на промежутке $t \in (0, T')$. Для краткости обозначим эти два промежутка как «прошлое» и «будущее».

Какова бы ни была система, генерирующая ряд $x(t)$, существуют распределения $P(x_p)$, $P(x_f)$, $P(x_f|x_p)$. Определим **предсказательную информацию** следующим образом:

$$J_{\text{pred}}(T, T') = \left\langle \log_2 \left(\frac{P(x_f|x_p)}{P(x_f)} \right) \right\rangle = \langle \log_2 P(x_f, x_p) \rangle - \langle \log_2 P(x_f) \rangle - \langle \log_2 P(x_p) \rangle,$$

где под $\langle \tau \rangle$ понимается усреднение величины τ по совместному распределению $P(x_p, x_f)$. Очевидно, что если будущие значения не зависят от прошлых,

то предсказательная информация равна 0, а если x_f , наоборот, однозначно определяется значениями x_p , то J_{pred} стремится к бесконечности.⁴

Если распределения ряда не зависят от времени, по крайней мере, до некоторого горизонта стационарности, то каждое слагаемое в последнем уравнении представляет собой энтропию распределения значений ряда на интервале фиксированной длины:

$$J_{\text{pred}}(T, T') = \dots = S(T) + S(T') - S(T + T'). \quad (5)$$

По аналогии с физическими процессами, энтропия $S(T)$ асимптотически линейна по времени, то есть:

$$S(T) = S_0 \cdot T + S_1(T), \quad (6)$$

где $S_0 = \text{const}$, $\frac{S_1(T)}{T} \rightarrow 0$, $S_1(T) \geq 0$. Слагаемые в (6) называются **экстенсивной** и **субэкстенсивной** составляющими энтропии соответственно.

Подставляя (6) в (5), получаем асимптотику предсказательной информации при прогнозировании на бесконечно долгий период:

$$I_{\text{pred}}(T) = \lim_{T' \rightarrow \infty} J_{\text{pred}}(T, T') = \lim_{T' \rightarrow \infty} (S_1(T) + S_1(T') - S_1(T + T')) = S_1(T).$$

Полученная величина называется **предсказательной сложностью**.

Следует отметить, что возможны только три случая асимптотического поведения $I_{\text{pred}}(T)$:

1. $\lim_{T \rightarrow \infty} I_{\text{pred}}(T) = \text{const}$, то есть наблюдение за системой в течение сколь угодно долгого времени позволяет извлечь лишь конечное количество информации о будущих значениях. Такое поведение встречается в регулярных системах, поведение которых можно точно предсказать на сколь угодно долгий срок вперед и в системах, где отсутствует зависимость от прошлого, за исключением последнего наблюденного значения (например, Марковские процессы).
2. $I_{\text{pred}}(T) \sim \log T$.
3. $I_{\text{pred}}(T) \sim T^\mu$, где $0 < \mu < 1$.

⁴Авторы данного документа, рассмотрев тривиальный случай, получили в результате $\log_2 \delta(0)$. Съезд постановил, что это значение вполне корректно считать бесконечностью.

11. Прогнозирование в моделях регрессии. Безусловное прогнозирование. Условное прогнозирование. Прогнозирование при наличии авторегрессии ошибок.

12. Временные ряды. Динамические модели. Единичные корни и коинтеграция. Автокорреляционная и частичная автокорреляционная функции.

При рассмотрении регулярных временных рядов независимые переменные (x_t) называются **экзогенными**, зависимые (y_t) – **эндогенными**.

Стандартной моделью является модель, в которой y_t линейным образом зависит от лагированных экзогенных и эндогенных переменных:

$$y_t = \delta + \sum_{i=1}^q \beta_i x_{t-i} + \sum_{j=1}^p \alpha_j y_{t-j} + \varepsilon_t = \delta + A(L)[y_t] + B(L)[x_t] + \varepsilon_t, \quad (7)$$

где L – оператор лага, A, B – некоторые многочлены от одной переменной, ε_t – случайные ошибки, удовлетворяющие условиям Гаусса-Маркова (то есть они независимы, обладают нулевым матожиданием и постоянной конечной дисперсией и некоррелированы с x_t). В случае, когда $p > 0$, то есть присутствует зависимость от лагированных эндогенных переменных, модель (7) называется **динамической**.

Опр. 8 (Стационарность). Ряд y_t называется **стационарным в узком смысле**, если для любых m, t для любого набора чисел t_1, t_2, \dots, t_m следующие распределения совпадают:

$$P(y_{t_1}, y_{t_2}, \dots, y_{t_m}) = P(y_{t_1+t}, y_{t_2+t}, \dots, y_{t_m+t}).$$

и **стационарным в широком смысле**, если его матожидание, дисперсия и ковариационная функция не зависят от времени:

$$\mathbb{E} y_t = \mu < \infty, \quad \forall y_t = \gamma = \text{const}, \quad \text{cov}(y_{t_1}, y_{t_2}) = K(t_1 - t_2).$$

Рассмотрим авторегрессионный процесс 1 порядка (AR(1)):

$$y_t = m + \phi y_{t-1} + \varepsilon_t \iff y_t = (1 - \phi L)^{-1} (m + \varepsilon_t) \quad (8)$$

Если $|\phi| < 1$, то последнее равенство также можно переписать в следующем виде:

$$\begin{aligned} y_t &= \dots = (1 + \phi L + \phi^2 L^2 + \dots) (m + \varepsilon_t) = \\ &= (1 + \phi + \phi^2 + \dots) m + \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \dots \end{aligned}$$

Можно показать, что в таком случае $\mathbb{E} y_t = \frac{m}{1 - \phi}$, $\mathbb{V} y_t = \frac{\sigma^2}{1 - \phi^2}$, $\text{cov}(y_t, y_{t-k}) = \frac{\phi^k \sigma^2}{1 - \phi^2}$, следовательно, ряд стационарен в широком смысле.

Оценку $\hat{\phi}$ можно получить при помощи метода наименьших квадратов:

$$\hat{\phi} = \frac{\sum_t y_t y_{t-1}}{\sum_s y_s^2} = \phi + \left(\frac{\sum_t y_{t-1} \varepsilon_t}{\sum_s y_{s-1}^2} \right),$$

причем в рассматриваемых предположениях эта оценка является состоятельной и асимптотически нормальной: $\sqrt{n}(\hat{\phi} - \phi) \rightarrow \mathcal{N}(0, 1 - \beta^2)$.

Однако если подставить в (8) значение $\phi = 1$, то получится процесс **случайного блуждания**, не являющийся стационарным, так как $\mathbb{V} y_t = \sigma^2 t$; тем более не являются стационарными процессами процессы с $|\phi| > 1$. В общем случае аналогичная проблема возникает, если многочлен $A(L)$ из (7) имеет единичный корень.

Кроме того, t -статистика $t = \frac{\hat{\phi} - \phi}{s_{\hat{\phi}}}$ в случае наличия единичного корня

имеет какое-то нетривиальное распределение, описанное Дики и Фуллером для нескольких различных модификаций уравнения. На этом распределении основывается **тест Дики-Фуллера** и **расширенный (augmented) тест Дики-Фуллера**, используемый для моделей авторегрессии более чем первого порядка.

Опр. 9 (Интегрируемый ряд). Рассмотрим нестационарный ряд x_t . Назовем ряд $\Delta x_t = x_t - x_{t-1}$ рядом из разностей x_t первого порядка. Разностью $k + 1$ порядка называется ряд из разностей ряда разностей k -го порядка. Исходный ряд x_t называется **интегрируемым порядком k** , $I(k)$, если он сам и его разности до $k - 1$ порядка нестационарны, а разности k -го порядка стационарны. Стационарный ряд также считается интегрируемым рядом порядка 0.

Опр. 10 (Коинтеграция). Пусть два ряда x_t, y_t таковы, что некоторая их линейная комбинация $y_t - \beta x_t$ стационарна. Тогда эти ряды называются **коинтегрированными**.

В таком случае можно получить состоятельную оценку для β , применив МНК к уравнению

$$y_t = \alpha + \beta x_t + \varepsilon_t.$$

Однако в отличие от обычной оценки распределение $\sqrt{n}(\hat{\beta} - \beta)$ не является нормальным.

Опр. 11 (Автокорреляционная функция). Автокорреляционной функцией (ACF) стационарного ряда y_t называется коэффициент корреляции между y_t и y_{t-k} в зависимости от k :

$$\rho_k = \frac{\text{cov}(y_t, y_{t-k})}{\text{var } y_t}.$$

Опр. 12 (Частичная автокорреляционная функция). Частичной автокорреляционной функцией (PACF) стационарного ряда y_t называется коэффициент корреляции между y_t и y_{t-k} в зависимости от k при исключении влияния промежуточных значений $y_{t-1}, \dots, y_{t-k+1}$. Более формально, значение $\text{PACF}(k)$ вычисляется как коэффициент β_k в МНК-решении следующего регрессионного уравнения:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_k y_{t-k} + \varepsilon_t$$

13. Вопрос по выбору.

См. выше.

14. ARIMA-модели. Идентификация ARIMA-моделей. Способы определения параметров моделей. Прогнозирование в ARIMA-моделях.

Опр. 13. Моделью авторегрессии и скользящего среднего ARMA (p, q) называется класс моделей стационарных временных рядов, задаваемый уравнением следующего вида:

$$y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q},$$

или, в записи с использованием оператора лага,

$$\Phi(L)y_t = \delta + \Theta(L)\varepsilon_t,$$

где ε_t – ошибки, удовлетворяющие условиям Гаусса-Маркова.

Рассмотрим нестационарный временной ряд y_t . Пусть этот ряд является интегрируемым порядка d , то есть ряд из разностей d -го порядка стационарен, и ряд из разностей удовлетворяет модели ARMA (p, q) . Тогда исходный ряд y_t называется **интегрированным процессом авторегрессии и скользящего среднего** ARIMA (p, d, q) . Понятно, что модель для ряда y_t легко строится на основе модели для ряда $\Delta^d y_t$, поэтому задача сводится к предыдущей.

Методология Бокса-Дженкинса подбора параметров модели включает три этапа.

Идентификация модели. На первом шаге подбирается параметр d . Последовательно берутся разности порядка d , где d меняется от нуля до какой-либо границы. В качестве параметра модели принимается наименьшее значение, при котором ряд разностей оказывается стационарным. Стационарность ряда проверяют, применяя к ряду и графикам ACF и PACF метод пристального взгляда или тест Дики-Фуллера. На практике обычно d не превышает 2.

Для полученного стационарного ряда строятся графики выборочных ACF, PACF. По количеству точек этих функций, статистически значимо отличных от нуля получаем несколько гипотез о параметрах p, q .

Выбор параметров модели Для оценки параметров модели применяется метод наименьших квадратов (в том числе нелинейный), полный или условный метод наибольшего правдоподобия.

Проверка адекватности модели данным Для проверки адекватности модели имеющимся данным используются несколько методов. Во-первых, оценки коэффициентов должны статистически значимо отличаться от нуля. Во-вторых, следует рассмотреть остатки регрессии $e_t = y_t - \hat{y}_t$, где \hat{y}_t – оценка соответствующего значения ряда при помощи найденных параметров регрессии. Если модель соответствует данным, то остатки также являются белым шумом, и выборочные автокорреляции остатков распределены как $\mathcal{N}\left(0, \frac{1}{n}\right)$, то есть близки к нулю.

Если ничто из описанного выше не помогло окончательно определиться с выбором, применяются различные эмпирические критерии, например, **информационный критерий Акаике**:

$$\text{AIC} = 2\frac{p+q}{n} + \ln \left(\frac{\sum_t e_t^2}{n} \right) \rightarrow \min$$

15. Временные ряды. Свойства AR(1)-процесса. Свойства автокорреляционного процесса второго порядка AR(2). Свойства процессов скользящего среднего.

Рассмотрим наиболее распространенные частные случаи моделей класса ARMA (p, q) .

Авторегрессионный процесс 1 порядка $\text{AR}(1) = \text{ARMA}(1, 0)$ Авторегрессионный процесс 1 порядка задается следующим уравнением:

$$y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t$$

Как будет показано ранее, этот процесс является стационарным при $|\phi_1| < 1$. В таком случае временной ряд обладает следующими свойствами:

$$\mathbb{E} y_t = \frac{\delta}{1 - \phi_1}, \quad \mathbb{V} y_t = \frac{\mathbb{V} \varepsilon = \sigma^2}{1 - \phi_1^2}, \quad \rho_k = \frac{\text{cov}(y_t, y_{t-k})}{\mathbb{V} y_t} = \phi_1^k$$

Частичная автокорреляционная функция процесса равна 0, за исключением начальных значений $\text{PACF}(0) = 1$, $\text{PACF}(1) = \text{ACF}(1) = \phi_1$. Аналогичное свойство выполняется для всех моделей авторегрессионного процесса $\text{PACF}_{\text{AR}(p)}(k) = 0, k > p$.

Авторегрессионный процесс 2 порядка $\text{AR}(2) = \text{ARMA}(2, 0)$ Будем считать, что $\delta = 0$, в таком случае матожидание ряда равно нулю, и модель задается следующим уравнением:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \tag{9}$$

Вычислим значения ковариационной функции при $k > 0$:

$$\begin{aligned}\gamma_k &= \text{cov}(y_t, y_{t-k}) = \text{cov}(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, y_{t-k}) = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} \\ \rho_k &= \frac{\gamma_k}{\gamma_0} = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}\end{aligned}\quad (10)$$

Так как из свойств автокорреляционной функции $\rho_0 = 1, \rho_{-1} = \rho_1$, то, подставляя в уравнение (10) значение $k = 1$, получаем:

$$\rho_1 = \frac{\phi_1}{1 - \phi_2} \implies \rho_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2 \implies \dots$$

Значение дисперсии найдем, умножив обе части уравнения (9) на y_t и взяв математическое ожидание:

$$\begin{cases} \gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma^2, \\ \gamma_0 \rho_1 = \gamma_1, \\ \gamma_0 \rho_2 = \gamma_2. \end{cases}$$

Решив полученную систему уравнений, находим

$$\mathbb{V} y_t = \gamma_0 = \frac{(1 - \phi_2) \sigma^2}{(1 + \phi_2)(1 - \phi_1 - \phi_2)(1 + \phi_1 - \phi_2)}$$

Отсюда получаем, что процесс стационарен, если

$$|\phi_2| < 1, \quad \phi_1 + \phi_2 < 1, \quad \phi_2 - \phi_1 < 1.$$

Заметим, что эти же условия можно вывести, потребовав, чтобы корни многочлена $1 - \phi_1 L - \phi_2 L^2$ лежали вне единичного круга.

Процесс скользящего среднего $\text{MA}(q) = \text{ARMA}(0, q)$ Процесс скользящего среднего задается уравнением вида:

$$y_t = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} = \delta + \Theta(L) \varepsilon_t.$$

Вычислим характеристики процесса, для удобства записи обозначив $\theta_0 = -1$:

$$\mathbb{E} y_t = \delta,$$

$$\mathbb{V} y_t = \mathbb{V} \delta + \mathbb{V} \varepsilon_t + \theta_1^2 \mathbb{V} \varepsilon_{t-1} + \dots + \theta_q^2 \mathbb{V} \varepsilon_{t-q} = \sigma^2 \cdot \left(1 + \sum_{i=1}^q \theta_i^2 \right),$$

$$\gamma_k = \sum_{i=0}^q \sum_{j=0}^q \theta_i \theta_j \text{cov}(\varepsilon_{t-i}, \varepsilon_{t-k-j}) = \sum_{i=0}^q \sum_{j=0}^q \theta_i \theta_j \sigma^2 \delta_{t-i, t-k-j}.$$

Видно, что ожидание, дисперсия и ковариации процесса не зависят от времени, то есть при любых параметрах процесс стационарен в широком смысле. Кроме того, при $k > q$ ковариация γ_k , а значит, и $\text{ACF}(k)$, равны нулю, а частичная автокорреляционная функция экспоненциально убывает.

Рассмотрим подробнее процесс $\text{MA}(1)$:

$$y_t = \delta + \Theta(L) \varepsilon_t, \Theta(L) = 1 - \theta_1 L$$

Если $|\theta_1| < 1$, то оператор $\Theta(L)$ обратим, и процесс можно представить в виде авторегрессионного процесса $\text{AR}(\infty)$:

$$\begin{aligned} y_t &= \delta + \Theta(L) \varepsilon_t, \\ \Theta(L)^{-1} y_t &= \Theta(L)^{-1} \delta + \varepsilon_t, \\ y_t + \theta_1 y_{t-1} + \theta_1^2 y_{t-2} + \dots &= \frac{\delta}{1 - \theta_1} + \varepsilon_t, \\ y_t &= \frac{\delta}{1 - \theta_1} - \theta_1 y_{t-1} - \theta_1^2 y_{t-2} - \dots + \varepsilon_t. \end{aligned}$$