







Конспект лекций и задания по курсу “Статистика случайных процессов”

1	Общие определения и обозначения 🌱	3
2	Список литературы 🌻	4
I	Теория	6
3	Марковские цепи 🌻	6
4	Скрытые марковские цепи 🍂	7
4.1	Наиболее вероятная траектория скрытого процесса 🪑	8
5	Выборочные статистики стационарных процессов 🌲	9
5.1	Оценивание математического ожидания стационарного процесса 🖋️	11
5.2	Оценивание автоковариации стационарного процесса 🚗	12
6	Процессы авторегрессии скользящего среднего 🌴	14
6.1	Представление через операторы сдвига и условие стационарности ⚖️	15
6.2	Визуальная оценка значений p и q на основе анализа графиков автокорреляции и частной автокорреляции 🪓	16
6.3	Оценивание $\varphi, \theta, \sigma_\varepsilon^2, a_0$ методом условного максимального правдоподобия 🪡	18
6.4	Подгонка p и q методом Bayesian Information Criterion (BIC) 🩺	20
6.5	Подгонка p и q на основе проверки гипотезы отсутствия автокорреляции в остатках модели 🪓	21
6.6	Процесс авторегрессии проинтегрированного скользящего среднего 🧪	22
6.7	Оценивание параметра d модели $ARIMA(p, d, q)$ 🗡️	23
6.8	Прогнозирование процесса АРСС 🔪	24
6.9	Прогнозирование процесса АРПСС 🛡️	26
7	Процесс Пуассона 🍀	27
8	Винеровский процесс 🌹	28
II	Вспомогательные темы	29
9	Общие сведения о работе с данными в Python 🌿	29
9.1	Краткий обзор numpy 📡	30
9.2	Краткий обзор matplotlib.pyplot 💊	31
9.3	Краткий обзор scipy 🔗	32
10	Ошибка оценивания и ошибка прогноза при неслучайном параметре 🌿	32
10.1	Метод моделирования реализаций оценки $\hat{\theta}$ при фиксированном θ 📡	32
10.2	Метод моделирования предполагаемых значений θ относительно известной оценки $\hat{\theta}$ 🔪	33
10.3	Метод моделирования предполагаемых значений θ исходя из функции правдоподобия 🏹	34
11	Монте-Карло на марковских цепях. Метод Метрополиса-Гастингса 🌱	34
III	Задания	38

12	Задание 1		38
13	Задание 2		38
14	Задание 3		39
15	Задание 4		40
16	Задание 5		41
17	Задание 6		42

1 Общие определения и обозначения

⟨1⟩ **Обозначение.** Через $(\Omega, \mathcal{A}, \mathbf{P})$ обозначим вероятностное пространство. \square

⟨2⟩ **Обозначение.** Греческие буквы (например, ξ) обозначают случайные процессы. Через T (или T_ξ) обозначим множество возможных значений t , на которых определён случайный процесс. Выражение вида $\xi(t)$ обозначает значение случайного процесса в момент времени t . \square

⟨3⟩ **Обозначение.** Большие латинские буквы (например, X) обозначают случайные величины или наборы случайных величин. Например, X_i (или $X_i^{(n_i)}$) может обозначать i -ую выборочную траекторию процесса ξ с наблюдениями $X_{i1} = \xi(\tau_{i1}), \dots, X_{in_i} = \xi(\tau_{in_i})$ в моменты времени $\tau_{i1}, \dots, \tau_{in_i}$. \square

⟨4⟩ **Обозначение.** Обозначим через $\text{dom } X$ — множество возможных значений случайной величины X . Обозначим через $\text{dom } \xi$ — множество возможных значений случайного процесса ξ так, что $\text{dom } \xi = \cup_{t \in T} \text{dom } \xi(t)$. \square

⟨5⟩ **Обозначение.** Выборку из m независимых траекторий $\xi(t)$ обозначаем как $D = \{X_1, \dots, X_m\}$ или $D^{(m)}$. Каждая из выборочных траекторий X_i (или $X_i^{(n_i)}$) может состоять из различного количества наблюдений в разный набор моментов времени:

$$X_i = (X_i(1), \dots, X_i(n_i)), \quad X_i(j) = \xi^{(i)}(\tau_{ij}), \quad \tau_{ij} \in T,$$

где $\xi^{(i)}$ — i -ая независимая копия ξ ; $\{\tau_{ij}\}$ — набор моментов времени.

Набор $(\tau_{i1}, \dots, \tau_{in_i})$ указывает на все моменты времени, для которых присутствуют наблюдения в выборочной траектории X_i . Будем полагать, что этот набор всегда упорядочен по возрастанию, т.е.:

$$\tau_{i1} \leq \dots \leq \tau_{in_i}.$$

В том случае, если рассматривается лишь одна выборочная траектория, будем её обозначать как $X = \{X(1), \dots, X(n)\}$, где $X(i) = \xi(\tau_i)$. \square

⟨6⟩ **Обозначение.** Через \mathcal{D} обозначаем случайную выборку независимых выборочных траекторий. Данная в задаче выборка D считается реализацией \mathcal{D} . Обычно будем считать, что количество траекторий, моменты наблюдений и неслучайные параметры модели у \mathcal{D} такие же, как у D . \square

⟨7⟩ **Обозначение.** Для множества A через $|A|$ обозначим количество элементов в A . \square

⟨8⟩ **Обозначение.** Для краткости записи при перечислениях мы будем неявно опускать те значения переменных перечисления, при которых выражение не имеет смысла / выходит за границы определения. Например, если $T = \{0, 1, \dots\}$, то в записи

$$\xi(t) = \xi(t-1) + \xi(t-2), \quad \forall t \in T,$$

неявно подразумевается, что значения $t \in \{0, 1\}$ не рассматриваются. \square

⟨9⟩ Обозначение. \mathbf{F} обозначает функцию распределения, а \mathbf{f} — функцию плотности. □

⟨10⟩ Обозначение. Для дискретной случайной величины X под функцией плотности $\mathbf{f}(x)$ мы будем понимать функцию вероятности $\mathbf{P}(X = x)$. По другому её можно понимать как функцию плотности относительно считающей меры. □

⟨11⟩ Обозначение. Если Y — непрерывная случайная величина, то, формально, $\mathbf{P}(X \in A | Y = y)$ не определена. Эту проблему разрешают, воспринимая такую условную вероятность как, например, предел условных вероятностей вида

$$\mathbf{P}(X \in A | Y = y) = \lim_{\Delta \rightarrow 0+} \mathbf{P}(X \in A | Y \in (y - \Delta; y + \Delta)).$$
□

⟨12⟩ Обозначение. $(X | Y = y) \sim \mathcal{F}(y)$ обозначает, что

$$\mathbf{F}_{X | Y}(x | Y = y) = \mathbf{P}(X \leq x | Y = y) = \mathbf{F}_{\mathcal{F}}(x; y),$$

где $\mathbf{F}_{\mathcal{F}}(x; y)$ — функция распределения закона распределения $\mathcal{F}(y)$. Если X — непрерывная с.в., то через $\mathbf{f}_{X|Y}(x | y)$ обозначается условная функция плотности, которая порождается $\mathbf{F}_{X | Y}(x | Y = y)$. □

⟨13⟩ Обозначение. Для краткости записи $\mathbf{f}_{X|Y}(x | y)$ будем также обозначать как $\mathbf{f}(X | Y)$ или $\mathbf{f}(X = x | Y = y)$. □

⟨14⟩ Обозначение. Под условной независимостью X_1, \dots, X_k при фиксированном значении Y будем понимать выполнение для всех x_1, \dots, x_k, y равенства

$$\mathbf{f}_{X_1, \dots, X_k | Y}(x_1, \dots, x_k | y) = \prod_{i=1}^k \mathbf{f}_{X_i | Y}(x_i | y).$$
□

2 Список литературы

[1] Barber, David. **Bayesian reasoning and machine learning**. Cambridge University Press, 2012.

https://github.com/snowdj/CS228_PGM/blob/master/books/Bayesian%20Reasoning%20and%20Machine%20Learning%20by%20David%20Barber.pdf

- 1) Марковские цепи \implies раздел 23.1
- 2) Скрытые марковские цепи \implies раздел 23.2
- 3) Монте-Карло на марковских цепях и алгоритм Метрополиса-Гастингса \implies раздел 27.4

[2] Duke University — **Computational Statistics in Python — Markov Chain Monte Carlo (MCMC)**

<https://people.duke.edu/~ccc14/sta-663/mcmc.html>

Небольшое руководство по методам Монте-Карло на марковских цепях — краткие теоретические описания и обоснования, примеры кода реализации на Python.

- [3] Duke University — презентация “**Bootstrap Confidence Intervals**”
<http://www2.stat.duke.edu/~banks/111-lectures.dir/lect13.pdf>
Введение в метод bootstrap и “pivot confidence interval”.
- [4] University of Michigan — **HOWTO estimate parameter-errors using Monte Carlo — an example with Python**
<http://www-personal.umd.umich.edu/~wiclarks/AstroLab/HOWTOs/NotebookStuff/MonteCarloHOWTO.html>
Оценивание разброса оценок методом Монте-Карло с примерами кода и вычисления на Python.
- [5] DataLearning.ru — курс “Статистическая обработка временных рядов” — презентация “**Стационарные и эргодические случайные процессы**”.
<http://datalearning.ru/study/Courses/ssp/lections/lection02.pdf>
Понятия и критерии эргодичности по среднему и по автоковариации \implies слайды 55–62
- [6] Симушкин С.В. **Многомерный статистический анализ. Часть 2.**
<https://kpfu.ru/portal/docs/F1218776922/Simushkin.S.V..Mnogomernyj.statisticheskij.analiz..Chast.2.pdf>
Асимптотическое распределение выборочного коэффициента корреляции \implies Теорема I.5.
- [7] Dufour J.M. **Estimation of ARMA models by maximum likelihood.** McGill University, Fevereiro de. — 1981.
https://jeanmariedufour.github.io/ResE/Dufour_2008_C_TS_ARIMA_Estimation.pdf
Методы максимального правдоподобия и условного максимального правдоподобия для оценивания параметров АРСС модели
- [8] The Book of Statistical Proofs — **Proof: Derivation of the Bayesian information criterion.**
<https://statproofbook.github.io/P/bic-der.html>
- [9] UW Homepage — **Unit Root Tests.**
<https://faculty.washington.edu/ezivot/econ584/notes/unitroot.pdf>

Часть I

Теория

3 Марковские цепи

⟨1⟩ **Определение.** Однородная марковская цепь $\xi(t)$ определяется как:

- 1) Процесс с дискретным временем и начальным моментом времени. Для определённости будем полагать, что $T = \{0, 1, 2, \dots\}$.
- 2) Множество возможных значений $V = \text{dom } \xi$ — конечное множество. Для определённости будем полагать, что $V = \{1, 2, \dots, q\}$.
- 3) Выполняется марковское свойство:

$$\mathbf{f}(\xi(t) = x \mid \xi(t-1) = x_1, \xi(t-2) = x_2, \dots) = \mathbf{f}(\xi(t) = x \mid \xi(t-1) = x_1) \quad \forall t \in T.$$

- 4) Выполняется свойство однородности:

$$\mathbf{f}(\xi(t) = x \mid \xi(t-1) = x_1) = \mathbf{f}(\xi(l) = x \mid \xi(l-1) = x_1) \quad \forall t, l \in T.$$

□

⟨2⟩ **Обозначение.** $\xi \sim MC(\pi, P)$.

□

⟨3⟩ **Замечание.** В общем случае марковская цепь может быть и не однородной, т.е. свойство 4) может не выполняться. Тогда для каждого $t \in T$ переходные вероятности $\mathbf{f}(\xi(t) = x \mid \xi(t-1) = y)$ могут задаваться по-разному. Мы будем рассматривать только однородные марковские цепи.

□

⟨4⟩ **Замечание.** В общем случае в качестве элементов V могут выступать любые (не обязательно числовые) объекты, например $V = \{ \text{Солнечно}, \text{Пасмурно}, \text{Дождь}, \text{Гроза}, \text{Ураган}, \text{Снег} \}$.

□

⟨5⟩ **Определение.** Для задания всех вероятностных свойств однородной марковской цепи $\xi(t)$ достаточно указать:

- 1) Вектор начального распределения π :

$$\pi = \{\pi_1, \dots, \pi_q\}, \quad \pi_i = \mathbf{f}(\xi(0) = i).$$

- 2) Матрицу переходных вероятностей P :

$$P = \{p_{ij} : 1 \leq i, j \leq q\}, \quad p_{ij} = \mathbf{f}(\xi(t) = j \mid \xi(t-1) = i).$$

□

⟨6⟩ В однородных марковских цепях, обычно, в качестве оцениваемых параметров выступают $\{\pi_i\}$ и $\{p_{ij}\}$. Можно получить их оценку в явном виде по выборке из m независимых траекторий $D = \{X_1, \dots, X_m\}$ методом максимального правдоподобия. При выводе можно воспользоваться, например, методом неопределённых множителей Лагранжа.

4 Скрытые марковские цепи

⟨1⟩ Определение. Пара случайных процессов с дискретным временем (ξ, ζ) называется скрытой марковской цепью, если:

- 1) $\zeta \sim MC(\pi, P)$ — однородная марковская цепь.
- 2) $(\xi(t) \mid \zeta(t) = y) \sim \mathcal{F}(y)$.
- 3) $\xi(t)$ попарно условно независима от $\zeta(l)$ и $\xi(l)$ для $l \neq t$ при фиксированном значении $\zeta(t)$.

Будем считать, что $T = \{0, 1, \dots\}$, $\text{dom } \zeta = \{1, \dots, q\}$. □

⟨2⟩ Замечание. Пусть рассматривается модель, состоящая из набора случайных величин X_1, \dots, X_k . Если совместное распределение модели описывается

- 1) введением набора отношений условной независимости между элементами X_1, \dots, X_k ;
- 2) описанием распределения части элементов X_i посредством задания условного распределения при фиксированных значениях других элементов;

то отношения условных независимостей и “направление” условных распределений часто описывают в виде графов. Такие модели иногда называют графовыми вероятностными моделями. Например, скрытые марковские цепи представляют как в рис. 1. □

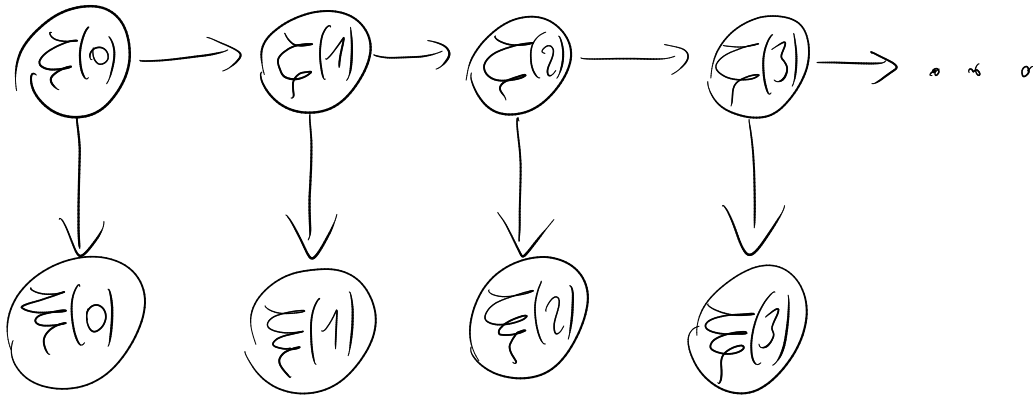


Рис. 1: Представление модели скрытой марковской цепи в виде графа.

⟨3⟩ Обычно в эксперименте наблюдаются только значения процесса $\xi(t)$, а целью исследования является предугадывание по ним значений $\zeta(t)$. В связи с этим $\zeta(t)$ называют скрытым (латентным) процессом, а $\xi(t)$ — видимым.

⟨4⟩ $\xi(t)$ можно интерпретировать как результат случайного зашумления / искажения (emission) значения $\zeta(t)$, поэтому задачу оценивания значений $\xi(t)$ по $\zeta(t)$ также называют задачей фильтрации.

⟨5⟩ Отдельно отметим, что распределение $\xi(t)$ может быть непрерывным. В этом отношении можно сказать, что скрытые марковские цепи позволяют пользоваться марковскими цепями при обработке непрерывных сигналов.

⟨6⟩ На данный момент не существует эффективного алгоритма оценивания параметров методом максимального правдоподобия по значениям $\xi(0), \dots, \xi(n)$. Возможно приближённое оценивание (с точностью до локального максимума) методами типа ЕМ-алгоритма.

⟨7⟩ Если параметры модели π , P и $\mathcal{F}(y)$ известны, то марковское свойство позволяет вычислительно легко находить характеристики условного распределения $\{\zeta(0), \zeta(1), \dots, \zeta(n)\}$ при известных $\{\xi(0), \xi(1), \dots, \xi(n)\}$.

4.1 Наиболее вероятная траектория скрытого процесса

⟨8⟩ Решается задача нахождения

$$\arg \max_{y_0, \dots, y_n} \mathbf{f}(\zeta(0) = y_0, \dots, \zeta(n) = y_n \mid \xi(0), \dots, \xi(n)),$$

при известных параметрах модели.

⟨9⟩ **Теорема.** Верно представление:

$$\max_{y_0, \dots, y_n} \mathbf{f}(\zeta(0) = y_0, \dots, \zeta(n) = y_n, \xi(0), \dots, \xi(n)) = \mu_0$$

где

$$\begin{aligned} \mu_n(y_{n-1}) &= \max_{y_n} \mathbf{f}(\xi(n) \mid \zeta(n) = y_n) \mathbf{f}(\zeta(n) = y_n \mid \zeta(n-1) = y_{n-1}); \\ \mu_t(y_{t-1}) &= \max_{y_t} \mathbf{f}(\xi(t) \mid \zeta(t) = y_t) \mathbf{f}(\zeta(t) = y_t \mid \zeta(t-1) = y_{t-1}) \mu_{t+1}(y_t), \\ t &= n-1, \dots, 1; \\ \mu_0 &= \max_{y_0} \mathbf{f}(\xi(0) \mid \zeta(0) = y_0) \mathbf{f}(\zeta(0) = y_0) \mu_1(y_0). \end{aligned}$$

□

Доказательство. Представим совместную функцию плотности в виде

$$\begin{aligned} \mathbf{f}(\zeta(0) = y_0, \dots, \zeta(n) = y_n, \xi(0), \dots, \xi(n)) &= \\ &= \mathbf{f}(\xi(0) \mid \zeta(0) = y_0) \mathbf{f}(\zeta(0) = y_0) \prod_{t=1}^n \mathbf{f}(\xi(t) \mid \zeta(t) = y_t) \mathbf{f}(\zeta(t) = y_t \mid \zeta(t-1) = y_{t-1}), \end{aligned}$$

Тогда

$$\begin{aligned} \max_{y_0, \dots, y_n} \mathbf{f}(\zeta(0) = y_0, \dots, \zeta(n) = y_n, \xi(1), \dots, \xi(n)) &= \\ &= \max_{y_0, \dots, y_n} \mathbf{f}(\xi(0) \mid \zeta(0) = y_0) \mathbf{f}(\zeta(0) = y_0) \cdot \\ &\quad \cdot \prod_{t=1}^n \mathbf{f}(\xi(t) \mid \zeta(t) = y_t) \mathbf{f}(\zeta(t) = y_t \mid \zeta(t-1) = y_{t-1}) \\ &= \max_{y_0} \mathbf{f}(\xi(0) \mid \zeta(0) = y_0) \mathbf{f}(\zeta(0) = y_0) \cdot \\ &\quad \cdot \max_{y_1} \mathbf{f}(\xi(1) \mid \zeta(1) = y_1) \mathbf{f}(\zeta(1) = y_1 \mid \zeta(0) = y_0) \cdot \\ &\quad \dots \end{aligned} \tag{10}$$

$$\cdot \max_{y_n} \mathbf{f}(\xi(n) \mid \zeta(n) = y_n) \mathbf{f}(\zeta(n) = y_n \mid \zeta(n-1) = y_{n-1}).$$

Можно увидеть, что выражения вида \max_{y_t} соответствуют $\mu_t(y_{t-1})$. Заменяв их соответствующим образом, получим доказываемое представление. \square

⟨11⟩ Исходя из представления в утверждении 9 имеем вычислительно-эффективный (с линейной сложностью по n) алгоритм нахождения наиболее вероятной траектории скрытого процесса. Пусть

- 1) y_0^* — значение y_0 , при котором достигается максимум в определении μ_0 ;
- 2) y_t^* , $t = 1, \dots, n$ — значение y_t , при котором достигается максимум в определении $\mu_t(y_{t-1}^*)$.

Тогда

$$\begin{aligned} \arg \max_{y_0, \dots, y_n} \mathbf{f}(\zeta(0) = y_0, \dots, \zeta(n) = y_n \mid \xi(0), \dots, \xi(n)) &= \\ &= \arg \max_{y_0, \dots, y_n} \frac{\mathbf{f}(\zeta(0) = y_0, \dots, \zeta(n) = y_n, \xi(0), \dots, \xi(n))}{\mathbf{f}(\xi(0), \dots, \xi(n))} = \\ &= \arg \max_{y_0, \dots, y_n} \mathbf{f}(\zeta(0) = y_0, \dots, \zeta(n) = y_n, \xi(0), \dots, \xi(n)) = \\ &= (y_0^*, \dots, y_n^*). \end{aligned}$$

5 Выборочные статистики стационарных процессов

⟨1⟩ Определение. Случайный процесс ξ называется стационарным в широком смысле (weakly stationary process), если существуют $\mathbf{E} \xi(t)$ и $\text{cov}(\xi(t), \xi(l))$ и

$$\mathbf{E} \xi(t) = \mathbf{E} \xi(t+h), \quad \text{cov}(\xi(t), \xi(l)) = \text{cov}(\xi(t+h), \xi(l+h)) \quad \forall t, l, h.$$

Везде в этом разделе под стационарным процессом будет пониматься стационарный в широком смысле процесс. \square

⟨2⟩ Определение. Пусть ξ — стационарный процесс. Автоковариационной функцией ξ называют

$$\gamma(h) = \text{cov}(\xi(t), \xi(t+h)).$$

\square

⟨3⟩ Свойства.

- 1) $\gamma(0) = \mathbf{D} \xi(t)$ и это максимальное значение γ .
- 2) $\gamma(-h) = \gamma(h)$.
- 3) $|\gamma(h)| \leq \gamma(0)$.
- 4) $\mathbf{D}(\xi(t+h) \pm \xi(t)) = 2(\gamma(0) \pm \gamma(h))$.
- 5) Если $\gamma(h_0) = \gamma(0)$ для $h_0 > 0$ и если у ξ достаточно “предсказуемые” траектории (например, непрерывные или непрерывные слева/справа), то у ξ п.н. периодические траектории с периодом h_0 .

- 6) Если $\gamma(h_0) = -\gamma(0)$ для $h_0 > 0$ и если у ξ достаточно “предсказуемые” траектории (например, непрерывные или непрерывные слева/справа), то у ξ п.н. периодические траектории с периодом $2 h_0$.
- 7) Если у ξ п.н. периодичные траектории с периодом l , то $\gamma(h)$ — периодичная функция с периодом l .

□

⟨4⟩ График $\gamma(h)$, обычно, исследуют с целью выявления периодичности в поведении случайных процессов и чтобы выявить, насколько $\xi(t)$ и $\xi(t+h)$ взаимозависимы как функцию $|h|$. Из свойств ковариации проистекает, что $\gamma(h_0) = \gamma(0)$ означает, что п.н. $\xi(t+h_0) = a\xi(t) + b$ для любого фиксированного t . $\gamma(h) = 0$ рассматривают как нечто близкое на независимость; однако действительную независимость это равенство означает только в частных случаях, например, если известно, что ξ — гауссовский процесс.

⟨5⟩ Если зависимость значений $\xi(t)$ и $\xi(t+h)$ стационарного процесса достаточно быстро убывает с ростом h , то выборка значений $\xi(\tau_1), \dots, \xi(\tau_n)$ (где между всеми τ_i и τ_j достаточно большое расстояние) по своим свойствам близка к выборке независимых наблюдений $\xi^{(1)}(t_0), \dots, \xi^{(n)}(t_0)$, где $\xi^{(i)}$ — независимые копии ξ . Тогда возможно оценивать характеристики ξ по выборке из одной траектории используя такие же оценки как для независимых выборок.

⟨6⟩ **Определение.** Явление, когда по значениям случайного процесса ξ за большой интервал времени $[t_1, t_2]$ можно судить о распределении значений набора случайных процессов $\xi^{(1)}(t), \dots, \xi^{(n)}(t)$ в любой фиксированный момент времени t , называют эргодичностью.

Это довольно “поэтический” и неконкретный термин. Более конкретно говорят об эргодичности по математическому ожиданию, эргодичности по автокорреляции и т.п. — когда выборочные оценки соответствующих моментов, сделанные по наблюдению из одной траектории, состоятельны, т.е. сходятся к истинным значениям моментов.

□

⟨7⟩ $\mu = \mathbf{E} \xi(t)$ и $\gamma(h)$ — это базовые характеристики, рассматриваемые при первичном разведочном анализе выборочных данных из случайного процесса ξ , когда есть подозрение о том, что ξ — стационарный или “около-стационарный” процесс.

В дальнейшем мы будем рассматривать вопрос их оценивания по выборкам из стационарных процессов и условиях их состоятельности.

⟨8⟩ **Определение.** Величину

$$r(h) = \frac{\gamma(h)}{\gamma(0)}$$

называют автокорреляцией случайного процесса. Её свойства в целом повторяют свойства корреляции.

□

⟨9⟩ **Определение.** При исследовании зависимости двух стационарных процессов ξ и η рассматривают их кросс-ковариацию

$$\gamma_{\xi, \eta}(h) = \mathbf{E}(\xi(t) - \mu_\xi)(\eta(t+h) - \mu_\eta).$$

□

⟨10⟩ В отличие от γ , кросс-ковариация $\gamma_{\xi,\eta}(h)$ не симметричная. Так же теряется особый смысл величины $\gamma_{\xi,\eta}(0)$.

5.1 Оценивание математического ожидания стационарного процесса

⟨11⟩ Рассмотрим выборочное математическое ожидание стационарного процесса ξ по одной выборочной траектории:

$$\hat{\mu}_{tr} = \frac{1}{n} \sum_{i=1}^n \xi(\tau_i).$$

Легко видеть, что $\hat{\mu}_{tr}$ — несмещённая оценка для μ , т.е.

$$\mathbf{E} \hat{\mu}_{tr} = \mu.$$

Рассмотрим при каких условиях $\lim_{n \rightarrow \infty} \mathbf{D} \hat{\mu}_{tr}(n) \rightarrow 0$ — при этих условиях оценка $\hat{\mu}_{tr}$ будет состоятельной. Исходя из свойств ковариации

$$\begin{aligned} \mathbf{D} \hat{\mu}_{tr} &= \text{cov} \left(\frac{1}{n} \sum_{i=1}^n \xi(\tau_i), \frac{1}{n} \sum_{j=1}^n \xi(\tau_j) \right) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \gamma(\tau_j - \tau_i) = \\ &= \frac{1}{n^2} \left(n \gamma(0) + \sum_{i,j: i < j} \gamma(\tau_j - \tau_i) \right) = \frac{\gamma(0)}{n} + \frac{2}{n^2} \sum_{i,j: i < j} \gamma(\tau_j - \tau_i), \end{aligned}$$

где $\tau_1 \leq \tau_2 \leq \dots \leq \tau_n$. $\mathbf{D} \hat{\mu}_{tr}$ отличается от дисперсии оценки математического ожидания по выборке независимых наблюдений на величину $(2/n^2) \sum_{i,j: i < j} \gamma(\tau_j - \tau_i)$ — ошибка прогноза, возникающая в силу зависимости наблюдений.

Таким образом, $\hat{\mu}_{tr}(n)$ — состоятельная оценка, если $\gamma(h)$ и τ_1, \dots, τ_n таковы, что

$$\frac{2}{n^2} \sum_{i,j: i < j} \gamma(\tau_j - \tau_i) \rightarrow 0 \quad \text{при} \quad n \rightarrow \infty. \tag{12}$$

⟨13⟩ **Теорема.** Пусть $\tau_i = \tau_{i-1} + \Delta\tau$. Тогда состоятельность оценки $\hat{\mu}_{tr}$ следует из

$$\lim_{h \rightarrow \infty} \gamma(h) = 0. \quad \square$$

Доказательство. (12) можно преобразовать и оценить сверху:

$$\frac{2}{n^2} \sum_{i=1}^{n-1} (n-i) \gamma(i \Delta\tau) = \frac{2}{n} \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) \gamma(i \Delta\tau) \leq \frac{2}{n} \sum_{i=1}^{n-1} \gamma(i \Delta\tau) \rightarrow \lim_{h \rightarrow \infty} \gamma(h) = 0,$$

где на последнем шаге для перехода от суммы к пределу мы воспользовались свойством, что если существует $\lim_{n \rightarrow \infty} x_n$, то $\lim_{n \rightarrow \infty} (1/n) \sum_{i=1}^n x_i = \lim_{n \rightarrow \infty} x_n$ (следствие теоремы Штольца). \square

⟨14⟩ **Определение.** Стационарный процесс, для которого выборочное среднее по одной выборочной траектории является состоятельной оценкой, иногда называют эргодичным по среднему (см., например, [5]). \square

⟨15⟩ Пусть имеется выборка $D = \{X_1, \dots, X_m\}$, состоящая из m независимых выборочных траекторий, и пусть n_1, \dots, n_m — количества наблюдений в них. С учётом несмещённости $\hat{\mu}_{tr}$ рассмотрим два пути:

$$1) \quad \hat{\mu}_1 = \frac{1}{m} \sum_i \hat{\mu}_{tr}(X_i).$$

$$2) \quad \hat{\mu}_2 = \frac{1}{\sum_i n_i} \sum_j \sum_{x \in X_j} x.$$

Оценки $\hat{\mu}_1$ и $\hat{\mu}_2$ дают похожий результат, если $n_i \approx n_j$. Если $\{n_i\}$ существенно различаются, то

- 1) при сильной взаимной зависимости $\xi(t)$ и $\xi(t+h)$ оценка $\hat{\mu}_1$ точнее (большой вес даётся независимым наблюдениям)
- 2) при слабой взаимной зависимости $\xi(t)$ и $\xi(t+h)$ оценка $\hat{\mu}_2$ точнее (вес по наблюдениям распределяется равномерно)

⟨16⟩ Для выбора между $\hat{\mu}_1$ и $\hat{\mu}_2$ на практике можно, например, сравнить выборочные распределения выборок X_i — если оно примерно совпадает во всех выборках, то это хороший признак для использования $\hat{\mu}_2$.

5.2 Оценивание автоковариации стационарного процесса 🚧

⟨17⟩ Оценивание автоковариации — проблема сложная, и универсального хорошего решения пока нет или оно не получило распространения. Для простоты мы будем рассматривать только случай процессов с дискретным временем и выборочными траекториями без пропусков.

⟨18⟩ Для процессов с дискретным временем (для определённости $T \subseteq \mathbb{Z}$) для оценивания обычно предлагают использовать выборочную автоковариацию. Пусть дана выборочная траектория $\xi(1), \dots, \xi(n)$. Выборочная оценка автоковариации $h \in \mathbb{Z}_+$ порядка имеет вид:

$$\hat{\gamma}(h) = \frac{1}{n-h} \sum_{i=1}^{n-h} (\xi(i) - \hat{\mu})(\xi(i+h) - \hat{\mu}).$$

Полагаем

$$\hat{\gamma}(-h) = \hat{\gamma}(h).$$

⟨19⟩ Пусть дана выборка независимых выборочных траекторий:

- 1) $D = \{X_1, \dots, X_m\}$.
- 2) Объёмы выборки n_1, \dots, n_m .
- 3) X_i — выборки без пропусков.

В качестве оценки автоковариации по нескольким независимым выборочным траекториям можно предложить:

$$\hat{\gamma}(h) = \frac{1}{\sum_i \max\{0, n_i - h\}} \sum_j \sum_{k=1}^{n_j-h} (X_j(k) - \hat{\mu})(X_j(k+h) - \hat{\mu}),$$

где $\hat{\mu}$ — оценка $\mathbf{E} \xi(t)$.

⟨20⟩ Некоторые свойства $\hat{\gamma}(h)$ при оценивании по одной выборочной траектории (без подробного разбора):

- 1) Асимптотически несмещённая оценка.
- 2) Состоятельная, если для всех $h \in \mathbb{Z}_+$

$$\lim_{l \rightarrow \infty} \text{cov} \left((\xi(t) - \mu)(\xi(t+h) - \mu), (\xi(t+l) - \mu)(\xi(t+h+l) - \mu) \right) = 0,$$

где $\mu = \mathbf{E} \xi(t)$.

- 3) При фиксированном объёме выборки точность оценивания падает с ростом оцениваемого порядка h в $\gamma(h)$.
- 4) В общем случае не образует положительно определённую функцию — выборочная автоковариация, формально говоря, может не быть автоковариацией. Как пишут, это может осложнить её использование при анализе, например, спектрального разложения процесса. Положительной определённостью обладает смещённая версия выборочной автоковариации:

$$\hat{\gamma}_b(h) = \frac{1}{n} \sum_{i=1}^{n-h} (\xi(i) - \hat{\mu})(\xi(i+h) - \hat{\mu}).$$

⟨21⟩ Определение. Стационарный процесс, для которого выборочная автоковариация по одной выборочной траектории является состоятельной оценкой, иногда называют эргодичным по автоковариации (см., например, [5]). \square

⟨22⟩ Для оценивания кросс-ковариации процессов ξ и η по выборке D применяется выборочная кросс-ковариация

$$\hat{\gamma}_{\xi, \eta}(h) = \frac{1}{|P_h|} \sum_{(x,y) \in P_h} (x - \hat{\mu}_\xi) \cdot (y - \hat{\mu}_\eta),$$

где

- 1) $P_h = \{(X_i(j), Y_i(j+h)) : i = 1, \dots, m; \quad X_i(j) \in X_i; \quad Y_i(j+h) \in Y_i\}$.
- 2) $D = \{(X_1, Y_1), \dots, (X_m, Y_m)\}$.
- 3) $X_i = \{\xi^{(i)}(\tau_{\xi i} + j - 1) : j = 1, \dots, n_{\xi i}\}$.
- 4) $Y_i = \{\eta^{(i)}(\tau_{\eta i} + j - 1) : j = 1, \dots, n_{\eta i}\}$.
- 5) Множество пар выборочных траекторий $\{(X_i, Y_i)\}$ — независимы в совокупности.
- 6) $\hat{\mu}_\xi$ и $\hat{\mu}_\eta$ — оценки математических ожиданий ξ и η , соответственно.

Условие состоятельности оценки аналогично таковому для автоковариации.

6 Процессы авторегрессии скользящего среднего

⟨1⟩ Определение. Процесс авторегрессии скользящего среднего (p, q) -го порядка (АРСС, Autoregressive Moving-Average, ARMA) ξ определяется как:

- 1) Процесс с дискретным временем и $T = \mathbb{Z}$. В обычной постановке у процесса нет начального и конечного моментов времени.
- 2) $\text{dom } \xi = \mathbb{R}$.
- 3) Значения $\xi(t)$ связаны с предыдущими согласно рекурсивной формуле:

$$\xi(t) = \varepsilon(t) + \sum_{i=1}^p \varphi_i \xi(t-i) - \sum_{j=1}^q \theta_j \varepsilon(t-j) + a_0, \quad \varepsilon(t) \sim \mathcal{N}(0, \sigma_\varepsilon^2), \quad (2)$$

где

- 1) $\varphi_i, \theta_j, a_0 \in \mathbb{R}$.
- 2) $\sigma_\varepsilon > 0$ — дисперсия белого шума.
- 3) $\{\varepsilon(t)\}$ — независимы в совокупности.

Обозначение. $\xi \sim ARMA(p, q)$. Если $p > 0$ и $q = 0$, то $\xi(t)$ называют авторегрессией p -го порядка и обозначают как $\xi \sim AR(p)$. Если $p = 0$ и $q > 0$, то $\xi(t)$ называют скользящим средним q -го порядка и обозначают как $\xi \sim MA(q)$.

Множество параметров модели: $\varphi = \{\varphi_i : i \in \{1, \dots, p\}\}$, $\theta = \{\theta_i : i \in \{1, \dots, q\}\}$, a_0 , σ_ε^2 . □

⟨3⟩ Определение. Последовательность независимых нормальных случайных величин $\{\varepsilon(t)\}$ с $\mathbf{E} \varepsilon(t) = 0$ в составе процесса АРСС называют белым шумом. □

⟨4⟩ При $p = 0$ АРСС всегда стационарен. При $p > 0$ стационарность зависит от значений $\{\varphi_i\}$.

⟨5⟩ Рекурсивная формула (2) неявным образом задаёт условное распределение

$$\left(\xi(t) \mid \xi(t-1), \dots, \xi(t-p), \varepsilon(t-1), \dots, \varepsilon(t-q) \right) \sim \mathcal{N}(\mathcal{M}, \sigma_\varepsilon^2),$$

где

$$\mathcal{M} = \sum_{i=1}^p \varphi_i x_i - \sum_{j=1}^q \theta_j e_j + a_0.$$

⟨6⟩ В зависимости от параметров $\{\varphi_i\}$, процесс ξ может быть или не быть стационарным. Условия стационарности будут приведены ниже. Когда процесс ξ не стационарен, для его траекторий становится характерным экспоненциальный, “взрывной” рост значений до ∞ .

⟨7⟩ Известно, что при некоторых условиях на спектральную плотность (которые в практических задачах, обычно, считаются выполненными), стационарный в широком смысле процесс может быть представлен в виде бесконечномерного скользящего суммирования:

$$\xi(t) = \sum_{i=0}^{\infty} \theta_i \tilde{\varepsilon}(t-i) + a_0,$$

где $\tilde{\varepsilon}(t)$ — последовательность белого шума с некоторым распределением и $\mathbf{E} \tilde{\varepsilon}(t) = 0$.

Это позволяет рассматривать процессы АРСС как естественное приближение для класса стационарных процессов в широком смысле с дискретным временем.

⟨8⟩ **Свойство.** Если $\xi \sim ARMA(p, q)$ стационарен, то

$$\mathbf{E} \xi(t) = \frac{a_0}{1 - \sum_{i=1}^p \varphi_i}. \quad \square$$

⟨9⟩ **Замечание.** Если и $p > 0$, и $q > 0$, то авторегрессионная часть и часть скользящего среднего могут взаимокompенсироваться / подавлять друг друга. Например, если $p = q$ и $\varphi_i = \theta_i$ для всех $i \in \{1, \dots, p\}$, то ξ вырождается в белый шум. Чтобы проверить это, достаточно рекурсивно подставлять выражения для $\xi(t-l)$, $l > 0$ в (2). \square

⟨10⟩ **Замечание.** При применении АРСС для моделирования реальных данных авторы рекомендуют рассматривать $p, q \leq 2$. \square

6.1 Представление через операторы сдвига и условие стационарности

⟨11⟩ Рекуррентная формула (2), характеризующая вероятностные свойства $\xi \sim ARMA(p, q)$, более компактно записывается с использованием операторов сдвига назад и вперёд. Дополнительная польза от такого представления — условия стационарности может быть выражено в терминах тех же конструкций, что и определение 2.

В дальнейших определениях операторов пусть \mathbf{A} обозначает некоторое выражение.

⟨12⟩ **Определение.**

1) Оператор сдвига вперёд F определим как: $F \mathbf{A}(t) = \mathbf{A}(t+1)$.

2) Оператор сдвига назад B определим как: $B \mathbf{A}(t) = \mathbf{A}(t-1)$.

Например, $B \xi(t) = \xi(t-1)$. \square

⟨13⟩ **Определение.** Для операторов F и B определим набор преобразований. Для $X \in \{F, B\}$ пусть:

1) Возведение в степень k : $X^k \mathbf{A} = \underbrace{X \cdots X}_{k \text{ раз}} \mathbf{A}$. Полагаем $X^0 \mathbf{A} = \mathbf{A}$.
Например, $B^k \xi(t) = \xi(t-k)$.

2) Линейное преобразование: $(aX_1 + bX_2) \mathbf{A} = aX_1 \mathbf{A} + bX_2 \mathbf{A}$, где $a, b \in \mathbb{R}$.
Например, $(aB^k + bB^l) \xi(t) = a\xi(t-k) + b\xi(t-l)$. \square

⟨14⟩ Определение.

1) Оператор авторегрессии p -го порядка $\Phi_p(\mathbf{A})$ определяется как:

$$\Phi_p(\mathbf{A}) = \mathbf{A}^0 - \sum_{i=1}^p \varphi_i \mathbf{A}^i.$$

Например,

$$\Phi_p(B) \xi(t) = \left(B^0 - \sum_{i=1}^p \varphi_i B^i \right) \xi(t) = \xi(t) - \sum_{i=1}^p \varphi_i \xi(t-i).$$

2) Оператор скользящего среднего q -го порядка $\Theta_q(\mathbf{A})$ определяется как:

$$\Theta_q(\mathbf{A}) = \mathbf{A}^0 - \sum_{i=1}^q \theta_i \mathbf{A}^i.$$

Например,

$$\Theta_q(B) \varepsilon(t) = \left(B^0 - \sum_{i=1}^q \theta_i B^i \right) \varepsilon(t) = \varepsilon(t) - \sum_{i=1}^q \theta_i \varepsilon(t-i).$$

3) Формула для значения процесса АРСС (2) может быть выражена через операторы Φ_p и Θ_q в виде:

$$\Phi_p(B) \xi(t) = \Theta_q(B) \varepsilon(t) + a_0.$$

□

⟨15⟩ Свойство. Процесс $\xi \sim ARMA(p, q)$ с $p > 0$ стационарен тогда и только тогда, когда все комплексные корни уравнения $\Phi_p(z) = 0$ ($z \in \mathbb{C}$) лежат вне единичного круга.

Для случая $p = 1$ условие стационарности сводится к неравенству $|\varphi_1| < 1$.

Для случая $p = 2$ условие стационарности сводится к системе неравенств:

$$\begin{cases} \varphi_2 + \varphi_1 < 1, \\ \varphi_2 - \varphi_1 < 1, \\ |\varphi_2| < 1. \end{cases}$$

□

6.2 Визуальная оценка значений p и q на основе анализа графиков автокорреляции и частной автокорреляции

⟨16⟩ Раньше оценивать p и q предлагалось по поведению графиков автокорреляции и частной автокорреляции рассматриваемого процесса. Этот метод даёт чёткие результаты только для случаев чистых AR и MA . Для смешанных моделей (и $p > 0$, и $q > 0$) результаты, обычно, неоднозначные.

Вкратце рассмотрим эту методику, т.к. про неё часто рассказывают в учебных материалах по АРСС.

⟨17⟩ Определение. Частная корреляция (Partial Correlation) между случайными величинами X и Y за вычетом влияния Z_1, \dots, Z_k называют корреляцию между остатками линейной регрессии Z_1, \dots, Z_k на X и остатками линейной регрессии Z_1, \dots, Z_k на Y , т.е.:

$$PartCorr(X, Y; Z_1, \dots, Z_k) = \text{cov}(\varepsilon_X, \varepsilon_Y),$$

где

- 1) $\varepsilon_X = X - l_X(Z_1, \dots, Z_k)$.
- 2) $\varepsilon_Y = Y - l_Y(Z_1, \dots, Z_k)$.
- 3) l_X , l_Y — соответствующие функции линейных регрессий.

Частная корреляции интерпретируется как корреляция между X и Y , из которой убрана та часть корреляции, проводниками которой являются Z_1, \dots, Z_k . \square

⟨18⟩ Определение. В общем случае стационарного в широком смысле процесса ξ частная автокорреляция (Partial Autocorrelation Function, PACF) h -го определяется как

$$\psi(h) = PartCorr(\xi(t), \xi(t+h); \{\xi(l) : t < l < t+h\}).$$

В случае стационарных процессов частная автокорреляция может быть вычислена как

$$\psi(h) = \left\{ \psi_h : \gamma(i) = \sum_{j=1}^h \gamma(i-j) \psi_j; \psi_1, \dots, \psi_h \in \mathbb{R}; i = 1, \dots, h \right\}. \quad (19)$$

\square

⟨20⟩ Если есть оценки $\hat{\gamma}(0), \dots, \hat{\gamma}(h)$, то $\psi(h)$ может быть оценена решением системы уравнения из (19) с подстановкой вместо $\gamma(i)$ их оценок $\hat{\gamma}(i)$.

⟨21⟩ Замечание. Пусть $\xi \sim AR(p)$. Система уравнений

$$\begin{cases} r(1) = \sum_{j=1}^p r(j-1) \varphi_j; \\ r(2) = \sum_{j=1}^p r(j-2) \varphi_j; \\ \dots \\ r(p) = \sum_{j=1}^p r(j-p) \varphi_j; \end{cases}$$

$\varphi_1, \dots, \varphi_p$ называется системой уравнений Юла-Уокера, и она связывает значения автокорреляции ξ и параметры модели. В формуле (19) величина $\psi(h)$ вычисляется как последний элемент решения системы уравнений Юла-Уокера, решаемой относительно коэффициентов $\{\varphi_j\}$ при имеющихся значениях $r(i)$. \square

⟨22⟩ Свойство.

- 1) Если $\xi \sim AR(p)$, то
 - $\psi(h) = 0$ для $h > p$.

- $r(h)$ экспоненциально/синусоидально затухает.

2) Если $\xi \sim MA(q)$, то

- $r(h) = 0$ для $h > q$.
- $\psi(h)$ экспоненциально/синусоидально затухает.

3) Если $\xi \sim ARMA(p, q)$ с $p, q > 0$, то

- $r(h)$ экспоненциально/синусоидально затухает с $h \geq q + 1$.
- $\psi(h)$ экспоненциально/синусоидально затухает с $h \geq p + 1$.

□

⟨23⟩ Наблюдая, с какого шага графики $r(h) = 0$ и $\psi(h) = 0$ или с какого шага начинают убывать, можно исходя из Свойств 22 попытаться угадать параметры p и q ARСС модели.

⟨24⟩ На практике мы, обычно, имеем дело только с выборочными оценками $\hat{r}(h)$ и $\hat{\psi}(h)$. Для проверки выполнения условия $r(h) = 0$ и $\psi(h) = 0$ мы вынуждены применять критерии проверки соответствующих гипотез.

Известно, что для $\xi \sim MA(q)$ для $h > q$ выборочные оценки $\hat{r}(h)$ асимптотически распределены как $\mathcal{N}(0, 1/(n - h))$. Аналогично, для $\xi \sim AR(p)$ при $h > p$ оценки $\hat{\psi}(h)$ асимптотически распределены как $\mathcal{N}(0, 1/(n - h))$.

6.3 Оценивание $\varphi, \theta, \sigma_\varepsilon^2, a_0$ методом условного максимального правдоподобия

⟨25⟩ Пусть $\xi \sim ARMA(p, q)$, известны p и q и пусть дана выборочная траектория $X = (\xi(1), \dots, \xi(n))$. Стоит задача оценивания набора параметров

$$\Theta = \{\varphi, \theta, \sigma_\varepsilon^2, a_0\}.$$

Метод максимального правдоподобия предписывает определить оценку как

$$\arg \max_{\Theta} \mathbf{f}(\xi(1), \dots, \xi(n); \Theta),$$

что подразумевает необходимость знания распределения $\xi(t)$, $t \leq \max\{p, q\} + 1$ при неизвестных, незафиксированных $\xi(l)$, $\varepsilon(l)$, $l \leq 0$. Эта функция плотности тяжело вычислима, что усложняет применение обычного максимума правдоподобия (хотя это и возможно, см., например, [7]).

Для простоты мы рассмотрим метод условного максимума правдоподобия

$$\hat{\Theta} = \arg \max_{\Theta} L(\Theta; X), \quad (26)$$

где

$$L(\Theta; X) = \mathbf{f}(\xi(p+1), \dots, \xi(n) \mid \xi(1), \dots, \xi(p), \varepsilon(p-q) = 0, \dots, \varepsilon(p-1) = 0; \Theta). \quad (27)$$

Очевидные минусы такого подхода — несколько хуже утилизируется информация о первых p элементах выборки, а также использование предположения о равенстве нулю белого шума. Оба эти недостатка нивелируются при росте объёма выборки n .

⟨28⟩ Теорема. Оценки по методу условного правдоподобия по значениям выборочной траектории $X = (\xi(1), \dots, \xi(n))$ имеют вид:

$$1) \quad \hat{\varphi}, \hat{\theta}, \hat{a}_0 = \min_{\varphi, \theta, a_0} \sum_{t=p+1}^n \hat{\varepsilon}(t; \varphi, \theta, a_0)^2,$$

где для $t \leq p$ полагаем $\hat{\varepsilon}(t) = 0$, а для $t \in \{p+1, \dots, n\}$ (сопоставьте с (2))

$$\hat{\varepsilon}(t; \varphi, \theta, a_0) = \xi(t) - \sum_{i=1}^p \varphi_i \xi(t-i) + \sum_{j=1}^q \theta_j \hat{\varepsilon}(t-j) - a_0.$$

$$2) \quad \hat{\sigma}_\varepsilon^2 = \frac{1}{n-p} S_e(\hat{\varphi}, \hat{\theta}, \hat{a}_0).$$

□

Доказательство. Распишем и упростим (26). Обозначим значения выборочной траектории как $x_t = \xi(t)$. Исходя из уравнения для АРСС процесса (2):

$$\begin{aligned} \xi(t) \in [a; b] &\iff \varepsilon(t) + \sum_{i=1}^p \varphi_i \xi(t-i) - \sum_{j=1}^q \theta_j \varepsilon(t-j) + a_0 \in [a; b] \iff \\ &\iff \varepsilon(t) \in [a - g_t; b - g_t], \end{aligned}$$

где

$$g_t = \sum_{i=1}^p \varphi_i \xi(t-i) - \sum_{j=1}^q \theta_j \varepsilon(t-j) + a_0.$$

Выразим совместную функцию плотности $\xi(p+1), \dots, \xi(n)$ в терминах функции плотности $\varepsilon(p+1), \dots, \varepsilon(n)$:

$$\begin{aligned} &\mathbf{f} \left(\xi(p+1) = x_{p+1}, \dots, \xi(n) = x_n \mid \xi(1) = x_1, \dots, \xi(p) = x_p, \varepsilon(p-q) = 0, \dots \right. \\ &\quad \left. \dots, \varepsilon(p-1) = 0; \varphi, \theta, a_0 \right) = \\ &= \lim_{\Delta_{p+1} \rightarrow 0+, \dots, \Delta_n \rightarrow 0+} \frac{\mathbf{P}(\xi(p+1) \in [x_{p+1}; x_{p+1} + \Delta], \dots \mid \dots; \dots)}{\Delta_{p+1} \cdot \dots \cdot \Delta_n} = \\ &= \lim_{\Delta_{p+1} \rightarrow 0+, \dots, \Delta_n \rightarrow 0+} \frac{\mathbf{P}(\varepsilon(p+1) \in [x_{p+1} - g_{p+1}; x_{p+1} + \Delta_{p+1} - g_{p+1}], \dots \mid \dots; \dots)}{\Delta_{p+1} \cdot \dots \cdot \Delta_n} = \\ &= \mathbf{f}(\varepsilon(p+1) = x_{p+1} - g_{p+1}, \dots \mid \dots; \dots). \end{aligned} \tag{29}$$

Заметим, что значения $\varepsilon(p+1), \dots, \varepsilon(n)$ однозначно задаются значениями $\varepsilon(p-q), \dots, \varepsilon(p), \xi(1), \dots, \xi(n)$ исходя из (2). Отсюда можно показать, что (29) можно заменить на

$$\mathbf{f}(\varepsilon(p+1) = x_{p+1} - \hat{g}_{p+1}, \dots, \varepsilon(n) = x_n - \hat{g}_n; \varphi, \theta, a_0),$$

где \hat{g}_t — это g_t , в которых $\{\xi(t)\}$ и $\{\varepsilon(t)\}$ выражены из x_1, \dots, x_n . В силу независимости в совокупности $\{\varepsilon(t)\}$ имеем:

$$\ln L(\varphi, \theta, a_0; X) = C - \frac{1}{2} \ln \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} S_e,$$

где

$$S_e = \sum_{t=p+1}^n (x_t - \hat{g}_t)^2.$$

Видно, что при минимизации $\ln L$ минимум S_e не зависит от выбора σ_ε^2 . Следовательно, минимизацию $\ln L$ можно осуществлять по шагам:

- 1) Минимизация S_e .
 - 2) Максимизация $\ln L$ по σ_ε^2 при найденном минимальном S_e .
-

6.4 Подгонка p и q методом Bayesian Information Criterion (BIC)

⟨30⟩ Пусть $\xi \sim ARMA(p, q)$ с набором неизвестных параметров $\Theta_{pq} = \{\varphi, \theta, \sigma_\varepsilon^2, a_0\}$. Пусть имеется выборочная траектория $X = \{\xi(1), \dots, \xi(n)\}$. Статистика BIC критерия имеет вид:

$$BIC(p, q; X) = -2 \ln L(\hat{\Theta}_{pq}; X) + |\Theta_{pq}| \ln n,$$

где L — функция правдоподобия, соответствующая $\xi \sim ARMA(p, q)$; $\hat{\Theta}_{pq} = \arg \max_{\Theta} L(\Theta; D)$; $|\Theta_{pq}|$ — количество неизвестных параметров модели ($p + q + 2$).

Оценка по критерию BIC :

$$\hat{p}, \hat{q} = \arg \min_{p, q} BIC(p, q; X).$$

⟨31⟩ Если в статистике BIC использовать в качестве L функцию условного правдоподобия (27), то, формально говоря, p элементов из выборочной траектории “тратятся” на условие, и $\ln n$ необходимо заменить на $\ln(n - p)$.

⟨32⟩ Если часть параметров из числа $\varphi, \theta, \sigma_\varepsilon^2, a_0$ известна, то на соответствующее количество надо уменьшить множитель $|\Theta_{pq}|$ в выражении для статистики BIC . Например, если σ_ε^2 и a_0 известны, то $|\Theta_{pq}| = p + q$.

⟨33⟩ Если имеется выборка из нескольких выборочных траекторий, то в формуле статистики BIC можно использовать функцию правдоподобия по всей выборке, а n положить равным общему объёму выборки.

⟨34⟩ **Замечание.** Кратко опишем идею, на которой основывается BIC критерий. Пусть рассматривается случайная величина / случайный процесс ξ , и пусть есть несколько предположений о распределении ξ , которые часто называют моделями: $\mathcal{M}_1, \dots, \mathcal{M}_m$. Пусть Θ_i обозначает набор параметров модели \mathcal{M}_i . Предполагается, что все модели принадлежат экспоненциальному семейству распределений. Для каждой модели предполагается наличие некоторого априорного распределения набора параметров Θ_i (отсюда в названии критерия слово “байесовский”). Пусть имеется набор выборочных данных \mathcal{D} с n наблюдениями.

В предположении, что функции плотности априорных распределений и функции правдоподобия каждой из моделей \mathcal{M}_i дважды дифференцируемы, для маргинального распределения выборки для заданной модели верно асимптотическое при $n \rightarrow \infty$

представление:

$$-2 \ln \mathbf{f}(\mathcal{D}; \mathcal{M}_i) = -2 \ln L(\hat{\Theta}_i; \mathcal{D}) + |\Theta_i| \ln n + O(1).$$

Отсюда

$$\ln \mathbf{f}(\mathcal{D}; \mathcal{M}_i) \approx -\frac{BIC(\mathcal{M}_i; \mathcal{D})}{2}.$$

Следовательно, если руководствоваться принципом наибольшего правдоподобия, то

$$\widehat{\mathcal{M}} = \arg \max_{i \in \{1, \dots, m\}} \mathbf{f}(\mathcal{D}; \mathcal{M}_i) = \arg \min_{i \in \{1, \dots, m\}} BIC(\mathcal{M}_i; \mathcal{D}).$$

Подробнее о выводе BIC см., например, в [8]. □

⟨35⟩ Замечание. Хотя BIC критерий и обладает некоторым обоснованием, его лучше воспринимать как некоторое полезное, но грубое приближение. Во-первых, асимптотика, стоящая в основе критерия, очень груба. Во-вторых, в обосновании критерия стоят предположения о распределении наблюдаемой случайной величины, которые во многих применениях выполняются лишь приближённо. □

6.5 Подгонка p и q на основе проверки гипотезы отсутствия автокорреляции в остатках модели

⟨36⟩ Пусть $\xi \sim ARMA(p, q)$ — стационарный. Уравнение

$$\xi(t) = \varepsilon(t) + \sum_{i=1}^p \varphi_i \xi(t-i) - \sum_{j=1}^q \theta_j \varepsilon(t-j) + a_0 \quad (37)$$

задаёт взаимно-однозначное соответствие между $\{\xi(t) : t \leq t_0\}$ и $\{\varepsilon(t) : t \leq t_0\}$.

Пусть нам даны значения $\mathcal{X} = \{\xi(t) : t \leq t_0\}$. Обозначим через $h : \{\xi(t) : t \leq t_0\} \rightarrow \{\varepsilon(t) : t \leq t_0\}$ отображение, обратное к (37). Пусть $\mathcal{E} = h(\mathcal{X})$ — значения ε , соответствующие значениям ξ из \mathcal{X} .

Если преобразование h совпадает с действительным распределением ξ (т.е. совпадают $p, q, \varphi, \theta, a_0$), то набор \mathcal{E} будет представлять собой действительную реализацию случайных величин $\{\varepsilon(t) : t \leq t_0\}$. В частности, для значения \mathcal{E} — это реализации последовательности независимых случайных величин.

С другой стороны, если преобразование h не совпадает с действительным распределением ξ (т.е. не совпадает что-то из числа параметров $p, q, \varphi, \theta, a_0$), то \mathcal{E} не будет являться реализацией $\{\varepsilon(t) : t \leq t_0\}$, а нечто другого. В частности, можно ожидать, что \mathcal{E} будет являться реализацией последовательности зависимых случайных величин.

На этом явлении основывается способ оценки соответствия выбранной модели АРСС выборочным данным — отсутствие независимости в остатках \mathcal{E} означает несоответствие модели действительному распределению наблюдаемого процесса.

⟨38⟩ Предположим, что $\xi \sim ARMA(p_0, q_0)$ с неизвестными p_0, q_0 , ε — соответствующий ξ белый шум, и пусть дана выборочная траектория $X = \{\xi(1), \dots, \xi(n)\}$. Пусть к X подогнана модель $ARMA(p_1, q_1)$, и $\hat{\varepsilon}(1), \dots, \hat{\varepsilon}(n)$ — оценки белого шума (остатки модели) исходя из этих p_1, q_1 .

Можно считать, что применение модели $ARMA(p_1, q_1)$ к ξ порождает случайный процесс ε^* , выборочной траекторией из которого является $\hat{\varepsilon}(1), \dots, \hat{\varepsilon}(n)$.

Для проверки на отсутствие автокорреляций в ε^* можно применить критерий Льюнга-Бокса, проверяющий гипотезу

$$H_0: r(h; \varepsilon^*) = 0 \quad \forall h \in \{1, \dots, K\},$$

где K — параметр критерия, определяющий максимально рассматриваемый критерием лаг.

Статистика критерия:

$$Q = (n + 2)n \sum_{h=1}^K \frac{(\hat{r}(h; \varepsilon^*))^2}{n - h}$$

При верности гипотезы $Q \rightsquigarrow \chi^2(K - p - q)$, где $\chi^2(K - p - q)$ — хи-квадрат распределение с $K - p - q$ степенями свободы. p -значение критерия вычисляется как

$$p = 1 - F(Q; \chi^2(K - p - q))$$

6.6 Процесс авторегрессии проинтегрированного скользящего среднего

⟨39⟩ Определение. Разностный оператор определяется как $\nabla = (1 - B)$, т.е. $\nabla \xi(t) = (1 - B)\xi(t) = \xi(t) - \xi(t - 1)$. \square

⟨40⟩ Определение. ξ называется процессом авторегрессии проинтегрированного скользящего суммирования p, d, q -го порядка, если $\nabla^d \xi \sim ARMA(p, q)$. В терминах уравнения, описывающего значение $\xi(t)$ определение выглядит как

$$\Phi_p(B) \nabla^d \xi(t) = \Theta_q(B) \varepsilon(t) + a_0.$$

Обозначение. $\xi \sim ARIMA(p, d, q)$ (АРПСС, ARIMA — Autoregressive Integrated Moving Average)

Параметры модели: $p, d, q, \varphi, \theta, a_0, \sigma_\varepsilon^2$. \square

⟨41⟩ Ключевой особенностью процессов АРПСС в сравнении с АРСС — они описывают нестационарные процессы, “среднее” значение которых может меняться с течением времени. Добиваются этого некоторой “недоопределённости” процесса. Приведём интерпретацию того, что модель описывает.

Применение разностного оператора ∇ — это приближённый аналог взятия производной от функции. Ассоциируем задание распределения ξ с заданием некоторой функции $f(x)$, а объявление $\nabla^d \xi \sim ARMA(p, q)$ с заданием соответствующей производной f равной функции $g(x)$. Тогда

- 1) $d = 0 \Rightarrow f(x) = g(x)$ — ARMA модель полностью задаёт поведение процесса и определяет его стационарным.
- 2) $d = 1 \Rightarrow f'(x) = g(x) \Rightarrow f(x) = G(x) + C$ — ARMA модель задаёт поведение процесса с точностью до 1-ой “производной”, и только 1-я “производная” обязана быть стационарной; у процесса может произвольно меняться средний уровень C .

- 3) $d = 2 \Rightarrow f''(x) = g(x) \Rightarrow f(x) = G(x) + ax + C$ — $ARMA$ модель задаёт поведение процесса с точностью до 2-ой “производной”, и только 2-я “производная” обязана быть стационарной; у процесса может произвольно меняться средний уровень C и линейный тренд at .

В целом, значение d указывает, на производную какого порядка “навешивается” $ARMA$ модель.

⟨42⟩ Замечание. При применении АРПСС для моделирования реальных данных авторы рекомендуют рассматривать $p, d, q \leq 2$. □

⟨43⟩ Пусть $\xi \sim ARIMA(p, d, q)$ и обозначим $\eta(t) = \nabla^d \xi(t)$. Преобразование от $\xi(t)$ к $\eta(t)$ задаёт и обратное преобразование от $\eta(t)$ к $\xi(t)$ в виде суммы значения $\eta(t)$ и линейной комбинации предыдущих значений $\xi(t)$. Например,

- 1) при $d = 1$ имеем $\eta(t) = \xi(t) - \xi(t-1) \Rightarrow \xi(t) = \eta(t) + \xi(t-1)$.
- 2) при $d = 2$ имеем $\eta(t) = \xi(t) - 2\xi(t-1) + \xi(t-2) \Rightarrow \xi(t) = \eta(t) + 2\xi(t-1) - \xi(t-2)$.

6.7 Оценивание параметра d модели $ARIMA(p, d, q)$

⟨44⟩ Ключевая особенность АРПСС в сравнении с АРСС — отсутствие стационарности. Это используется в качестве критерия для оценивания d по выборочным данным. Если нам дана выборочная траектория $\xi(1), \dots, \xi(n)$, то преобразование $\eta(t) = \xi(t) - \xi(t-1)$ для $t \in \{2, \dots, n\}$ нам даст реализацию процесса $\eta(t) = \nabla \xi(t)$. Исходя из этого, оценку d исходя из выборочной траектории можно определить как

$$\hat{d} = \min\{d \geq 0: \nabla^d \xi(t) \text{ для } t \in \{d+1, \dots, n\} \text{ — выборка из стационарного процесса}\}.$$

Конкретные методы отличаются способом проверки гипотезы стационарности. Универсального общепринятого способа проверка на стационарность нет, в источниках часто рекомендуют применять либо критерий *Augmented Dickey–Fuller*, либо критерий *KPSS* (см. [9]). В целом, оба этих и другие критерии могут надёжно проверять на стационарность только если выполняются определённые предположения на структуру модели.

⟨45⟩ Определение стационарности “на глаз”

Характерная особенность стационарного процесса АРСС — быстрое убывание зависимости $\xi(t)$ и $\xi(t+h)$ с ростом h . Этим можно воспользоваться в качестве критерия стационарности процесса по выборочной траектории. Для этого необходимо:

- 1) Построить график $\hat{r}(h)$ для $h \in \{1, \dots, H\}$ для H существенно большего, чем ожидаемое значение $\max p, q$. Например, можно положить $H = 10$.
- 2) Если $\hat{r}(h)$ существенно не убывает (например, всегда больше 0.5), то рассматривать это как признак отсутствия стационарности.

⟨46⟩ Критерий KPSS

В рамках критерия предполагается, что для ξ верна модель:

$$\xi(t) = b_0 + \eta(t) + u(t),$$

где

- 1) $\eta(t) = \eta(t-1) + \varepsilon(t)$.
- 2) ε — белый шум с $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$.
- 3) u — некоторый стационарный процесс.

η называют случайным блужданием (дискретный аналог винеровского процесса). При $\sigma_\varepsilon^2 = 0$ процесс η вырождается в константу, а $\xi(t)$ “превращается” в стационарный процесс.

Критерий KPSS проверяет гипотезу $H_0: \sigma_\varepsilon^2 = 0$. Статистика критерия имеет

$$T = \frac{1}{n^2 \hat{\sigma}^2} \sum_{t=1}^n \left(\sum_{l=1}^t (\xi(l) - \hat{\mu}_\xi) \right)^2,$$

где

- 1) $\hat{\mu}_\xi$ — выборочное математическое ожидание ξ .
- 2) $\hat{\sigma}^2$ — некоторая состоятельная оценка дисперсии.

Доказано, что при верности H_0

$$T \rightsquigarrow \int_0^1 \beta(t) dt,$$

где β — броуновский мост, т.е. $\beta(t) = w(t) - tw(1)$, где w — стандартный винеровский процесс.

6.8 Прогнозирование процесса APSS

⟨47⟩ Пусть

- 1) $\xi \sim ARMA(p, q)$.
- 2) Значения параметров $p, q, \varphi, \theta, a_0, \sigma_\varepsilon^2$ считаются уже оценёнными или известными.
- 3) Дана выборочная траектория $\xi(1), \dots, \xi(n)$.
- 4) Значения $\varepsilon(1), \dots, \varepsilon(n)$ считаются или уже оценёнными, или известными.

В качестве прогноза можно рассматривать условное распределение

$$(\xi(t) \mid \xi(1), \dots, \xi(n), \varepsilon(1), \dots, \varepsilon(n))$$

при $t > n$.

⟨48⟩ Обозначим за $\mathcal{A} = \{\xi(1), \dots, \xi(n), \varepsilon(1), \dots, \varepsilon(n)\}$ — условие (минимальную σ -алгебру, порождённую выборочной траекторией и белым шумом). Не вдаваясь в строгое математическое обоснование, про значения $\xi(t)$ с $t > n$ при фиксированном \mathcal{A} можно сказать, что они формируются путём добавления к константе, обусловленной \mathcal{A} , последовательности независимых нормальных случайных величин $\varepsilon(n+1), \dots, \varepsilon(t)$. Следовательно,

$$(\xi(t) \mid \mathcal{A}) \sim \mathcal{N}(m_t, s_t^2),$$

и прогнозирование сводится к нахождению условных математического ожидания m_t и дисперсии s_t^2 .

⟨49⟩ Выразим значения m_t . Пусть $t > n$. Имеем

$$m_t = \mathbf{E}(\xi(t) \mid \mathcal{A}) = \sum_{i=1}^p \varphi_i \mathbf{E}(\xi(t-i) \mid \mathcal{A}) - \sum_{j=1}^q \theta_j \mathbf{E}(\varepsilon(t-j) \mid \mathcal{A}) + a_0 + \mathbf{E}(\varepsilon(t) \mid \mathcal{A}).$$

Исходя из свойств условного математического ожидания получаем

$$\mathbf{E}(\xi(t-i) \mid \mathcal{A}) = \begin{cases} \xi(t-i), & \text{если } t-i \leq n; \\ m_{t-i}, & \text{если } t-i > n; \end{cases}$$

и

$$\mathbf{E}(\varepsilon(t-j) \mid \mathcal{A}) = \begin{cases} \varepsilon(t-j), & \text{если } t-j \leq n; \\ 0, & \text{если } t-j > n; \end{cases}$$

⟨50⟩ Выразим значения s_t^2 . Пусть $t > n$. Выражение под знаком условной дисперсии можно представить в виде

$$s_t^2 = \mathbf{D}(\xi(t) \mid \mathcal{A}) = \mathbf{D}(A + \mathcal{B} \mid \mathcal{A}),$$

где

- 1) A содержит в себе все слагаемые, состоящие из констант и случайных величин из \mathcal{A} .
- 2) \mathcal{B} содержит слагаемые, зависящие от $\xi(t)$ и $\varepsilon(t)$ с $t > n$.

Заменяя рекурсивно в \mathcal{B} все вхождения $\xi(t)$ с $t > n$ по модельной формуле (2), мы придём к представлению вида

$$s_t^2 = \mathbf{D} \left(A + \sum_{l=n+1}^t b_l \varepsilon(l) \right) = \sum_{l=n+1}^t b_l^2 \sigma_\varepsilon^2.$$

6.9 Прогнозирование процесса АРПСС

⟨51⟩ Пусть

- 1) $\xi \sim ARIMA(p, d, q)$.
- 2) Значения параметров $p, d, q, \varphi, \theta, a_0, \sigma_\varepsilon^2$ считаются уже оценёнными или известными.
- 3) Дана выборочная траектория $\xi(1), \dots, \xi(n)$.
- 4) $\eta(t) = \nabla^d \xi(t)$ для $t \in \{d+1, \dots, n\}$.
- 5) Значения $\varepsilon(d+1), \dots, \varepsilon(n)$ белого шума, соответствующего η , считаются или уже оценёнными, или известными.

В качестве прогноза рассматривается условное распределение

$$(\xi(t) \mid \xi(1), \dots, \xi(n), \eta(d+1), \dots, \eta(n), \varepsilon(d+1), \dots, \varepsilon(n))$$

при $t > n$. Обозначим условие за \mathcal{A} .

⟨52⟩ Верно представление

$$\xi(t) = \eta(t) + \sum_{i=1}^d b_i \xi(t-i). \quad (53)$$

Рассуждая аналогично случаю АРСС, можно придти к выводу, что верно представление

$$(\xi \mid \mathcal{A}) = C + \sum_{l=n+1}^t c_l \varepsilon(l),$$

что означает $(\xi(t) \mid \mathcal{A}) \sim \mathcal{N}(\tilde{m}_t, \tilde{s}_t^2)$ при $t > n$.

⟨54⟩ Выразим \tilde{m}_t . Исходя из (53) имеем

$$\tilde{m}_t = \mathbf{E}(\xi(t) \mid \mathcal{A}) = \mathbf{E}(\eta(t) \mid \mathcal{A}) + \sum_{i=1}^d b_i \mathbf{E}(\xi(t-i) \mid \mathcal{A}),$$

где

$$\mathbf{E}(\eta(t) \mid \mathcal{A}) = \begin{cases} \eta(t), & \text{если } t \leq n; \\ m_t, & \text{если } t > n; \end{cases}$$

и

$$\mathbf{E}(\xi(t-i) \mid \mathcal{A}) = \begin{cases} \xi(t-i), & \text{если } t-i \leq n; \\ \tilde{m}_t, & \text{если } t-i > n; \end{cases}$$

а m_t — условное математическое ожидание прогноза процесса АРСС (см. предыдущий раздел).

⟨55⟩ Выразим \tilde{s}_t^2 . Как и в случае АРСС, всё сводится к представлению будущих значений процесса в виде суммы элементов белого шума. Если представить $\xi(t)$ в виде

$$\xi(t) = f(\xi(1), \dots, \xi(n), \eta(d+1), \dots, \eta(n), \varepsilon(d+1), \dots, \varepsilon(n)) + \sum_{l=n+1}^t b_l \varepsilon(l),$$

то, очевидно,

$$\tilde{s}_t^2 = \mathbf{D}(\xi(t) \mid \mathcal{A}) = \sigma_\varepsilon^2 \sum_{l=n+1}^t b_l^2.$$

7 Процесс Пуассона

⟨1⟩ **Определение.** Случайный процесс $\xi(t)$ называется пуассоновским процессом с параметром интенсивности λ , если:

- 1) $T = [0; +\infty)$ — непрерывное время.
- 2) $\xi(0) = 0$.
- 3) $\xi(t)$ — однородный процесс с независимыми приращениями.
- 4) $(\xi(t+h) - \xi(t)) \sim \text{Poisson}(h\lambda)$ — пуассоновская случайная величина.

□

⟨2⟩ Пусть выборка представляет собой значения выборочной траектории $\xi(t_1), \dots, \xi(t_n)$, $t_1 \leq \dots \leq t_n$. Для более компактной записи положим $t_0 = 0$. Функция правдоподобия относительно такой выборки имеет вид

$$L(\lambda) = \mathbf{f}(\xi(t_1), \dots, \xi(t_n); \lambda) = \prod_{i=1}^n \frac{((t_i - t_{i-1})\lambda)^{x_i - x_{i-1}}}{(x_i - x_{i-1})!} e^{-(t_i - t_{i-1})\lambda}.$$

Оценка максимума правдоподобия

$$\hat{\lambda} = \frac{\xi(t_n)}{t_n}.$$

⟨3⟩ Пусть процесс наблюдается до момента t^* . В ходе эксперимента наблюдаются моменты возникновения пуассоновских событий $\mathcal{T}_1, \dots, \mathcal{T}_\nu$, где $\nu = \xi(t^*)$ — количество пуассоновских событий, произошедших к моменту t^* . Функция правдоподобия относительно такой выборки имеет вид

$$L(\lambda) = \mathbf{f}(\mathcal{T}_1, \dots, \mathcal{T}_\nu, \nu; \lambda) = \mathbf{f}(\mathcal{T}_1, \dots, \mathcal{T}_\nu \mid \nu) \mathbf{f}(\nu; \lambda) = \frac{\nu!}{(t^*)^\nu} \frac{(t^*\lambda)^\nu}{\nu!} e^{-t^*\lambda}.$$

Оценка максимума правдоподобия

$$\hat{\lambda} = \frac{\xi(t_n)}{t_n}.$$

8 Винеровский процесс

⟨1⟩ **Определение.** Случайный процесс $\xi(t)$ называется винеровским процессом со сносом μ и коэффициентом диффузии σ^2 , если:

- 1) $T = [0; +\infty)$ — непрерывное время.
- 2) $\xi(0) = 0$.
- 3) $\xi(t)$ — однородный процесс с независимыми приращениями.
- 4) $(\xi(t+h) - \xi(t)) \sim \mathcal{N}(h\mu, h\sigma^2)$ — нормальная случайная величина.

□

⟨2⟩ Пусть выборка представляет собой значения выборочной траектории $\xi(t_1), \dots, \xi(t_n)$, $t_1 \leq \dots \leq t_n$. Для более компактной записи положим $t_0 = 0$. Функция правдоподобия относительно такой выборки имеет вид

$$\begin{aligned} L(\mu, \sigma^2) &= \mathbf{f}(\xi(t_1), \dots, \xi(t_n); \mu, \sigma^2) = \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi(t_i - t_{i-1})\sigma^2}} \exp \left\{ - \frac{(\xi(t_i) - \xi(t_{i-1}) - (t_i - t_{i-1})\mu)^2}{2(t_i - t_{i-1})\sigma^2} \right\}. \end{aligned}$$

Оценка максимума правдоподобия

$$\hat{\mu} = \frac{\xi(t_n)}{t_n}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \frac{(\xi(t_i) - \xi(t_{i-1}) - (t_i - t_{i-1})\hat{\mu})^2}{t_i - t_{i-1}}.$$

9 Общие сведения о работе с данными в Python

В этом разделе приводится очень краткий обзор библиотек, на основе которых можно выполнять вычислительные части заданий курса. Здесь приводятся лишь названия основных методов и типов, а также их примерная связь между собой. Для того, чтобы суметь написать реальный код, понадобится смотреть подробную справку по названным методам или по этим названиям искать примеры в Интернете.

⟨1⟩ При анализе данных, работе с моделями машинного обучения и итеративном программировании может быть удобным заниматься разработкой в Jupyter Notebook'ах. Основное преимущество такого подхода — можно запускать дополнительные программы без необходимости перезагрузки всей текущей программы. Это сильно ускоряет работу методом научного тыка. Стандартное расширение для Notebook файлов: `.ipynb`.

Варианты способов использования Notebook'ов:

- 1) Установить JupyterLab, который представляет собой локально запускаемый web-сервис, позволяющий открывать в окне браузера локально находящиеся Jupyter Notebook'и.
- 2) Использовать встроенную поддержку в VS Code. VS Code самостоятельно устанавливает и запускает сервер Jupyter Notebook'ов и позволяет работать с ним через свой интерфейс.
- 3) Использовать Google Colab — этот web-сервис использует клон Jupyter Notebook'ов в качестве интерфейса взаимодействия с пользователем.

⟨2⟩ При работе в Python основные библиотеки для численных расчётов и визуализации:

- 1) `numpy` — эффективно упакованные числовые массивы (с произвольной размерностью), векторные и почленные операции, реализованные на компилированном языке (что позволяет достигать высокой производительности операций).
- 2) `matplotlib.pyplot` — самая, кажется, распространённая библиотека для рисования графиков в Python. Интерфейс и принципы взаимодействия схожи с аналогами в Matlab / Octave, с R и т.п.
- 3) `scipy` — набор математических функций и алгоритмов. Для нас, кажется, будут наиболее интересны методы для вычисления функций плотности различных распределений и методы численной оптимизации / минимизации / максимизации функции.

9.1 Краткий обзор numpy

⟨3⟩ При работе с `numpy` его часто импортируют с псевдонимом `np` (команда `import numpy as np`). В дальнейшем будем предварять функции и типы из этой библиотеки префиксом `np..`

Основой интересующий нас тип данных — это `np.array`. Тип является массивом с заданным фиксированным размером и количеством измерений и единым фиксированным числовым типом данных. Если заранее известно, какой нам понадобится размер массива, то рекомендуется заранее создать его с нужным размером командой `np.zeros`.

К элементам массива типа `np.array` можно обращаться по индексу. Например, если `x` — это 3-х мерный массив, то доступ к его элементам можно осуществлять командой `x[i1, i2, i3]`, где `i1, i2, i3` — номера координат элемента вдоль соответствующих измерений с нумерацией от нуля. В качестве индекса (в качестве `i1, i2` или `i3` в примере) может выступать

- 1) знак `:` — выбрать все элементы из измерения.
- 2) массив индексов — одновременный выбор нескольких элементов с указанными значениями в измерении.

Т.к. `Python` — интерпретируемый язык, то работа с массивами в циклах (в `for` / `while`) работает медленно, в тысячи или даже сотни тысяч раз медленнее, чем можно было бы ожидать в, например, программе на `C` / `C++`. В связи с этим, когда возможно, трансформировать и преобразовывать значения в массивах желательно сразу скопом, используя встроенные в `numpy` операции.

Основная часть преобразования делается через векторные по-элементные операции над массивами. Например, если `a` и `b` — два массива одинаковой размерности, то `a + b` будет выдавать массив той же размерности и с элементами, равными сумме соответствующих элементов (с теми же индексами) в `a` и `b`. Аналогично `a * b` — по-элементное перемножение, `a ** b` — по-элементное возведение элементов из `a` в степени порядков, указанных в соответствующих элементах в `b`. Кроме того, во всех этих операциях `a` или `b` может быть числом — тогда `numpy` “притворяется”, что это не одно число, а массив с совместимой размерностью, заполненный этим числом.

Операция матричного перемножения в `numpy` вызывается через `np.matmul` и `np.dot`. Другие варианты семантики матричного умножения представлены в функциях и методах типа `inner` / `outer`.

Другие три, как мне кажется полезные, способа создания заполненных `np.array`:

- 1) `np.fill` — создание массива, наполненного заданным числом.
- 2) `np.arange` — создание одномерного массива (вектора), заполненного (в частности) последовательностью всех целых чисел от 0 до заданного числа `n` (без включения). Возможно, также, явно задать начальное число и шаг изменения числа в последовательности
- 3) `np.linspace` — создание одномерного массива (вектора), заполненного сектой чисел заданного интервала и с заданным количеством элементов.

Модуль `np.random` содержит функции для генерирования массивов заданной величины, заполненных независимыми реализациями заданных распределений вероятностей. Например, функция `np.random.normal` позволяет генерировать выборки из нормального

распределения. Для более быстрой работы программы, если заранее известно, сколько случайных величин понадобится для дальнейших вычислений, желательно их генерировать одним запуском функции генерации с заданием подходящей размерности выходного массива.

9.2 Краткий обзор `matplotlib.pyplot`

⟨4⟩ При работе с `matplotlib.pyplot` его часто импортируют с псевдонимом `plt` (команда `import matplotlib.pyplot as plt`). В дальнейшем будем предварять функции и типы из этой библиотеки префиксом `plt..`

Основная интересующая нас команда: `plt.plot`. Базовые способы запуска:

- 1) `plt.plot(y)`, где `y` — одномерный `np.array`, — рисует график в виде ломанной прямой. `x` координаты точек получаются как набор целых чисел от 0 до количества элементов в `y` (т.е. `y.shape[0]` или `len(y)`), а в качестве `y` координат — соответствующие элементы из `y`.
- 2) `plt.plot(x, y)`, где `x, y` — одномерные `np.array`. Аналогично предыдущему, только теперь `x` координаты берутся из вектора `x`.

Другие параметры команды `plt.plot` контролируют цвет, толщину линий, оформление точек ломанной кривой, тип графика (например, переключение от ломанной прямой к отображению лишь точек).

Если на одном графике необходимо отобразить несколько кривых, то достаточно последовательно запустить несколько `plt.plot` команд. Для того, чтобы нарисовать финальный график может понадобится вызвать команду `plt.show`. Более расширенный и явный способ совмещения нескольких линий, графиков и объектов на одной картинке основывается на:

- 1) Создании объекта типа `Figure` — объект-холст, который содержит в себе все остальные элементы.
- 2) Создание в `Figure` одного или нескольких объектов `Axis` — подграфик со своей системой координат и своими графическими объектами. Также в `Figure` могут быть добавлены вспомогательные объекты, например, `Legend` — список подписей к разным элементам графика.
- 3) Линии и точки рисуются вызовами методов, например, `plot` у объекта типа `Axis`.
- 4) Также у `Axis` есть методы, позволяющие задавать подпись к подграфику, названия осей и т.п.
- 5) После того, как `Figure` полностью наполнен необходимым содержанием, он отрисовывается методом `show` или сохраняется в файл методом `savefig`.

9.3 Краткий обзор `scipy`

⟨5⟩ Нас будут интересовать методы и типы из модуля `scipy.stats`. Этот модуль содержит набор типов, соответствующих различным распределениям вероятностей. Создание объекта соответствующего типа позволяет указать параметры распределения, а методы объекта позволяют вычислять различные характеристики этого распределения (функцию плотности, функцию распределения, квантили и т.п.)

Например, тип `scipy.stats.norm` указывает на нормальное распределение, аргументы конструктора типа задают среднее (μ) и стандартное отклонение (σ) нормального распределения. Полезные для нас сейчас методы:

- 1) `cdf` — функция распределения. `pdf` — функция плотности. `ppf` — квантили (обратная функция распределения).

10 Ошибка оценивания и ошибка прогноза при неслучайном параметре

⟨1⟩ Пусть мы имеем дело со случайным процессом $\xi \sim \mathcal{F}(\theta)$, где θ — некоторый неслучайный неизвестный параметр. Предположим, что мы хотим вычислить некую функцию $h(\theta)$, зависящую от θ (например, спрогнозировать будущие значения ξ исходя из известных текущих). Пусть даны выборочные траектории D и мы получили по ним оценку $\hat{\theta}$ параметра θ . Тогда в качестве естественной оценки значения функции h выступает $h(\hat{\theta})$.

Возникает характерный для математической статистики вопрос — насколько точна оценка $h(\hat{\theta})$? Если представить, что $\hat{\theta} = \theta + \varepsilon$, то вместо $h(\theta)$ мы вычисляем $h(\theta + \varepsilon)$, и величина отклонения $h(\theta) - h(\theta + \varepsilon)$ зависит как от свойств функции h , так и от вероятностных свойств ошибки оценивания ε .

Мы рассмотрим несколько методов оценки отклонения $h(\hat{\theta})$ от истинного значения $h(\theta)$. Во всех случаях будем формироваться “выборка” $\theta_1^*, \dots, \theta_k^*$ со сколь угодно большим k предполагаемых значений θ на основе выборочных данных D . Значения $h(\theta_1^*), \dots, h(\theta_k^*)$ образует “выборку”, которую будем интерпретировать как набор независимых реализаций случайной величины $h(\hat{\theta})$. По значениям $\{h(\theta_i^*)\}$ обычными статистическими методами возможно оценить любые вероятностные свойства распределения $h(\hat{\theta})$.

⟨2⟩ Далее мы рассмотрим несколько схем генерации значений $\{\theta_i^*\}$.

10.1 Метод моделирования реализаций оценки $\hat{\theta}$ при фиксированном θ

⟨3⟩ Зададим $\theta = \theta_0$. Возможны две ситуации:

- 1) Если распределение $\hat{\theta} \sim G(\theta_0)$ известно, то в качестве значений $\{\theta_i^*\}$ задаём k независимых реализаций из распределения $G(\theta_0)$.
- 2) Если распределение $\hat{\theta}$ неизвестно, то нужно сгенерировать k (D_1, \dots, D_k) независимые реализации D в эксперименте с параметром $\theta = \theta_0$. Полагаем $\theta_i^* = \hat{\theta}(D_i)$.

⟨4⟩ Обычно нам не известно истинное параметра θ , а только его оценка $\hat{\theta}$. При этом при анализе разброса оценки нам необходимо зафиксировать некоторое значение θ . Естественно положить $\theta = \hat{\theta}$. Если, однако, истинное значение θ существенно отличается от $\hat{\theta}$, а распределение ϵ сильно асимметрично, то такой метод может давать большую погрешность.

10.2 Метод моделирования предполагаемых значений θ относительно известной оценки $\hat{\theta}$

⟨5⟩ Пусть дана выборка D , по которой нами получена оценка $\hat{\theta}_0$. Пусть $\hat{\theta}_1^*, \dots, \hat{\theta}_k^*$ — выборка реализаций оценки $\hat{\theta}$ в экспериментах с $\theta = \hat{\theta}_0$, полученная аналогично как в п. 3. Представим значения оценок через отклонения от истинных значений:

$$\hat{\theta}_0 = \theta + \epsilon, \quad \hat{\theta}_i^* = \hat{\theta}_0 + \epsilon_i^*, \quad \text{где } \epsilon \sim \mathcal{G}(\theta), \quad \epsilon_i^* \sim \mathcal{G}(\hat{\theta}_0).$$

Обычно можно ожидать, что распределения ошибок оценивания $\mathcal{G}(\theta)$ и $\mathcal{G}(\hat{\theta}_0)$ будут мало отличаться, если θ и $\hat{\theta}_0$ близки.

В силу предполагаемой близости распределений $\{\epsilon_i^*\}$ можно воспринимать как выборку из ϵ . Отсюда получаем формулу для формирования выборки $\{\hat{\theta}_i\}$ предполагаемых значений θ :

$$\theta = \hat{\theta}_0 - \epsilon \quad \Rightarrow \quad \hat{\theta}_i = \hat{\theta}_0 - \epsilon_i^* = 2\hat{\theta}_0 - \hat{\theta}_i^*.$$

⟨6⟩ По построению видно, что этот метод иногда может порождать $\hat{\theta}_i \notin \Theta$. В этом случае приходится корректировать значения $\hat{\theta}_i$.

⟨7⟩ **Замечание.** Попробуем интерпретировать идею метода. Когда мы получаем оценку $\hat{\theta}_0$ по выборке D при истинном значении параметра $\theta = \theta_0$, её отклонение от θ_0 определяется реализацией случайной величины ϵ — величины случайного отклонения оценки от истинного значения параметра. В частности, если бы мы знали ϵ , то могли бы “исправить” оценку $\hat{\theta}_0$ на величину ϵ и получили бы точное значение параметра, т.е. $\hat{\theta}_0 - \epsilon = \theta_0$. Но значения ϵ мы в реальных задачах не знаем.

Полезно хотя бы знать распределение вероятностей ϵ — тогда возможно осознавать, какие отклонения $\hat{\theta}$ от θ наиболее вероятные, а, значит, предполагать наиболее вероятные значения для θ исходя из имеющегося значения $\hat{\theta}$.

Распределение ϵ можно оценить методом Монте-Карло генерируя значения при некотором задаваемом нами значении параметра θ_1 . Распределение ϵ может зависеть от значения θ , поэтому симуляцию желательно проводить со значением θ_1 как можно более близким к θ_0 (которое мы не знаем). Поэтому естественно положить $\theta_1 = \hat{\theta}_0$. Наблюдая, насколько в симуляции оценка $\hat{\theta}$ отклоняется от θ_1 , мы можем предположить, что подобного же рода отклонение произошло и при оценивании θ_0 по D . \square

10.3 Метод моделирования предполагаемых значений θ исходя из функции правдоподобия

⟨8⟩ Мы можем притвориться, что неслучайный параметр $\theta \in \Theta$ — это на самом деле реализация случайного параметра $\vartheta \sim Uniform(A)$, $A \subseteq \Theta$ с A достаточно большим, чтобы охватить все интересующие нас значения θ , но с конечной мерой. Тогда функция $f(\vartheta)$ константна и

$$f(\vartheta | \mathcal{D} = D) = \frac{f(\mathcal{D} = D | \vartheta) f(\vartheta)}{f(\mathcal{D} = D)} = cL(\vartheta; D),$$

где $L(\vartheta; D) = f(\mathcal{D} = D | \vartheta)$ — функция правдоподобия, а c — некий нормирующий множитель, не зависящий от значения ϑ . Т.е. $L(\vartheta; D)$ — это апостериорная функция плотности случайного параметра ϑ с точностью до нормирующей константы.

В предположении случайности ϑ апостериорная функция плотности параметра даёт нам достоверную информацию о том, каким может быть значение параметра ϑ исходя из имеющихся у нас выборочных данных D . Поэтому в этих предположениях естественно в качестве θ_i^* взять реализации случайной величины, задаваемой апостериорной функцией плотности $f(\vartheta | \mathcal{D} = D) = cL(\vartheta; D)$.

⟨9⟩ Для генерации выборок из многомерных распределений или распределений, функция плотности которых известна с точностью до нормирующего множителя, обычно, применяют метод Монте-Карло на марковских цепях (Markov chain Monte Carlo, сокращённо МСМС). Генерация из $L(\vartheta; D)$ — как раз тот случай. Алгоритмы МСМС поясняются в разделе 11.

⟨10⟩ **Замечание.** Рассмотрение вместо неслучайного параметра θ байесовской модели со случайным ϑ , конечно же, является волюнтаристским решением, неким удобным приближением, которое в конкретных задачах может приводить к искажениям в выводах. Тем не менее равномерное априорное распределение постулирует отсутствие у нас каких-либо предпочтений в отношении значений θ — каждое значение $\theta \in \Theta$ имеет тот же “вес” и настолько же ожидаемо, что и любое другое значение параметра. В этом отношении байесовская модель с равномерным априорным является близким аналогом постановки с неслучайным параметром. Ещё один довод в пользу близости — стандартные интервальные оценки параметра совпадают для случая неслучайного параметра и рассматриваемого случая случайного параметра с равномерным распределением. \square

11 Монте-Карло на марковских цепях. Метод Метрополиса-Гастингса

⟨1⟩ Методами Монте-Карло на марковских цепях называют большое семейство алгоритмов генерации случайных величин по заданной функции плотности $f(x)$. Пусть $f(x)$ задаёт распределение случайного вектора размерности p . Алгоритмы сильно отличаются друг от друга по механике и начальным условиям, но есть ряд объединяющих их признаков:

- 1) Генерируемая выборка образует марковский случайный процесс с дискретным временем ξ , $T = \{1, 2, \dots\}$.

- 2) Алгоритмы задают переходные распределения таким образом, чтобы у ξ существовало эргодическое распределение и оно совпадало с $\mathbf{f}(x)$.
- 3) Требуется задание произвольного начального значения, т.е. значения для $\xi(1)$.
- 4) Элементы сгенерированной выборки зависимы от соседних элементов.

Мы рассмотрим только метод Метрополиса-Гастингса (Metropolis-Hastings sampler).

⟨2⟩ Пусть дана функция плотности

$$\mathbf{f}(x) = c q(x),$$

где нормирующий множитель $c > 0$ сложно вычислим и неизвестен. Стоит задача — сгенерировать выборку из \mathbf{f} .

В рамках метода Метрополиса-Гастингса выборка формируется в виде случайного процесса ξ , $T = \{1, 2, \dots\}$. Пусть

- 1) $\eta_i \sim \text{Uniform}([0; 1])$, $i \in \{2, 3, \dots\}$ и $\{\eta_i\}$ независимы в совокупности.
- 2) $(\gamma_i | \xi(i-1)) \sim \mathcal{G}(\xi(i-1))$, $i \in \{2, 3, \dots\}$ и γ_i при фиксированном $\xi(i-1)$ не зависит от $\{\gamma_j: j < i\}$ и $\{\xi(j): j < i-1\}$. Величины γ_i называют предлагаемыми значениями (proposed value).

Перед применением метода должны быть заданы параметры:

- 1) Начальное значение $\xi(1)$.
- 2) Условная функция плотности

$$g(\gamma_i, \xi(i-1)) = \mathbf{f}(\gamma_i | \xi(i-1)) = \mathbf{f}(\gamma_i; \mathcal{G}(\xi(i-1))),$$

задающее так называемое распределение предлагаемых значений (proposal distribution).

Тогда, согласно методу, полагаем для $i \in \{2, 3, \dots\}$

$$\xi(i) = \gamma_i \mathbb{I}(A_i) + \xi(i-1) (1 - \mathbb{I}(A_i)), \quad \text{где} \quad A_i = \left\{ \eta_i \leq \frac{g(\xi(i-1), \gamma_i) q(\gamma_i)}{g(\gamma_i, \xi(i-1)) q(\xi(i-1))} \right\}.$$

Событие A_i называют принятие предлагаемого значения γ_i в качестве нового элемента выборки $\xi(i)$ (acceptance). В случае непринятия, новый элемент выборки принимается равным прошлому значению $\xi(i)$.

⟨3⟩ Реализация на компьютере. При проведении большого количества операций умножения и деления чисел, а также при работе с большим количеством маленьких дробных чисел, часто возникают сложности с точностью вычислений. Поэтому часто бывает выгоднее производить вычисления с логарифмами величин, чем с исходным выражением — часто благодаря этому большая часть операций произведения и возведения в степень превращаются в более щадящие и быстро вычисляемые сложения и произведения, соответственно.

В наших задачах при вычислении A_i рекомендуется применять логарифмированную версию:

$$A_i = \left\{ \ln \eta_i \leq \ln g(\xi(i-1), \gamma_i) - \ln g(\gamma_i, \xi(i-1)) + \ln q(\gamma_i) - \ln q(\xi(i-1)) \right\}. \quad \square$$

⟨4⟩ Если известен $\arg \max_x \mathbf{f}(x)$ (как бывает при оценивании методом максимального правдоподобия или методом максимума апостериорного риска), то в качестве начального значения лучше взять его: $\xi(1) = \arg \max_x \mathbf{f}(x)$.

⟨5⟩ Скорость работы и степень независимости элементов в выходной выборке $\xi(1), \xi(2), \dots$ сильно зависит от выбора распределения предлагаемых значений $g(\gamma_i, \xi(i-1))$ и его “соответствия” распределению $q(x)$, из которого мы хотим сгенерировать выборку.

В качестве хорошего варианта по умолчанию в литературе рекомендуют использовать нормальный вектор с независимыми компонентами:

$$\mathcal{G} = \mathcal{N}(\mu = \xi(i-1), \Sigma = \sigma^2 I_{p \times p}),$$

где p — размерность генерируемой случайной величины (случайного вектора, распределение которого задаёт функция плотности $\mathbf{f}(x)$); $I_{p \times p}$ — единичная матрица размерности $p \times p$; σ^2 — параметр дисперсии, который в этом методе обретает смысл величины шага случайного блуждания. При таком выборе $g(x, y) = g(y, x)$, так что $A_i = \{\eta_i \leq q(\gamma_i) / q(\xi(i-1))\}$.

При выборе σ^2 (и \mathcal{G} в целом) рекомендуют ориентироваться на величину частоты принятия (acceptance rate)

$$\alpha = \lim_{n \rightarrow \infty} \frac{\sum_{i=2}^n \mathbb{I}(A_i)}{n-1},$$

т.е. на долю предлагаемых значений, которые принимаются в качестве выходных выборочных значений. В литературе считается, что для хорошей работы метода желательно, чтобы $\alpha \in [0.50; 0.85]$. При меньших или больших α будет увеличиваться степень зависимости значений в выходной выборке $\xi(1), \xi(2), \dots$.

⟨6⟩ Значения в выборке $\xi(1), \dots, \xi(n)$, получаемой из метода, формально говоря, являются зависимыми. Однако, зависимость между $\xi(i)$ и $\xi(j)$ падает с ростом расстояния $|i-j|$. Поэтому при достаточно большом объёме выборки n (достаточная величина зависит от удачности выбора \mathcal{G}) после перемешивания выборку $\xi(1), \dots, \xi(n)$ можно считать приближённо независимой (и степень независимости будет расти с ростом n).

⟨7⟩ **Алгоритм.** Подытоживая, мы будем использовать следующий алгоритм генерации, в котором предполагается выбор нормального распределения предложений и наличие знания $\arg \max_x q(x)$ или его приближения:

- 1) Выбираем требуемый объём генерируемой выборки n .
- 2) Фиксируем некоторое значение σ^2 — дисперсии распределения предложений. Для выбора σ^2 может понадобиться совершить несколько пробных генераций, чтобы добиться значения частоты принятия $\alpha \in [0.50; 0.85]$.
- 3) $\xi(1) \leftarrow \arg \max_x q(x)$.
- 4) Генерируем значения $\eta_j \sim \text{Uniform}([0; 1])$ и многомерного нормального вектора с независимыми компонентами $\gamma_j \sim \mathcal{N}(0, \sigma^2 I_{p \times p})$ для $j \in \{2, \dots, n\}$.
- 5) $i \leftarrow 2$.

6) Если $\ln \eta_i \leq \ln q(\gamma_i + \xi(i-1)) - \ln q(\xi(i-1))$, то $\xi(i) \Leftarrow \gamma_i + \xi(i-1)$; иначе $\xi(i) \Leftarrow \xi(i-1)$.

7) Если $i = n$, то завершаем работу алгоритма; иначе $i \Leftarrow i + 1$ и переходим к пункту 6). □

⟨8⟩ Обосновать корректность метода можно показав, что стационарное распределение марковского процесса ξ совпадает с целевым распределением $\mathbf{f}(x)$. Это можно показать, подставив $\mathbf{f}(x)$ и $\mathbf{f}(\xi(t+1) | \xi(t))$ в определение стационарности и показав выполнение соответствующего равенства. Т.к. нужно уметь орудовать со стационарностью марковских процессов с непрерывными значениями, то здесь доказательство приводить не будем. Обоснование можно посмотреть в [1, п. 27.4.2].

⟨9⟩ **Замечание.** В том случае, когда начальное значение $\xi(1)$ выбирается случайно, оно может оказаться в области с низкой вероятностью рассматриваемого распределения \mathbf{f} . В этом случае первые сгенерированные элементы выборки могут неадекватно представлять значения из \mathbf{f} . В связи с этим, при случайном выборе $\xi(1)$ первые элементы выходной выборки принято отбрасывать, то момента стабилизации распределения. Этот процесс называют прогревом (burn-in).

В качестве метода для определения, с какого момента наблюдения в выборке уже “настоящие”, запускают сразу несколько параллельно существующих процесса генерации выборки с разными (случайно выбранными) начальными значениями. Как только графики значений $q(x)$ на сгенерированных значениях начинают “перемешиваться” — это выступает признаком, что выборки во всех процессах генерации обрели “одинаковое” распределение (и, можно надеяться, желаемое нами). С этого момента последующие сгенерированные элементы выборки считаются годными для дальнейшего использования. □

Часть III

Задания

12 Задание 1

⟨1⟩ Дано.

- 1) Рассматривается марковская цепь $\xi \sim MC(\pi, P)$, $T = \{1, 2, \dots\}$.
- 2) В файле `data.csv` даны несколько независимых выборочных траекторий процесса ξ . Столбцы в `data.csv`:
 - 1) `sample` — номер траектории.
 - 2) `t` — момент наблюдения (из T).
 - 3) `x` — значение процесса в этот момент времени.

□

⟨2⟩ Задание.

- 1) Нарисуйте (в каком-то виде) графики траекторий из `data.csv`.
- 2) Получите в явном виде оценки максимального правдоподобия для π и P по нескольким независимым выборочным траекториям.
- 3) Оцените π и P по методу максимального правдоподобия исходя из данных в `data.csv`.
- 4) Пусть X_1, \dots, X_n — значения из первой выборочной траектории (n — последний момент времени, для которого присутствует наблюдение в выборочной траектории). Исходя из полученных оценок для π и P вычислите $\mathbf{P}(\xi(n+1) = x \mid X_1, \dots, X_n)$ и $\mathbf{P}(\xi(n+2) = x \mid X_1, \dots, X_n)$ для всех $x \in \text{dom } \xi$.

□

13 Задание 2

⟨1⟩ Дано. Рассматривается постановка и данные Задания 1.

□

⟨2⟩ Задание.

- 1) Суть задания — оценить степень ошибки предсказания $\mathbf{P}(\xi(n+1) = x \mid X_1, \dots, X_n)$ и $\mathbf{P}(\xi(n+2) = x \mid X_1, \dots, X_n)$, вызванной неточностью оценивания параметров модели марковской цепи π и P .
- 2) Сформируйте выборку предполагаемых значений $\theta_1^*, \dots, \theta_k^*$ параметра $\theta = (\pi, P)$ с $k = 1000$ исходя из метода, описанного в п. 10.1. Используя $\theta_1^*, \dots, \theta_k^*$, оцените стандартные отклонения и постройте 0.9-квантильные интервалы (т.е. от 0.05 до 0.95 выборочной квантили) для оценок прогноза — вероятностей будущих значений процесса (см. задание 1)

$$\mathbf{P}(\xi(n+1) = x \mid X_1, \dots, X_n; \hat{\pi}, \hat{P}) \quad \text{и} \quad \mathbf{P}(\xi(n+2) = x \mid X_1, \dots, X_n; \hat{\pi}, \hat{P}),$$

где X_1, \dots, X_n — значения из первой выборочной траектории. Нарисуйте гистограммы для 2–3 параметров из числа $\{\pi_i\}$ и для 2–3 параметров из числа $\{p_{ij}\}$.

- 3) Сделайте то же самое, что и в 2), но используя для генерации $\theta_1^*, \dots, \theta_k^*$ метод, описанный в п. 10.2. □

14 Задание 3

⟨1⟩ Дано. Рассматривается постановка и данные Задания 1. □

⟨2⟩ Задание. Задание в целом аналогично Заданию 2, но с использованием метода Метрополиса-Гастингса (см. п. 11) для генерации выборки из распределения, порождаемого функцией правдоподобия (см. п. 10.3)

- 1) Суть задания — оценить степень ошибки предсказания $\mathbf{P}(\xi(n+1) = x \mid X_1, \dots, X_n)$ и $\mathbf{P}(\xi(n+2) = x \mid X_1, \dots, X_n)$, вызванной неточностью оценивания параметров модели марковской цепи π и P .
- 2) Сформируйте выборку предполагаемых значений $\theta_1^*, \dots, \theta_k^*$ параметра $\theta = (\pi, P)$ с $k = 10000$ исходя из функции прав, используя метод Метрополиса-Гастингса (см. п. 11). Используя $\theta_1^*, \dots, \theta_k^*$, оцените стандартные отклонения и постройте 0.9-квантильные интервалы (т.е. от 0.05 до 0.95 выборочной квантили) для оценок прогноза — вероятностей будущих значений процесса (см. задание 1)

$$\mathbf{P}(\xi(n+1) = x \mid X_1, \dots, X_n; \hat{\pi}, \hat{P}) \quad \text{и} \quad \mathbf{P}(\xi(n+2) = x \mid X_1, \dots, X_n; \hat{\pi}, \hat{P}),$$

где X_1, \dots, X_n — значения из первой выборочной траектории. Нарисуйте гистограммы для 2–3 параметров из числа $\{\pi_i\}$ и для 2–3 параметров из числа $\{p_{ij}\}$.

- 3) Сравните результат с тем, что было получено в Задании 2. □

⟨3⟩ Замечание. В этом задании мы применяем метод Метрополиса-Гастингса для генерации выборки параметров из функции правдоподобия по марковской цепи. Чтобы получить выборку с корректным распределением, важно помнить, что функция правдоподобия для марковской цепи

$$L(\pi_1, \dots, \pi_p, p_{11}, \dots, p_{1p}, p_{21}, \dots, p_{2p}, \dots, p_{p1}, \dots, p_{pp})$$

определена (имеет ненулевое значение) только при

$$\sum_{i=1}^p \pi_i = 1, \quad \sum_{j=1}^p p_{kj} = 1 \quad \forall k \in \{1, 2, \dots, p\}. \quad (4)$$

В этой ситуации, если попытаться воспользоваться методом Метрополиса-Гастингса для генераций значений напрямую для всех параметров $\pi_1, \dots, \pi_p, p_{11}, \dots, p_{pp}$, то окажется, что предлагаемые значения с нулевой вероятностью будут удовлетворять (4), а значит будут всегда отвергаться. Выход из этой ситуации — перейти к более

сжатой параметризации марковской цепи, в которой по одному параметру из каждой суммы из (4) будет выражаться линейно из оставшихся. Например, можем положить

$$\pi_p = 1 - \sum_{i=1}^{p-1} \pi_i, \quad p_{kp} = 1 - \sum_{j=1}^{p-1} p_{kj}, \quad k = 1, \dots, p.$$

Тогда функция правдоподобия в модели с сжатым параметрическим пространством выражается как

$$\begin{aligned} \tilde{L}(\pi_1, \dots, \pi_{p-1}, p_{11}, \dots, p_{1p-1}, \dots, p_{p1}, \dots, p_{pp-1}) = \\ = L\left(\pi_1, \dots, \pi_{p-1}, 1 - \sum_{i=1}^{p-1} \pi_i, p_{11}, \dots, p_{1p-1}, 1 - \sum_{i=1}^{p-1} p_{1i}, p_{21}, \dots, p_{p1}, \dots, p_{pp-1}, 1 - \sum_{i=1}^{p-1} p_{pi}\right). \quad \square \end{aligned}$$

15 Задание 4

Рассматривается задача фильтрации траектории движения объекта в двухмерном пространстве. Информация о положении объекта в каждый момент времени поступает со случайными искажениями.

⟨1⟩ Дано.

- 1) Рассматривается скрытая марковская цепь (ξ, ζ) про которую известно, что:
 - 1) $T = \{0, 1, \dots\}$.
 - 2) $\text{dom } \zeta = \mathbb{Z} \times \mathbb{Z}$, $\zeta(t) = (\zeta_x(t), \zeta_y(t))$, $\zeta \sim MC(\pi, P)$.
 - 3) $\xi(t) = (\xi_x(t), \xi_y(t))$, $(\xi(t) \mid \zeta(t)) \sim \mathcal{N}(\zeta(t) + \mu, \sigma^2 I_{2 \times 2})$, где $I_{2 \times 2}$ — единичная матрица размера 2×2 .
 - 4) $\xi_x(t)$ и $\zeta_x(t)$ имеют смысл x координат объекта, а $\xi_y(t)$ и $\zeta_y(t)$ — y координат.
 - 5) $\mathbf{P}(\zeta(0) = (0, 0)) = 1$.
 - 6) $\mathbf{P}(\zeta(t+1) = \zeta(t) + (0, 1) \mid \zeta(t)) = a_1$.
 $\mathbf{P}(\zeta(t+1) = \zeta(t) + (1, 0) \mid \zeta(t)) = a_2$.
 $\mathbf{P}(\zeta(t+1) = \zeta(t) + (0, -1) \mid \zeta(t)) = a_3$.
 $\mathbf{P}(\zeta(t+1) = \zeta(t) + (-1, 0) \mid \zeta(t)) = a_4$.
 - 7) $a_1 + a_2 + a_3 + a_4 = 1$.
- 2) В файле `train.csv` даны несколько независимых выборочных траекторий процессов ξ и ζ . Столбцы в `train.csv`:
 - 1) `sample` — номер траектории.
 - 2) `t` — момент наблюдения (из T).
 - 3) `x` — значение $\xi_x(t)$.
 - 4) `y` — значение $\xi_y(t)$.
 - 5) `hidden_x` — значение $\zeta_x(t)$.
 - 6) `hidden_y` — значение $\zeta_y(t)$.

- 3) В файле `current_visible.csv` дана выборочная траектория процесса ξ . Столбцы в `current_visible.csv`:
- 1) t — момент наблюдения (из T).
 - 2) x — значение $\xi_x(t)$.
 - 3) y — значение $\xi_y(t)$.

□

⟨2⟩ Задание.

- 1) Пусть $\theta = (a_1, a_2, a_3, a_4, \mu, \sigma^2)$ — параметры модели.
- 2) По данным из `train.csv` получите их оценку $\hat{\theta}$.
- 3) Полагая $\theta = \hat{\theta}$, для выборки из `current_visible.csv` найдите наиболее вероятную траекторию скрытого процесса (т.е. по данным значениям $\xi(0), \dots, \xi(n)$ найдите наиболее вероятную комбинацию значений $\zeta(0), \dots, \zeta(n)$). Нарисуйте график наиболее вероятной траектории с наложением значений видимого процесса.
- 4) Исходя из полученной оценки наиболее вероятной траектории скрытого процесса найдите наиболее вероятные значения для $\zeta(n+1), \zeta(n+2), \xi(n+1), \xi(n+2)$.

□

16 Задание 5 🐶

В этом задании рассматриваются вопросы оценивания математического ожидания стационарного процесса, а также оценивание и применение при анализе функций автокорреляции и кросс-ковариации. Все случайные процессы предполагаются с дискретным временем.

⟨1⟩ Дано.

- 1) Рассматриваются стационарные случайные процессы ξ и ζ . В файле `mean_1.csv` представлены значения нескольких независимых траекторий процесса ξ , а в файле `mean_2.csv` — процесса ζ . В обеих таблицах поля означают:
 - 1) `sample` — номер траектории.
 - 2) t — момент наблюдения.
 - 3) x — значение процесса.
- 2) Рассматривается стационарный случайный процесс η . В файле `periodic.csv` даны значения одной выборочной траектории. Поля таблицы:
 - 1) t — момент наблюдения.
 - 2) x — значение процесса.
- 3) Рассматривается пара стационарных случайных процессов α и β . В файле `cross.csv` даны значения одной выборочной траектории пары этих процессов. Поля таблицы:
 - 1) t — момент наблюдения.
 - 2) x — значение процесса α .

3) y — значение процесса β .

□

⟨2⟩ Задание.

- 1) Требуется оценить математическое ожидание процессов ξ и ζ по выборочным данным. Один из процессов сильно зависим — в его случае полезнее использовать оценку, полученную усреднением оценок по каждой траектории (см. оценку $\hat{\mu}_1$ из пункта 15 раздела 5.1). Другой процесс слабо зависим — для него полезнее оценивать математическое ожидание, сложив наблюдения со всех траекторий в одну “кучу” (см. оценку $\hat{\mu}_2$ из пункта 15 раздела 5.1). Постройте гистограммы выборочных траекторий и, исходя из их схожести, определите, для какого из процессов какую оценку использовать. Оцените математические ожидания исходя из выбранной оценки.
- 2) Про η известно, что его траектории имеют выраженную периодичность. Ожидается, что период из множества $\{5, 6, \dots, 10\}$. Построив график выборочной автоковариации, угадайте по нему период траекторий процесса. Попробуйте угадать период и по графику траектории процесса.
- 3) Про процессы α и β известно, что их траектории схожи, если траекторию одного из них сместить на l единиц времени относительно другого. Попробуйте определить это смещение / лаг l , построив график кросс-ковариации. Ожидается, что $l \in \{-6, -5, \dots, 6\}$. Попробуйте угадать l и по графику траекторий процессов α и β .

□

17 Задание 6

В этом задании рассматриваются вопросы оценивания параметров модели АРСС / АРПСС и предсказания её будущих значений.

⟨1⟩ Дано.

- 1) Рассматривается случайный процесс $\xi \sim ARIMA(p, d, q)$. В файле `1.csv` дана выборочная траектория из ξ . Поля таблицы означают:
 - 1) t — момент наблюдения.
 - 2) x — значение процесса.
- 2) Рассматривается случайный процесс $\eta \sim ARMA(p, q)$. В файле `2.csv` дана выборочная траектория из η . Поля таблицы означают:
 - 1) t — момент наблюдения.
 - 2) x — значение процесса.
- 3) В файле `2_parameters.txt` даны истинные значения параметров процесса η .
- 4) В файле `2_prediction_d.txt` дано значение d^* (смысл величины см. ниже в задании).

□

⟨2⟩ Задание.

1) По выборочной траектории оцените параметры $p, d, q, \varphi, \theta, a_0, \sigma_\varepsilon^2$ процесса $\xi \sim ARIMA(p, d, q)$. Известно, что $p, d, q \leq 2$. Порядок оценивания:

- 1) Оцените d на основе визуальной оценки стационарности процесса по графику выборочной автокорреляции. Стройте график $|\hat{r}(h)|$ для $h \in \{0, \dots, 10\}$.
- 2) Оцените p, q либо на основе BIC критерия, либо на основе проверки некоррелированности остатков критерием Льюнга-Бокса (с $K = 10$). Для подгонки параметров $\varphi, \theta, a_0, \sigma_\varepsilon^2$ при фиксированных p, q используйте метод максимума условного правдоподобия.

При подгонке параметров методом максимума условного правдоподобия можете применять любой метод численной минимизации (в том числе, уже реализованный в библиотеках). Например, можно воспользоваться функцией численной минимизации `optimize.minimize` из пакета `scipy` с использованием в качестве метода минимизации (параметр `method`) метода “Nelder-Mead” с максимальным количеством итераций 10000.

3) В качестве ответа выведите значения всех полученных оценок.

2) Для процесса $\eta \sim ARMA(p, q)$ даны истинные значения параметров $p, q, \varphi, \theta, a_0, \sigma_\varepsilon^2$ и выборочная траектория $\eta(1), \dots, \eta(n)$. Сделайте следующее:

- 1) Сделайте прогноз значений процесса η на 2 шага вперёд, т.е. оцените условное математическое ожидание и условную дисперсию $\eta(n+1)$ и $\eta(n+2)$.
- 2) Предполагая, что $\eta(t) = \nabla^{d^*} \zeta(t)$ с $\zeta \sim ARIMA(p, d^*, q)$, $\zeta(n-1) = 1$ и $\zeta(n) = 2$, сделайте прогноз о значениях $\zeta(n+1)$ и $\zeta(n+2)$, т.е. оцените их условные математические ожидания и условные дисперсии.

□