ФКН ВШЭ, 3 курс, 3 модуль

Задание 2. Марковские цепи. Авторегрессионные и условно-гауссовские модели временных рядов

Вероятностные модели и статистика случайных процессов, весна 2017

Время выдачи задания: 6 февраля (понедельник).

Срок сдачи: **20 февраля (понедельник), 23:59.**

Среда для выполнения практического задания – PYTHON 2.x.

Правила сдачи

Инструкция по отправке:

1. Домашнее задание необходимо отправить до дедлайна на почту hse.cs.stochastics@gmail.com.

2. В письме укажите тему «[ФКН ССП17] Задание 2, Фамилия Имя».

3. Решения задач следует присылать единым файлом формата .pdf, набранным в LaTeX. Допускается отправка последней задачи в виде отдельной ipython-тетрадки.

Оценивание и штрафы:

1. **Каждая из задач имеет стоимость 2 балла**, при этом за задачу можно получить 0, 1 или 2 балла. Максимально допустимая оценка за работу – 10 баллов. Баллы, набранные сверх максимальной оценки, считаются бонусными и влияют на освобождение от задач на экзамене.

2. Дедлайн жесткий. Сдавать задание после указанного срока сдачи нельзя.

3. Задание выполняется самостоятельно. «Похожие» решения считаются плагиатом и все задействованные студенты (в том числе те, у кого списали) не могут получить за него больше 0 баллов (подробнее о плагиате см. на странице курса). Если вы нашли решение какого-то из заданий (или его часть) в открытом источнике, необходимо указать ссылку на этот источник в отдельном блоке в конце Вашей работы (скорее всего вы будете не единственным, кто это нашел, поэтому чтобы исключить подозрение в плагиате, необходима ссылка на источник).

Необходимые теоретические сведения

1. Ковариационной функцией $R_X(t_1, t_2)$ случайного процесса $X = (X_t)_{t \geqslant 0}$ называется неслучайная функция

$$R_X(t_1, t_2) = \operatorname{cov}(X_{t_1} - \operatorname{E} X_{t_1})(X_{t_2} - \operatorname{E} X_{t_2}).$$

Корреляционной функцией $r_X(t_1, t_2)$ случайного процесса X называется неслучайная функция

$$r_X(t_1, t_2) = \frac{R_X(t_1, t_2)}{\sqrt{\operatorname{V} X_{t_1} \operatorname{V} X_{t_2}}},$$

где $\operatorname{V} X_t = \operatorname{E}(X_t - \operatorname{E} X_t)^2$ – функция дисперсии случайного процесса X. Взаимной ковариационной функцией $R_{XY}(t_1, t_2)$ пары случайных процессов $X = (X_t)_{t \geqslant 0}$ и $Y = (Y_t)_{t \geqslant 0}$ называется неслучайная функция

$$R_{XY}(t_1, t_2) = \operatorname{cov}(X_{t_1} - \operatorname{E} X_{t_1})(Y_{t_2} - \operatorname{E} Y_{t_2}).$$

Взаимная корреляционная функция $r_{XY}(t_1, t_2)$ определяется аналогично равенству для $r_X(t_1, t_2)$ выше.

2. Случайный процесс $X = (X_t)_{t \geqslant 0}$ называется процессом с независимыми приращениями, если для любых $t_0, t_1, \ldots, t_n$, таких, что $0 = t_0 < t_1 < \ldots < t_n$ случайные величины $X_1 - X_0, \ldots, X_n - X_{n-1}$ независимы в совокупности.

3. Вектор $X = (X_1, \ldots, X_n)$ называется гауссовским, если для любого набора коэффициентов $(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$ случайная величина $Y \overset{\text{def}}{=} \sum_{k=1}^{n} \lambda_k X_k$ имеет нормальное распределение.

4. Процесс $W = (W_t)_{t \geqslant 0}$ на $(\Omega, \mathcal{F}, \mathrm{P})$ называется винеровским (или броуновским движением), если

- $W_0 = 0$ P-п.н.,
- W_t имеет независимые приращения $\forall t$,

- $W_t - W_s \sim \mathcal{N}(0, t-s)\ \ \forall t > s \geq 0.$

процесс ARMA (p,q) уравнения Юла-Уолкера частичная автокорреляционная функция

Модель ARMAX(p, q, r) (autoregressive moving average with exogenous inputs) – это стохастическая модель временного ряда, в которой наблюдения $X = (X_t)_{t \geqslant 0}$ задаются соотношением

$$X_t = a_0 + \sum_{i=1}^{p} a_i X_{t-i} + \sum_{i=1}^{p} b_i \varepsilon_{t-i} + \sigma \varepsilon_t + \sum_{i=1}^{r} c_i u_{t-i},$$

где имеются члены AR(p), MA(q), и $u = (u_t)_{t \geqslant 0}$ – это некоторая заданная (возможно, случайная) последовательность.

Вариант 1

1. Пусть $\xi_1, \xi_2, \ldots$ – независимые одинаково распределенные случайные величины, причем

$$\mathrm{P}(\xi_t = 1) = p, \quad \mathrm{P}(\xi_t = -1) = 1 - p.$$

Проверить, являются ли цепями Маркова следующие последовательности случайных величин:

   (a) $\eta_t = \xi_t \xi_{t+1}$,

   (b) $\eta_t = \xi_1 \xi_2 \ldots \xi_t$.

Для цепей Маркова найти вероятности перехода за один шаг.

2. Движение частицы по целым точкам отрезка $[0, N]$ описывается цепью Маркова с $N+1$ состояниями и вероятностями перехода за один шаг $p_{00} = p_{NN} = 1$; $p_{i,i+1} = p, p_{i,i-1} = 1-p = q$; $i = 1, \ldots, N-1$ («случайное блуждание с поглощающими стенками»). Пусть ξ_t – положение частицы в момент t, и $\tau = \min\{t : \xi_t = N\}$ – первый момент времени частицей верхней стенки. Показать, что

$$m_k = \mathrm{E}[\tau \mid \xi_0 = k] = \begin{cases} \frac{2pq}{(p-q)^2}\left(\left(\frac{q}{p}\right)^N - \left(\frac{q}{p}\right)^k\right) - \frac{N-k}{q-p}, & \text{если } p \neq q, \\ (N-k)(N+k), & \text{если } p = q. \end{cases}$$

3. Скачайте в интернете музыкальный midi файл, обучите марковскую цепь и сгенерируйте импровизацию

4. Для случайного процесса $h = (h_t)_{t=1,2,\ldots}$, заданного выражением:

   (a) $h_t = \varepsilon_t - 1.3\varepsilon_{t-1} + 0.4\varepsilon_{t-2}$,

   (b) $h_t - h_{t-1} = \varepsilon_t - 1.3\varepsilon_{t-1} + 0.3\varepsilon_{t-2}$, где $\varepsilon = (\varepsilon_t)_{t\geqslant 0}$ – последовательность независимых одинаково стандартно нормально распределенных случайных величин, (1) записать эквивалентные представления с использованием оператора сдвига L, (2) провести исследование на стационарность, (3) вычислить первые четыре автокорреляции $\rho_0, \rho_1, \rho_2, \rho_3$, (4) вычислить дисперсию случайной величины h_t, (5) определить тип процесса в терминах ARMA(p, q), (6) записать уравнения Юла-Уолкера, (7) решить эти уравнения с определением коэффициентов автокорреляции ρ_1, ρ_2, (8) получить частичную автокорреляционную функцию.

5. Рассматривается процесс ARMAX, заданный уравнением

$$X_t - 1.5X_{t-1} + 0.7X_{t-2} = u_{t-1} + 0.5u_{t-2} + \varepsilon_t - \varepsilon_{t-1} + 0.2\varepsilon_{t-2},$$

   где $\varepsilon = (\varepsilon_t)_{t\geqslant 0}$ – процесс белого шума с дисперсией 0.25, последовательность $u = (u_t)_{t\geqslant 0}$ – случайная бинарная (± 1) последовательность.

   (a) Сгенерировать траекторию длины $N = 500$ указанного процесса;

   (b) Выписать функционал правдоподобия указанного процесса и выражения для оценок максимального правдоподобия его параметров в предположении, что известны порядки частей AR, MA и X процесса;

   (c) По сгенерированной выборке оценить параметры процесса ARMAX в предположении, что полностью известна модель (известны порядки частей AR, MA и X процесса).

   (d) Построить графики зависимости оценок параметров процесса ARMAX (предполагается, что полностью известна модель процесса) от объема использованной выборки. Сходятся ли эти оценки к настоящим значениям?

(e) Подогнать к сгенерированной траектории процесс авторегрессии порядка $p = 4$. Использовать batch-оценку параметров процесса авторегрессии. Сравнить передаточную функцию (нарисовать передаточные функции на одном графике) полученного фильтра авторегрессии с передаточной функцией фильтра исходного процесса ARMAX.

6. Доходность акций часто оценивают в процентах согласно следующей формуле. Если S_n – цена закрытия акции в день $n, n \geqslant 1$, то ее доходность равна величине $X_n = \frac{S_n - S_{n-1}}{S_n}$ (отношение прироста цены закрытия за текущий день к цене предыдущего дня, daily returns in percent).

   Для моделирования временного ряда, соответствующего процентной доходности некоторого фондового индекса, использовалось 1000 наблюдений, к которым была подогнана GARCH-модель. Результатом подгонки являются уравнения:

$$X_n = 0.102 + h_n, \quad h_n = \sigma_n \varepsilon_n,$$
$$\sigma_n^2 = 0.146 + 0.1068 h_{n-1}^2 + 0.8212 \sigma_{n-1}^2.$$

   Здесь X_n – доходность в день n, $\varepsilon = (\varepsilon_t)_{t=1,2,\dots}$ – процесс гауссовского белого шума.

   Доходность в день $n = 1000$ (2013-05-31) составляла $X_{1000} = -1.354$, причем $\sigma_{1000}^2 = 2.27$. В день $n + 1 = 1001$ (2013-06-03) наблюдаемая доходность составила $X_{1001} = -10.47$, в то время как прогноз дисперсии в этот день составил 2.24.

   (a) Записать формулу для подсчета σ_{1001}^2, подставить числовое значение, подсчитать результат;

   (b) Подсчитать распределение доходности X_{1001};

(c) Обосновать ожидаемость или невозможность наблюдений фактической доходности $X_{1000} = -1.354$;

(d) Используя приведенный пример, обосновать утверждение о том, что GARCH является моделью с условной дисперсией;

(e) Используя уравнения модели, охарактеризовать *безусловную* дисперсию доходности.

7. Вам выдан файл `aapl.txt`, содержащий значения наблюдаемой доходности $X_n = \frac{S_n - S_{n-1}}{S_n}$ акций компании Apple в период с по .

(a) Записать формулу для правдоподобия условно-гауссовской модели ARCH(p);

(b) Используя несколько различных значений p, оценить параметры модели ARCH(p), прокомментировать качество оценки для различных p;

(c) Определить, характеризуются ли данные кластерами волатильности (volatility clustering), т.е. дает ли модель ARCH(p) преимущества по сравнению с однородной моделью, для которой $\sigma_n^2 = \text{const}$;

(d) Для выбранного вами значения p нарисовать график волатильности σ_n^2;

(e) Описать процедуру выбора оптимального значения p с помощью какого-либо информационного критерия.

Вариант 2

1. Пусть $\xi_1, \xi_2, \ldots$ – независимые одинаково распределенные случайные величины, причем

$$\mathrm{P}(\xi_t = 1) = p, \quad \mathrm{P}(\xi_t = -1) = 1 - p = q.$$

Положим $\eta_0 = 0, \eta_{t+1} = \eta_t + \xi_{t+1}$. Выяснить, является ли последовательность η_t цепью Маркова. Найти $\mathrm{P}(\eta_t = m), m = 0, 1, \ldots$

2. Движение частицы по целым точкам отрезка $[0, N]$ описывается цепью Маркова с $N + 1$ состояниями и вероятностями перехода за один шаг $p_{00} = p_{NN} = 1; p_{i,i+1} = p, p_{i,i-1} = 1-p = q; i = 1, \ldots, N-1$ («случайное блуждание с поглощающими стенками»). Пусть ξ_t – положение частицы в момент t, и $p_{ij}(t) = \mathrm{P}(\xi_t = j|\xi_0 = i)$. Найти вероятности поглощения частицы в точках 0 и N:

$$\pi_k^{(0)} = \lim_{t\to\infty} p_{k0}(t), \qquad \pi_k^{(N)} = \lim_{t\to\infty} p_{kN}(t).$$

3. Скачайте в интернете музыкальный midi файл, обучите марковскую цепь и сгенерируйте импровизацию

4. Для случайного процесса $h = (h_t)_{t=1,2,\ldots}$, заданного выражением:

   (a) $h_t - 0.5h_{t-1} = \varepsilon_t - 1.3\varepsilon_{t-1} + 0.4\varepsilon_{t-2}$,

   (b) $h_t - 1.5h_{t-1} + 0.6h_{t-2} = \varepsilon_t$,

   где $\varepsilon = (\varepsilon_t)_{t\geqslant 0}$ – последовательность независимых одинаково стандартно нормально распределенных случайных величин, (1) записать эквивалентные представления с использованием оператора сдвига L, (2) провести исследование на стационарность, (3) вычислить первые четыре автокорреляции $\rho_0, \rho_1, \rho_2, \rho_3$, (4) вычислить дисперсию случайной величины h_t, (5) определить тип процесса в терминах ARMA(p, q), (6) записать уравнения Юла-Уолкера, (7) решить эти уравнения с определением коэффициентов автокорреляции ρ_1, ρ_2, (8) получить частичную автокорреляционную функцию.

5. использование ar(p) для моделирования сигнала речи (или еще чего-нить)

6. Доходность акций часто оценивают в процентах согласно следующей формуле. Если S_n – цена закрытия акции в день $n, n \geqslant 1$, то ее доходность равна величине $X_n = \frac{S_n - S_{n-1}}{S_n}$ (отношение прироста цены закрытия за текущий день к цене предыдущего дня, daily returns in percent).

   Для моделирования временного ряда, соответствующего процентной доходности некоторого фондового индекса, использовалось 1000 наблюдений, к которым была подогнана GARCH-модель. Результатом подгонки являются уравнения:

$$X_n = 0.102 + h_n, \quad h_n = \sigma_n \varepsilon_n,$$
$$\sigma_n^2 = 0.146 + 0.1068 h_{n-1}^2 + 0.8212 \sigma_{n-1}^2.$$

   Здесь X_n – доходность в день n, $\varepsilon = (\varepsilon_t)_{t=1,\,2,\,\ldots}$ – процесс гауссовского белого шума.

   Доходность в день $n = 1000$ (2013-05-31) составляла $X_{1000} = -1.354$, причем $\sigma_{1000}^2 = 2.27$. В день $n + 1 = 1001$ (2013-06-03) наблюдаемая доходность составила $X_{1001} = -10.47$, в то время как прогноз дисперсии в этот день составил 2.24.

   (a) Записать формулу для подсчета σ_{1001}^2, подставить числовое значение, подсчитать результат;

   (b) Подсчитать распределение доходности X_{1001};

(c) Обосновать ожидаемость или невозможность наблюдений фактической доходности $X_{1000} = -1.354$;

(d) Используя приведенный пример, обосновать утверждение о том, что GARCH является моделью с условной дисперсией;

(e) Используя уравнения модели, охарактеризовать *безусловную* дисперсию доходности.

7. Вам выдан файл `goog.txt`, содержащий значения наблюдаемой доходности $X_n = \frac{S_n - S_{n-1}}{S_n}$ акций компании Google в период с по .

   (a) Записать формулу для правдоподобия условно-гауссовской модели GARCH(p, q);

   (b) Используя несколько различных значений p и q, оценить параметры модели GARCH(p, q), прокомментировать качество оценки для различных p и q;

   (c) Определить, характеризуются ли данные кластерами волатильности (volatility clustering), т.е. дает ли модель GARCH(p, q) преимущества по сравнению с однородной моделью, для которой $\sigma_n^2 = \text{const}$;

   (d) Для выбранных вами значений p, q нарисовать график волатильности σ_n^2;

   (e) Описать процедуру выбора оптимального значения параметров p и q с помощью какого-либо информационного критерия.