

Санкт-Петербургский государственный университет
Прикладная математика и информатика

Отчет по учебной практике 1 (проектно-технологическая) (семестр 1)

Выполнил:

Потешкин Егор Павлович

группа 24.М22-мм

Научный руководитель:

д. ф.-м. н., доцент

Голяндина Нина Эдуардовна

Кафедра Статистического Моделирования

Оглавление

Введение	3
Глава 1. Теория случайных процессов	4
1.1. Вспомогательные определения	4
1.2. Процессы с длинной памятью	5
1.2.1. Возникновение процессов с длинной памятью	7
1.3. Оценка параметров	7
1.3.1. Maximum likelihood estimation (MLE)	8
1.3.2. Whittle estimation	8
1.3.3. Численное сравнение методов оценки параметров	9
1.3.4. Сходимость оценок к истинным значениям	12
Глава 2. Метод Monte Carlo SSA	16
2.1. Проверка статистических гипотез	16
2.1.1. Поправка неточных критериев	16
2.1.2. Сравнение критериев	17
2.2. Monte Carlo SSA	17
2.2.1. Метод SSA	17
2.2.2. Постановка задачи	18
2.2.3. Множественный тест	18
2.2.4. Используемый вариант MC-SSA	19
2.2.5. Сравнение MC-SSA по мощности при разных моделях шума	20
2.3. Применение MC-SSA на реальных временных рядах	21
2.3.1. Nile Minima	21
2.3.2. Ireland Wind	24
Заключение	26
Список литературы	27
Приложение А. Графики	29
А.1. Сравнение <code>arfima_mle</code> и <code>arfima</code>	29

Введение

TODO

Глава 1

Теория случайных процессов

1.1. Вспомогательные определения

Для начала введем некоторые обозначения, которые будем использовать в дальнейшем.

Определение 1.1. Случайный процесс $\{Y_t : t \in \mathbb{Z}\}$ называют стационарным (в широком смысле), если

1. $EY_t \equiv \text{const}$ (среднее постоянно по времени);
2. $\text{cov}(Y_t, Y_{t+h}) = \gamma(h)$ (ковариация зависит только от лага h).

Замечание 1.1. Поскольку $\gamma(0) = \text{cov}(Y_t, Y_t) = DY_t$, то дисперсия также не меняется со временем.

Замечание 1.2. Далее под стационарностью будет подразумеваться именно стационарность в широком смысле.

Определение 1.2. Случайный процесс $\{\varepsilon_t\}$ называют белым шумом $\text{WN}(0, \sigma^2)$, если он стационарный, $E\varepsilon_t = 0$, $\gamma(h) = 0 \ \forall h \neq 0$ и $D\varepsilon_t = \sigma^2$.

Определение 1.3. Моделью $\text{ARMA}(p, q)$, где $p, q \in \mathbb{N} \cup \{0\}$ называют случайный процесс $\{X_t\}$, удовлетворяющий соотношению

$$X_t = \varepsilon_t + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i},$$

где $\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2)$.

Замечание 1.3. Модель $\text{ARMA}(p, q)$ является стационарным и обратимым процессом, если корни характеристических полиномов

$$\Phi(L) = 1 - \sum_{i=1}^p \phi_i L^i, \quad \Theta(L) = 1 + \sum_{i=1}^q \theta_i L^i$$

лежат вне единичной окружности $\{z : |z| = 1\}$ [2, Section 3.4.1].

Определение 1.4. Процесс $\{X_t\}$ называют красным шумом с параметрами ϕ и σ^2 , если $\{X_t\}$ — стационарная модель ARMA(p, q) с $p = 1$, $q = 0$ и $\phi = \phi_1 \in (0, 1)$.

Определение 1.5. Спектральной плотностью стационарного процесса называется такая функция $f(\omega)$, что

$$\gamma(h) = 2 \int_0^{1/2} e^{2\pi h \omega i} f(\omega) d\omega.$$

Определение 1.6. Пусть $\{Y_t\}$ — стационарный процесс. Функцию

$$I(\omega) = \frac{1}{n} \left| \sum_{j=1}^n Y_j e^{-2\pi \omega j i} \right|^2$$

называют периодограммой выборки размера n процесса $\{Y_t\}$.

Замечание 1.4. Для любой фиксированной частоты ω_0

$$\begin{aligned} E(I(\omega_0)) &\rightarrow f(\omega_0), \quad n \rightarrow \infty; \\ D(I(\omega_0)) &\rightarrow f^2(\omega_0) \neq 0, \quad n \rightarrow \infty. \end{aligned}$$

Таким образом периодограмма является асимптотически несмещенной, но несостоятельной, оценкой спектральной плотности [3, Section 4.5].

1.2. Процессы с длинной памятью

Определение 1.7. Говорят, что стационарный процесс $\{Y_t\}$ обладает длинной памятью, если

$$\sum_{h=0}^H |\gamma(h)| \rightarrow \infty,$$

при $H \rightarrow \infty$. Иначе говорят, что $\{Y_t\}$ обладает короткой памятью:

$$\sum_{h=0}^{\infty} |\gamma(h)| < \infty.$$

Существуют и альтернативные определения процессов с длинной памятью, которые можно найти в [1, Section 3.1]. Там же показано, что они согласованы с определением 1.7.

Пример 1.1. Процессом с короткой памятью является, например, стационарная модель ARMA(p, q), поскольку $|\gamma(h)| \leq CR^h$, где $C > 0$ и $0 < R < 1$ [2, Section 10.4].

Введем понятие дробного интегрирования $(1 - L)^d$, где L — оператор сдвига. Например, для $d = 1$ имеем $(1 - L)Y_t = Y_t - Y_{t-1}$, для $d = 2$ — $(1 - L)^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$, и так далее. Обобщим этот оператор для нецелых d с помощью разложения в ряд Тейлора функции $(1 - x)^d$ в нуле:

$$\begin{aligned} (1 - x)^d &= 1 - dx - \frac{d(1-d)}{2}x^2 - \frac{d(1-d)(2-d)}{3!}x^3 - \dots \\ &= \sum_{j=0}^{\infty} \pi_j(d)x^j = \sum_{j=0}^{\infty} \binom{d}{j}(-1)^j x^j, \end{aligned}$$

где $\binom{d}{j}$ — обобщенный биномиальный коэффициент. Коэффициенты $\pi_j(d)$ удовлетворяют соотношению

$$\pi_j(d) = (-1)^j \binom{d}{j} = \frac{j-1-d}{j} \pi_{j-1}(d) = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}, \quad (1.1)$$

где $\Gamma(x)$ — гамма функция. Заметим, что второе равенство в формуле (1.1) верно для любых d , третье же верно только для $d \notin \mathbb{N} \cup \{0\}$, поскольку гамма функция не определена для неположительных целых чисел.

Определение 1.8. Пусть процесс $\{Y_t\}$ определен соотношением

$$Y_t = (1 - L)^{-d} X_t = \sum_{k=0}^{\infty} \pi_k(-d) X_{t-k}, \quad d < 1/2,$$

где $\pi_k(-d)$ из формулы (1.1), $\{X_t\}$ — стационарная и обратимая модель ARMA(p, d). Процесс $\{Y_t\}$ называют дробно интегрированной моделью ARMA или ARFIMA(p, d, q).

Предложение 1.1. Процесс $\{Y_t\}$ из определения 1.8 является стационарным процессом с нулевым средним. Его спектральная плотность определяется выражением

$$\begin{aligned} f_Y(\omega) &= 4^{-d} \sin^{-2d}(\pi\omega) f_X(\omega) \\ &= 4^{-d} \sin^{-2d}(\pi\omega) \sigma^2 \frac{|\Theta(e^{-2\pi\omega i})|^2}{|\Phi(e^{-2\pi\omega i})|^2}, \quad \omega > 0 \\ &\sim \omega^{-2d} \sigma^2 \frac{|\Theta(1)|^2}{|\Phi(1)|^2}, \quad \omega \rightarrow 0, \end{aligned} \quad (1.2)$$

где $\Phi(L)$, $\Theta(L)$ — характеристические полиномы процесса $\{X_t\}$.

Доказательство. См. [3, Proposition 6.1]. □

Замечание 1.5. Из формулы (1.2) видно, что монотонность спектральной плотности процесса $\{Y_t\}$ зависит от поведения спектральной плотности процесса $\{X_t\}$.

Следствие 1.1. В условиях предложения 1.1 при $0 < d < 1/2$

$$\gamma(h) \sim C_{\gamma,d} h^{2d-1}, \quad h \rightarrow \infty,$$

где

$$C_{\gamma,d} = \sigma^2 \frac{|\theta(1)|^2}{|\phi(1)|^2} \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}.$$

Доказательство. См. [3, Corollary 6.1]. □

Замечание 1.6. Из следствия 1.1 сразу следует, что ARFIMA(p, d, q) с $d \in (0, 1/2)$ обладает длинной памятью.

1.2.1. Возникновение процессов с длинной памятью

Нас интересуют процессы с монотонной спектральной плотностью, поскольку они довольно распространены в реальном мире. Такими процессами являются процессы со степенной спектральной плотностью $f(\omega) \sim \omega^{-\alpha}$, имеющие большое применение в различных областях, например, в физике, биологии, астрофизике, геофизике и экономике.

Процессы с длинной памятью, являющиеся частным случаем процессов со степенной спектральной плотностью, довольно распространены. Например, в работе [4] обнаружена длинная память в таких среднегодовых гидрологических временных рядах, как количество осадков, температура и данных о речном стоке. В работе [5] на наличие длинной памяти исследовалась скорость ветра в Ирландии, в работе [6] исследовался эффект длинной памяти у сейсмических данных. Помимо геофизики, длинная память встречается также в финансах [7, 8].

1.3. Оценка параметров

Пусть $Y_t = (1 - L)^{-d} X_t$, $d < 1/2$. Будем считать, что $\{X_t\}$ представляет собой модель ARMA(p, q) с нормально распределенным белым шумом $\{\varepsilon_t\}$. Тогда представим его спектральную плотность в параметрическом виде: $f_X(\omega) = f_X(\omega; \boldsymbol{\psi}, \sigma)$, где

$$\boldsymbol{\psi} = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)^T.$$

Поставим задачу оценить параметры $\boldsymbol{\varphi}^T = (d, \boldsymbol{\psi}^T)$ и σ^2 .

1.3.1. Maximum likelihood estimation (MLE)

Поскольку $\{\varepsilon_t\}$ — гауссовский белый шум, вектор

$$\mathbf{Y} = (Y_1, \dots, Y_n)^T \sim \mathcal{N}_n(\mathbf{0}, \Sigma_n),$$

где $\Sigma_n = (\gamma(|i - j|))_{i,j=1}^n$ — ковариационная матрица \mathbf{Y} . Совместная плотность распределения \mathbf{Y} равна

$$(2\pi)^{-n/2} |\Sigma_n|^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{Y}^T \Sigma_n^{-1} \mathbf{Y} \right\}.$$

Рассмотрим логарифм функции правдоподобия. Отбрасывая аддитивные константы, получаем

$$\ell(\boldsymbol{\varphi}, \sigma^2) = -\frac{1}{2} \ln |\Sigma_n| - \frac{1}{2} \mathbf{Y}^T \Sigma_n^{-1} \mathbf{Y}.$$

Положим $\Gamma_n = \Sigma_n / \sigma^2$ и, максимизируя ℓ по σ^2 , получаем

$$\ell_c(\boldsymbol{\varphi}) = -\frac{n}{2} \ln (S(\boldsymbol{\varphi})/n) - \frac{1}{2} \ln g_n(\boldsymbol{\varphi}), \quad (1.3)$$

где $S(\boldsymbol{\varphi}) = \mathbf{Y}^T \Gamma_n \mathbf{Y}$, $g_n(\boldsymbol{\varphi}) = |\Gamma_n|$. Тогда

$$\hat{\boldsymbol{\varphi}}_{\text{ML}} = \operatorname{argmax}_{\boldsymbol{\varphi}} \ell_c(\boldsymbol{\varphi}), \quad \hat{\sigma}_{\text{ML}}^2 = S(\hat{\boldsymbol{\varphi}}_{\text{ML}}).$$

Замечание 1.7. В случае ненулевого матожидания $\mathbf{E}Y_t = \mu$, для получения оценок максимального правдоподобия нужно вместо \mathbf{Y} рассматривать $\mathbf{Y} - \mu$.

Замечание 1.8. Для вычисления ℓ_c можно использовать алгоритм Левинсона-Дурбина, имеющий временную трудоемкость $O(n^2)$ [9].

1.3.2. Whittle estimation

Метод максимального правдоподобия применим, когда известно матожидание μ . При неизвестном μ обычно используют его оценку $\bar{\mathbf{Y}}$, однако, помимо этого, существует проблема вычислительной сложности метода при больших n .

Обе эти проблемы можно решить, используя оценку Уиттла (Whittle): вместо логарифма функции правдоподобия рассматривается ее оценка (с точностью до константы) [10]. Пусть $f(\omega; \boldsymbol{\varphi}, \sigma^2)$ — спектральная плотность $\{Y_t\}$, $I(\omega)$ — периодограмма \mathbf{Y} , тогда

$$\ell_W(\boldsymbol{\varphi}, \sigma^2) = -\frac{1}{m} \sum_{j=1}^m \left(\ln f(\omega_j; \boldsymbol{\varphi}, \sigma^2) + \frac{I(\omega_j)}{f(\omega_j; \boldsymbol{\varphi}, \sigma^2)} \right),$$

где $m = \lfloor (n-1)/2 \rfloor$, $\omega_j = j/n$, $j = 1, 2, \dots, m$. Заметим, что $f(\omega; \varphi, \sigma^2) = \sigma^2 g(\omega; \varphi)$. Тогда, максимизируя ℓ_W по σ^2 , получаем

$$\hat{\varphi}_W = \operatorname{argmax}_{\varphi} Q(\varphi), \quad \hat{\sigma}_W^2 = \frac{1}{m} \sum_{j=1}^m \frac{I(\omega_j)}{g(\omega_j; \hat{\varphi}_W)},$$

где

$$Q(\varphi) = -\ln \frac{1}{m} \sum_{j=1}^m \frac{I(\omega_j)}{g(\omega_j; \varphi)} - \frac{1}{m} \sum_{j=1}^m \ln g(\omega_j; \varphi).$$

Замечание 1.9. Такой метод оценки параметров можно использовать при неизвестном среднем, поскольку при ее вычислении не используется значение периодограммы в нуле.

Замечание 1.10. Периодограмму временного ряда можно вычислить за $O(n \log n)$ с помощью быстрого преобразования Фурье, что делает этот метод значительно быстрее MLE для больших n .

1.3.3. Численное сравнение методов оценки параметров

Сравним качество оценок параметров следующих моделей:

1. $d = q = 0$, $p = 1$ — модель AR(1);
2. $p = q = 0$ — модель ARFIMA(0, d, 0).
3. $p = 1$, $q = 0$ — модель ARFIMA(1, d, 0).

Для оценки параметров этих моделей на языке R [11] были реализованы функции `arfima_mle` и `arfima_whittle`, которые соответствуют методам MLE и Whittle соответственно. Для реализации MLE были использованы функции `tacvfARFIMA` из пакета `arfima` [12] и `DLLoglikelihood` из пакета `ltsa` [9], вычисляющие автоковариационную функцию модели ARFIMA и функцию $\ell_c(\varphi)$ из формулы (1.3) соответственно.

Помимо функций `arfima_mle` и `arfima_whittle`, будем использовать следующие функции:

1. Функцию `arma` из пакета `stats`, соответствующую MLE модели ARMA;
2. Функцию `fracdiff` из пакета `fracdiff` [13], соответствующую аппроксимации MLE модели ARFIMA, описанной в работе [5] (обозначим ее за H&R).

Замечание 1.11. Помимо вышеперечисленных функций, в пакете `arfima` есть функция `arfima`, которая вычисляет MLE модели ARFIMA. Но данная реализация MLE в некоторых случаях дает оценки хуже, чем `arfima_mle`. Сравнение оценок обеих реализаций можно найти в разделе A.1.

Поскольку для реальных временных рядов матожидание μ неизвестно, будем рассматривать MLE с известным средним и с его оценкой — выборочным средним (будем обозначать их $\text{MLE}(\mu)$ и $\text{MLE}(\bar{x})$ соответственно). Не умаляя общности, пусть $\mu = 0$.

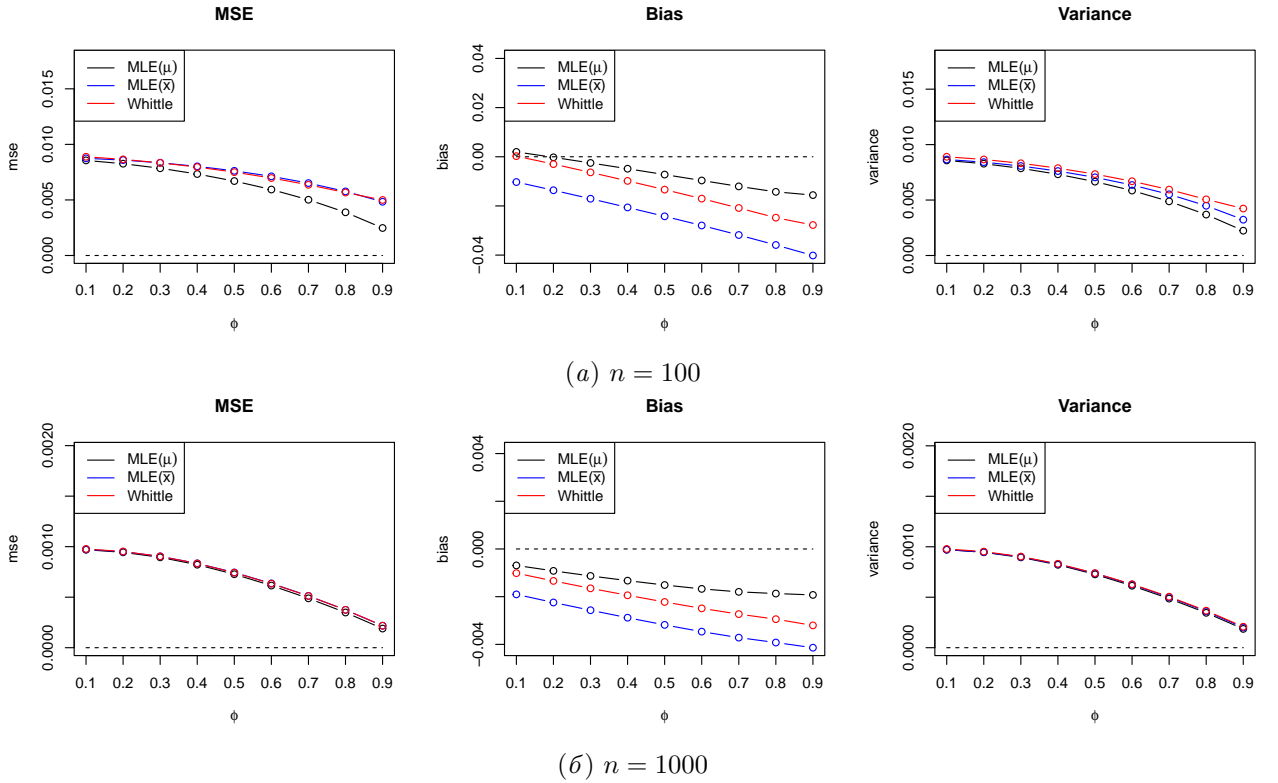


Рис. 1.1. Среднеквадратичное отклонение, смещение и дисперсия оценок параметра ϕ модели AR(1) (500 повторений)

На рис. 1.1 и 1.2 изображены среднеквадратичное отклонение, смещение и дисперсия оценок параметров ϕ и d моделей AR(1) и ARFIMA(0, d , 0). Отметим, что все оценки имеют в большинстве своем отрицательное смещение и отличаются между собой в основном степенью смещения. Как и ожидалось, оценка параметров методом максимального правдоподобия с известным средним дает оценку с наименьшим MSE. С другой стороны, если использовать вместо известного среднего выборочное среднее, оценки становятся сильно смещенными. Whittle же, в свою очередь, дает менее смещенную оценку, чем $\text{MLE}(\bar{x})$, а в случае оценки d имеет смещение даже меньше, чем у $\text{MLE}(\mu)$.

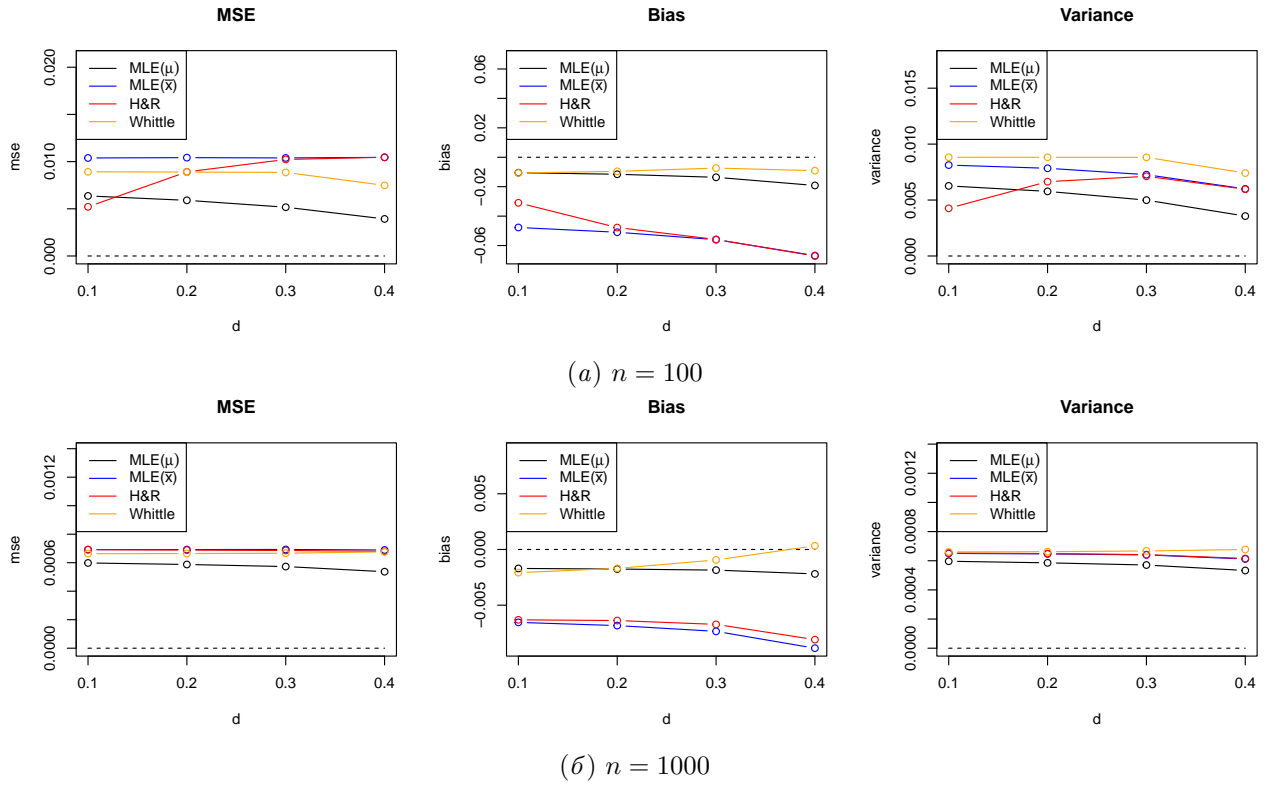


Рис. 1.2. Среднеквадратичное отклонение, смещение и дисперсия оценок параметра d модели ARFIMA(0, d , 0) (500 повторений)

Однако оценки Whittle обладают наибольшей дисперсией среди всех рассмотренных методов, но разница не такая значительная, как в случае смещения.

В таблицах 1.1 и 1.2 представлены значения среднеквадратичной ошибки и смещения оценок параметров d и ϕ модели ARFIMA(1, d). Заметим, что в оценках присутствует смещение, для $\phi = 0.1$ и $\phi = 0.5$ оценки d имеют отрицательное смещение, а оценки ϕ , наоборот, положительное. Также в таблицах синим цветом выделена лучшая по строке оценка d , а красным — лучшая оценка ϕ . Видно, что в случае коротких рядов ($n = 100$) метод H&R в большинстве случаев дает оценки с наименьшим MSE, однако наименьшее смещение имеют оценки MLE(μ). В случае же длинных рядов ($n = 1000$) наименьшую среднеквадратичную ошибку и смещение дает MLE(μ). Отметим, что даже для длинных рядов оценки MLE(\bar{x}) имеют, в большинстве случаев, наибольшее смещение и MSE. Оценки H&R, хотя и дают наименьшее после MLE(μ) MSE, также сильно смещены. Оценки методом Whittle выглядят наиболее привлекательными, поскольку имеют наименьшее после MLE(μ) смещение и имеют MSE меньше, чем MLE(\bar{x}).

Подведем итоги численного сравнения. Если для рассматриваемого ряда извест-

но его матожидание μ (что, конечно, редкость на практике), наилучшим методом является $\text{MLE}(\mu)$. Если же среднее неизвестно, для коротких рядов оценивать параметры моделей $\text{AR}(1)$ и $\text{ARFIMA}(0, d, 0)$ следует методом Whittle, а параметры модели $\text{ARFIMA}(1, d, 0)$ — методом H&R. В случае длинных рядов параметры рассматриваемых моделей следует оценивать методом Whittle.

1.3.4. Сходимость оценок к истинным значениям

Известно [3, Theorem 8.1], что

$$\sqrt{n}(\hat{\varphi}_{\text{ML}} - \varphi_0) \xrightarrow{d} \mathcal{N}_{k+1}(0, \mathcal{I}^{-1}(\varphi_0)), \quad (1.4)$$

где φ_0 — истинный вектор параметров, $\mathcal{I}(\varphi)$ — информационная матрица Фишера. Также известно [3, Proposition 8.3], что вектор $\hat{\varphi}_W$ имеет такое же асимптотическое распределение, что и $\hat{\varphi}_{\text{ML}}$.

Покажем, что методы $\text{MLE}(\mu)$ и Whittle реализованы корректно, посмотрев на дисперсию оценок для $n = 10000$. В таблице 1.3 представлены оценки дисперсий \hat{d} и $\hat{\phi}$, в скобках указана теоретическая дисперсия. Как видим, для разных значений параметров значение дисперсий оценок близки к теоретическим.

Таблица 1.1. Смещение и среднеквадратичное отклонение оценок параметров d и ϕ модели ARFIMA(1, d , 0) ($n = 100, 500$ повторений)

d	ϕ	MSE								Bias							
		MLE(μ)				MLE(\bar{x})				H&R				Whittle			
		\hat{d}	$\hat{\phi}$	\hat{d}	$\hat{\phi}$	\hat{d}	$\hat{\phi}$	\hat{d}	$\hat{\phi}$	\hat{d}	$\hat{\phi}$	\hat{d}	$\hat{\phi}$	\hat{d}	$\hat{\phi}$	\hat{d}	$\hat{\phi}$
0.1	0.1	0.049	0.056	0.119	0.114	0.009	0.018	0.069	0.067	-0.077	0.066	-0.229	0.199	-0.054	0.035	-0.086	0.068
0.2	0.1	0.047	0.055	0.151	0.141	0.025	0.032	0.077	0.073	-0.078	0.067	-0.265	0.232	-0.119	0.099	-0.094	0.074
0.3	0.1	0.041	0.049	0.183	0.165	0.049	0.055	0.084	0.081	-0.076	0.066	-0.301	0.266	-0.179	0.161	-0.109	0.09
0.4	0.1	0.029	0.038	0.211	0.187	0.081	0.089	0.179	0.194	-0.072	0.065	-0.34	0.305	-0.243	0.23	-0.26	0.241
0.1	0.5	0.045	0.041	0.086	0.053	0.010	0.015	0.057	0.054	-0.071	0.034	-0.222	0.151	-0.07	0.034	-0.066	0.024
0.2	0.5	0.042	0.038	0.092	0.055	0.031	0.025	0.074	0.058	-0.081	0.046	-0.244	0.171	-0.154	0.108	-0.153	0.107
0.3	0.5	0.040	0.036	0.1	0.059	0.063	0.043	0.098	0.062	-0.093	0.060	-0.267	0.192	-0.232	0.174	-0.209	0.161
0.4	0.5	0.037	0.033	0.115	0.067	0.104	0.065	0.104	0.066	-0.103	0.073	-0.304	0.226	-0.306	0.235	-0.228	0.177
0.1	0.9	0.029	0.029	0.014	0.01	0.007	0.007	0.034	0.025	0.075	-0.089	0.01	-0.049	0.001	-0.043	0.049	-0.069
0.2	0.9	0.019	0.018	0.011	0.006	0.009	0.004	0.026	0.019	0.046	-0.065	-0.011	-0.035	-0.037	-0.026	0.02	-0.056
0.3	0.9	0.012	0.01	0.009	0.004	0.014	0.003	0.022	0.015	0.016	-0.043	-0.033	-0.023	-0.076	-0.011	-0.024	-0.039
0.4	0.9	0.008	0.006	0.009	0.002	0.025	0.002	0.028	0.01	-0.016	-0.024	-0.061	-0.008	-0.121	0.003	-0.095	-0.016

Таблица 1.2. Смещение и среднеквадратичное отклонение оценок параметров \hat{d} и $\hat{\phi}$ модели ARFIMA(1, d , 0) ($n = 1000$, 500 повторов)

d	ϕ	MSE · 100						Bias · 100									
		MLE(μ)			MLE(\bar{x})			H&R			Whittle						
		\hat{d}	$\hat{\phi}$	$\hat{\phi}$	\hat{d}	$\hat{\phi}$	$\hat{\phi}$	\hat{d}	$\hat{\phi}$	$\hat{\phi}$	\hat{d}	$\hat{\phi}$	$\hat{\phi}$				
0.1	0.1	0.186	0.290	0.27	0.37	0.224	0.317	0.232	0.33	-0.581	0.448	-1.941	1.732	-1.729	1.508	-0.687	0.519
0.2	0.1	0.181	0.287	0.271	0.371	0.267	0.366	0.232	0.329	-0.599	0.465	-2.026	1.824	-1.981	1.773	-0.615	0.468
0.3	0.1	0.174	0.282	0.272	0.372	0.268	0.367	0.232	0.328	-0.639	0.505	-2.177	1.987	-2.176	1.975	-0.476	0.372
0.4	0.1	0.156	0.267	0.272	0.373	0.274	0.371	0.232	0.325	-0.795	0.666	-2.606	2.443	-2.725	2.542	-0.263	0.233
0.1	0.1	0.761	0.8	1.429	1.3	0.504	0.552	1.104	1.05	-2.047	1.588	-5.86	5.102	-3.256	2.741	-2.555	1.904
0.2	0.1	0.710	0.759	1.432	1.302	0.978	0.969	1.213	1.125	-2.018	1.571	-6.08	5.337	-5.157	4.556	-2.721	2.065
0.3	0.1	0.617	0.675	1.462	1.323	1.246	1.175	1.23	1.15	-1.984	1.560	-6.506	5.78	-6.129	5.463	-2.578	1.948
0.4	0.1	0.473	0.539	1.499	1.353	1.507	1.354	1.294	1.18	-2.226	1.861	-7.514	6.838	-7.622	6.905	-2.695	2.099
0.1	0.9	0.338	0.155	0.288	0.122	0.259	0.097	0.387	0.193	0.623	-0.774	-0.095	-0.56	-0.176	-0.504	0.583	-0.92
0.2	0.9	0.273	0.106	0.233	0.077	0.239	0.077	0.326	0.128	0.42	-0.611	-0.388	-0.355	-0.627	-0.303	0.041	-0.652
0.3	0.9	0.241	0.093	0.217	0.068	0.268	0.075	0.326	0.097	0.287	-0.53	-0.667	-0.204	-1.395	-0.029	-1.003	-0.285
0.4	0.9	0.173	0.067	0.182	0.051	0.381	0.076	0.545	0.09	-0.129	-0.295	-1.4	0.178	-2.602	0.359	-3.357	0.389

Метод	$nD\hat{d}$ (1.83167)	$nD\hat{\phi}$ (2.98284)	Метод	$nD\hat{d}$ (1.83167)	$nD\hat{\phi}$ (2.98284)
MLE(μ)	1.72733	2.92604	MLE(μ)	1.69348	2.90374
Whittle	1.78350	2.94929	Whittle	1.84265	2.96828
(a) $d = 0.2, \varphi = 0.1$			(b) $d = 0.4, \varphi = 0.1$		
Метод	$nD\hat{d}$ (4.91219)	$nD\hat{\phi}$ (6.06018)	Метод	$nD\hat{d}$ (4.91219)	$nD\hat{\phi}$ (6.06018)
MLE(μ)	5.00869	6.305	MLE(μ)	4.75975	6.06711
Whittle	5.08405	6.35985	Whittle	5.28104	6.4936
(c) $d = 0.2, \varphi = 0.5$			(d) $d = 0.4, \varphi = 0.5$		
Метод	$nD\hat{d}$ (2.49203)	$nD\hat{\phi}$ (0.77885)	Метод	$nD\hat{d}$ (2.49203)	$nD\hat{\phi}$ (0.77885)
MLE(μ)	2.42011	0.78209	MLE(μ)	2.26718	0.749318
Whittle	2.44394	0.77549	Whittle	2.57117	0.77311
(e) $d = 0.2, \varphi = 0.9$			(f) $d = 0.4, \varphi = 0.9$		

Таблица 1.3. Дисперсия оценок \hat{d} и $\hat{\phi}$, $n = 10000$, 100 повторений

Глава 2

Метод Monte Carlo SSA

2.1. Проверка статистических гипотез

Рассмотрим некоторый критерий со статистикой T . Введем обозначения.

Определение 2.1. Ошибка первого рода — вероятность отвергнуть нулевую гипотезу, если она верна: $\alpha_I(\alpha) = P_{H_0}(T \in A_{\text{крит}}(\alpha))$.

Определение 2.2. Если $\alpha_I = \alpha$, то говорят, что критерий точный при уровне значимости α , иначе говорят, что критерий неточный. При $\alpha_I < \alpha$ критерий является консервативным, а при $\alpha_I > \alpha$ — радикальным.

Определение 2.3. Мощность критерия против альтернативы H_1 — вероятность отвергнуть нулевую гипотезу, если верна альтернативная: $\beta(\alpha) = P_{H_1}(T \in A_{\text{крит}}(\alpha))$.

2.1.1. Поправка неточных критериев

Зафиксируем некоторый неточный (консервативный или радикальный) критерий и уровень значимости α^* . Пусть дана зависимость ошибки первого рода от уровня значимости $\alpha_I(\alpha) = P_{H_0}(p < \alpha)$. Тогда критерий с формальным уровнем значимости $\tilde{\alpha}^* = \alpha_I^{-1}(\alpha^*)$ является точным: ошибка первого рода $\alpha_I(\tilde{\alpha}^*) = \alpha^*$.

Если зависимость $\alpha_I(\alpha)$ неизвестна, она оценивается с помощью моделирования. Приведем алгоритм поправки в этом случае. Помимо критерия и уровня значимости, зафиксируем количество выборок M для оценки $\alpha_I(\alpha)$ и их объем N .

Алгоритм 1. Поправка уровня значимости по зависимости $\alpha_I(\alpha)$ [14]

1. Моделируется M выборок объема N при верной H_0 .
2. По моделированным данным строится оценка зависимости ошибки первого рода от уровня значимости $\alpha_I(\alpha)$.
3. Рассчитывается формальный уровень значимости: $\tilde{\alpha}^* = \alpha_I^{-1}(\alpha^*)$. Критерий с таким уровнем значимости является асимптотически точным при $M \rightarrow \infty$.

2.1.2. Сравнение критериев

Точные критерии, проверяющие одну и ту же гипотезу, можно использовать и сравнивать по мощности: критерий с наибольшей мощностью предпочтительней. Если критерий является консервативным, использовать и сравнивать его с другими критериями по мощности также можно, учитывая, что его мощность будет занижена. Радикальный же критерий, без поправки, введенной в разделе 2.1.1, нельзя использовать и сравнивать по мощности с другими критериями. Поэтому введем понятие ROC-кривой, соответствующей мощности критерия, к которому была применена поправка.

Определение 2.4. ROC-кривая — это кривая, задаваемая параметрически

$$\begin{cases} x = \alpha_I(\alpha) \\ y = \beta(\alpha) \end{cases}, \quad \alpha \in [0, 1]$$

Замечание 2.1. С помощью ROC-кривых можно сравнивать по мощности неточные (в частности, радикальные) критерии. Отметим, что для точного критерия ROC-кривая совпадает с графиком мощности, так как $\alpha_I(\alpha) = \alpha$.

2.2. Monte Carlo SSA

Метод Monte Carlo SSA (MC-SSA) тесно связан с методом SSA (Singular Spectrum Analysis), состоящим из четырех этапов: *вложения, разложения, группировки и диагонального усреднения*. Поэтому опишем сначала его.

2.2.1. Метод SSA

Пусть $\mathbf{X} = (x_1, \dots, x_N)$ — временной ряд длины N . Зафиксируем длину окна L , $1 < L < N$. Рассмотрим $K = N - L + 1$ векторов вложения $X_i = (x_i, \dots, x_{i+L-1})$ и составим из столбцов X_i так называемую траекторную матрицу:

$$\mathbf{X} = [X_1 : \dots : X_K].$$

Далее траекторная матрица \mathbf{X} разбивается в сумму матриц единичного ранга. В базовом SSA используются собственные векторы матрицы $\mathbf{X}\mathbf{X}^T$, в Toeplitz SSA используются собственные векторы матрицы \mathbf{T} с элементами

$$t_{ij} = \frac{1}{N - |i - j|} \sum_{n=1}^{N-|i-j|} x_n x_{n+|i-j|}, \quad i, j \leq L. \quad (2.1)$$

Обозначим за P_1, \dots, P_L собственные векторы матрицы $\mathbf{X}\mathbf{X}^T$ либо матрицы \mathbf{T} . Тогда получаем следующее разложение:

$$\mathbf{X} = \sum_{i=1}^L \sigma_i P_i Q_i^T = \mathbf{X}_1 + \dots + \mathbf{X}_L,$$

где $S_i = \mathbf{X}^T P_i$, $Q_i = S_i / \|S_i\|$, $\sigma_i = \|S_i\|$.

После этого полученные матрицы группируются и каждая из группированных матриц преобразовывается обратно во временной ряд. Таким образом, результатом SSA является разложение временного ряда.

2.2.2. Постановка задачи

Рассмотрим задачу поиска сигнала (неслучайной составляющей) во временном ряде. Модель выглядит следующим образом:

$$\mathbf{X} = \mathbf{S} + \boldsymbol{\xi},$$

где \mathbf{S} — сигнал, $\boldsymbol{\xi}$ — стационарный процесс с нулевым средним. Тогда нулевая гипотеза $H_0 : \mathbf{S} = 0$ (отсутствие сигнала, ряд состоит из чистого шума) и альтернатива $H_1 : \mathbf{S} \neq 0$ (ряд содержит сигнал, например, периодическую составляющую).

2.2.3. Множественный тест

Зафиксируем длину окна L и модель шума $\boldsymbol{\xi}$. Пусть $\mathbf{R}_1, \dots, \mathbf{R}_G$ — реализации $\boldsymbol{\xi}$, которые в дальнейшем будем называть суррогатными. Обозначим за \mathbf{X} и $\boldsymbol{\Xi}_i$, $i = 1, \dots, G$, траекторные матрицы ряда \mathbf{X} и каждой суррогатной реализации соответственно. Рассмотрим H проекционных векторов W_1, \dots, W_H , соответствующих некоторой частоте ω_k , $\|W_k\| = 1$, $k = 1, \dots, H$.

Алгоритм 2. Multiple MC-SSA [15]

1. Для $k = 1, \dots, H$ вычисляется статистика $\hat{p}_k = \|\mathbf{X}^T W_k\|^2$, выборка $P_k = \{p_{ki}\}_{i=1}^G$ с элементами $p_{ki} = \|\boldsymbol{\Xi}_i^T W_k\|^2$, ее среднее μ_k и стандартное отклонение σ_k .
2. Вычисляется $\eta = (\eta_1, \dots, \eta_G)$, где

$$\eta_i = \max_{1 \leq k \leq H} (p_{ki} - \mu_k) / \sigma_k, \quad i = 1, \dots, G.$$

3. Находится q как выборочный $(1 - \alpha)$ -квантиль η , где α — уровень значимости.

4. Нулевая гипотеза не отвергается, если

$$t = \max_{1 \leq k \leq H} (\hat{p}_k - \mu_k) / \sigma_k < q.$$

5. Если H_0 отвергнута, вклад W_k (и соответствующей частоты) значим, если \hat{p}_k превосходит $\mu_k + q\sigma_k$. Таким образом, $[0, \mu_k + q\sigma_k]$ считаются скорректированными интервалами прогнозирования.

Замечание 2.2. Для модели модели шума ξ важно, чтобы спектральная плотность процесса была строго монотонной. Это связано с тем, что собственные векторы автоковариационной матрицы стационарного процесса ведут себя как синусоиды с равностоящими частотами, и при строгой монотонности спектральной плотности вклад этих векторов будет попарно различным, делая их сильно разделимыми. Если же опустить требование строгой монотонности, компоненты могут смешаться, что делает невозможным определение доминирующей частоты значимого вектора.

Замечание 2.3. Примерами процессов со строго монотонной спектральной плотностью являются красный шум и модель ARFIMA(0, d , 0).

2.2.4. Используемый вариант MC-SSA

В разделе 2.2.3 предполагалось, что векторы W_1, \dots, W_H фиксированные и не зависят от исходного ряда. Такой критерий MC-SSA является точным, то есть ошибка первого рода равна заданному уровню значимости. В этой работе будут рассматриваться векторы W_k , порожденные рядом X , при этом по-прежнему при вычислении p_{ki} используются те же W_k , что и при вычислении \hat{p}_k .

Поскольку в этом варианте векторы W_k не заданы заранее, а порождены исходным рядом, критерий MC-SSA становится, вообще говоря, радикальным. Борьба с этой проблемой позволяет метод эмпирической поправки критерия, описанный в разделе 2.1.1.

В качестве W_1, \dots, W_H будем брать собственные векторы матрицы XX^T или T (см. формулу (2.1)). Такой способ выбора векторов для проекции самый распространенный, поскольку, если есть значимые векторы, можно восстановить сигнал с помощью SSA на их основе. Будем под MC-SSA подразумевать именно этот вариант критерия. Варианты критерия будут определяться конкретным разложением траекторной матрицы. Заметим, что обычно используется сингулярное разложение.

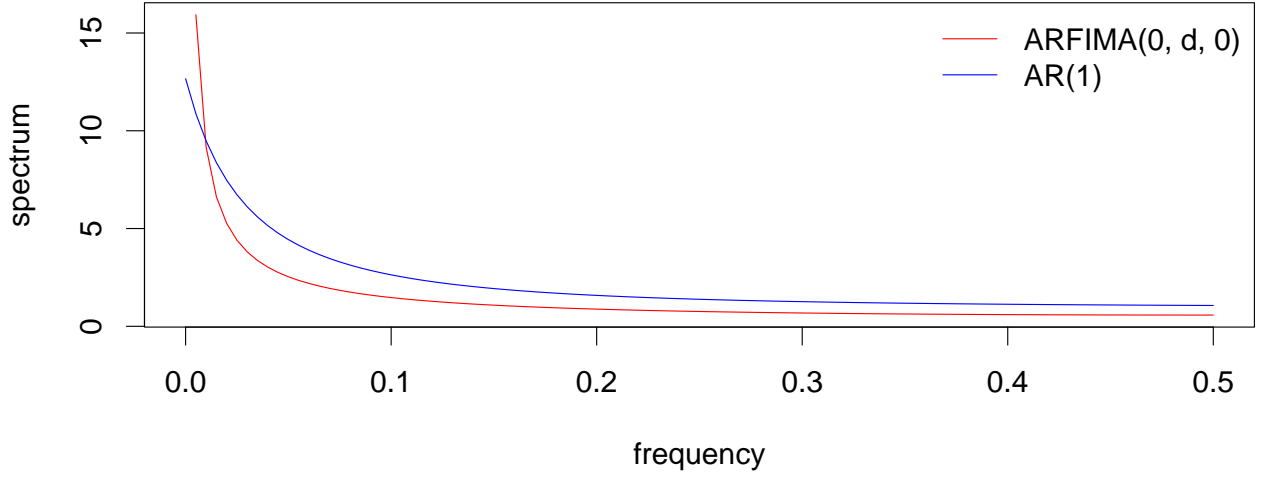


Рис. 2.1. Спектральная плотность процессов с одинаковой дисперсией

2.2.5. Сравнение MC-SSA по мощности при разных моделях шума

Пусть ξ — красный шум, а η — модель ARFIMA(0, d , 0). Будем считать дисперсию белого шума одинаковой для обоих процессов и равной σ^2 . Дисперсии ξ и η соответственно равны

$$D\xi = \frac{\sigma^2}{1 - \phi^2}, \quad D\eta = \sigma^2 \frac{\Gamma(1 - 2d)}{\Gamma(1 - d)^2}.$$

Тогда дисперсии процессов равны тогда и только тогда, когда

$$\phi = \pm \sqrt{1 - \frac{\Gamma(1 - d)^2}{\Gamma(1 - 2d)}}.$$

Пусть $d = 0.4$. Тогда при $\phi \approx 0.719$ процессы ξ и η имеют одинаковую дисперсию. На рис. 2.1 изображены спектральные плотности процессов. На нем видно, что процесс η имеет меньшее значение плотности для всех значений ω , за исключением близких к нулю.

Предполагается, что если рассмотреть в качестве альтернативы сигнал с частотой $\omega : f_\eta(\omega) < f_\xi(\omega)$, то мощность критерия MC-SSA против этой альтернативы при модели шума η больше, чем при модели шума ξ . Убедимся в этом. Пусть длина ряда $N = 100$, $\sigma^2 = 1$ и

$$S = \{A \cos(2\pi n\omega)\}_{n=1}^N, \quad A = 1, \quad \omega = 0.075.$$

На рис. 2.2, *a* изображены график ошибок первого рода критериев MC-SSA для разных длин окна L . По нему видно, что рассматриваемые критерии являются радикальными, поэтому сравнивать их по мощности будем с помощью ROC-кривых, которые

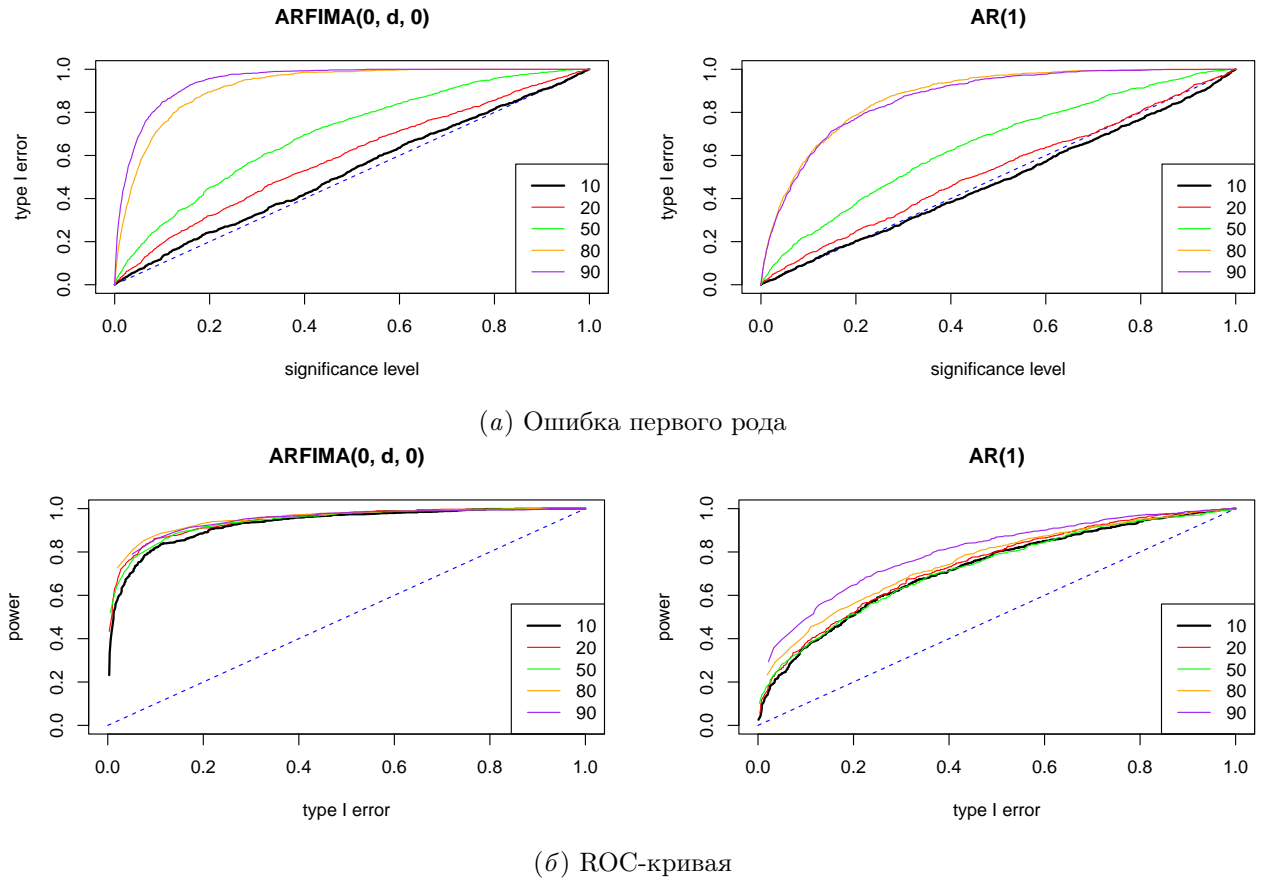


Рис. 2.2. Сравнение мощностей MC-SSA

являются графиком мощности критериев, к которым была применена поправка из раздела 2.1.1. По рис. 2.2, б видно, что, действительно, мощность критерия против данной альтернативы при модели шума $ARFIMA(0, d, 0)$ больше, чем при модели шума $AR(1)$ с такой же дисперсией.

2.3. Применение MC-SSA на реальных временных рядах

Рассмотрим несколько примером реальных временных рядов с длинной памятью и применим к ним MC-SSA. Оценивать параметры будем теми же методами, что и в разделе 1.3.3.

2.3.1. Nile Minima

На рис. 2.3, а изображен ежегодный минимальный уровень воды реки Нил за период с 622 по 1284 год (663 наблюдения), данные были взяты из [16]. Нерегулярные циклы или тенденции в этом временном ряду, обусловленные длинной памятью, впер-

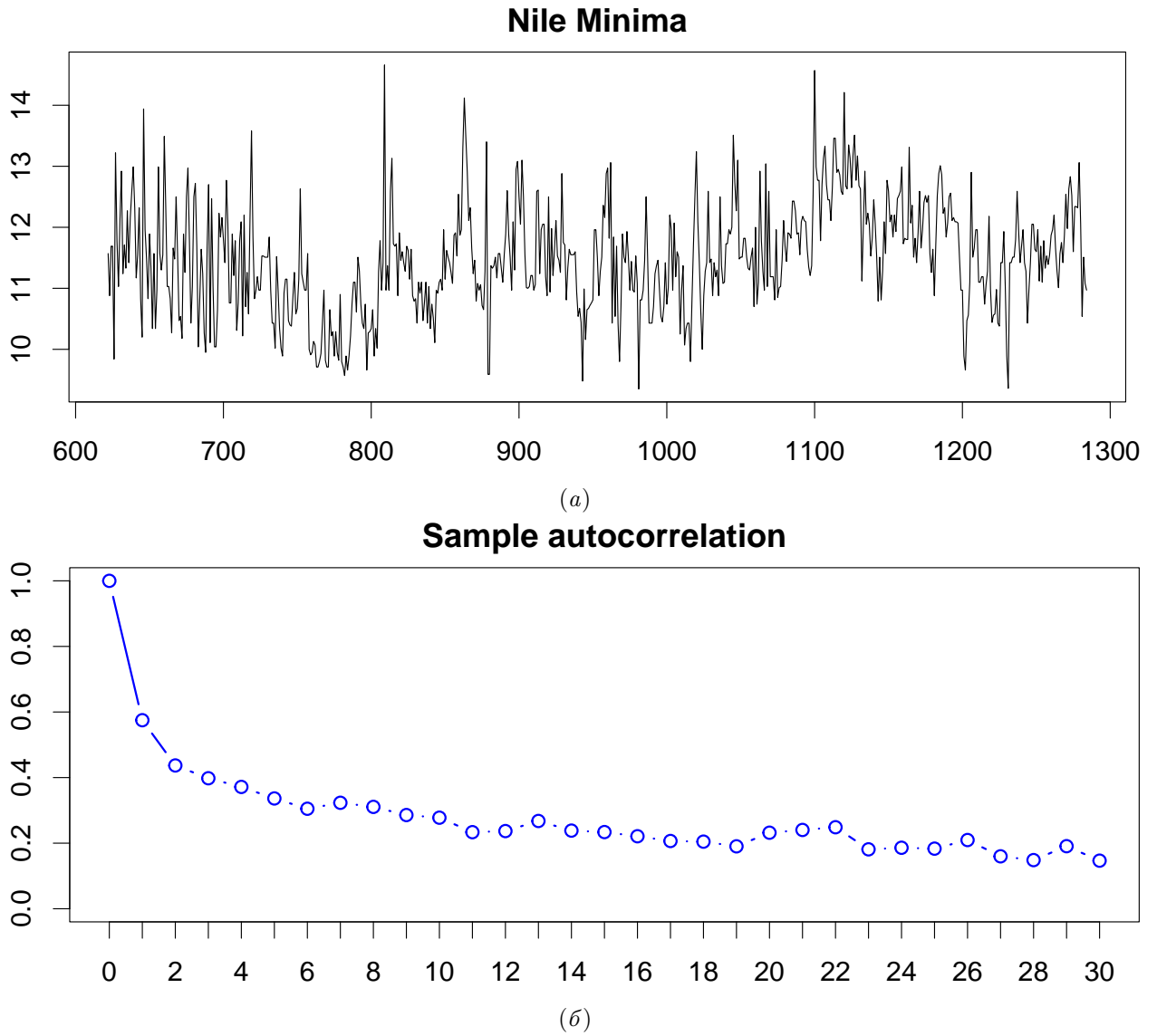


Рис. 2.3. Ежегодный минимальный уровень воды реки Нил

вые были обнаружены и обсуждены Хёрстом, британским инженером, который работал гидрологом на реке Нил. Подтверждает присутствие длинной памяти график медленно угасающей автокорреляционной функции на рис. 2.3, б.

Оценим параметры модели $ARFIMA(0, d, 0)$. В таблице 2.1 представлены оценки параметров d и σ^2 . Поскольку истинное среднее неизвестно и оценка d по Whittle дает наименьшее смещение (см. рис. 1.2), в качестве нулевой гипотезы MC-SSA выберем модель $ARFIMA(0, d, 0)$ с $d = 0.40547$ и $\sigma^2 = 0.48971$, на рис. 2.4 изображена периодограмма ряда вместе с оцененной спектральной плотностью.

Применим MC-SSA с длиной окна $L = 330 \approx N/2$. На рис. 2.5 изображены 95%-ные доверительные интервалы статистик \hat{p}_k , $k = 1, \dots, L$ (см. алгоритм 2). Ни одна из

Таблица 2.1. Оценка параметров модели ARFIMA(0, d , 0) ряда Nile Minima

Метод	\hat{d}	$\hat{\sigma}^2$
MLE(\bar{x})	0.39264	0.48939
H&R	0.39327	0.48934
Whittle	0.40547	0.49026

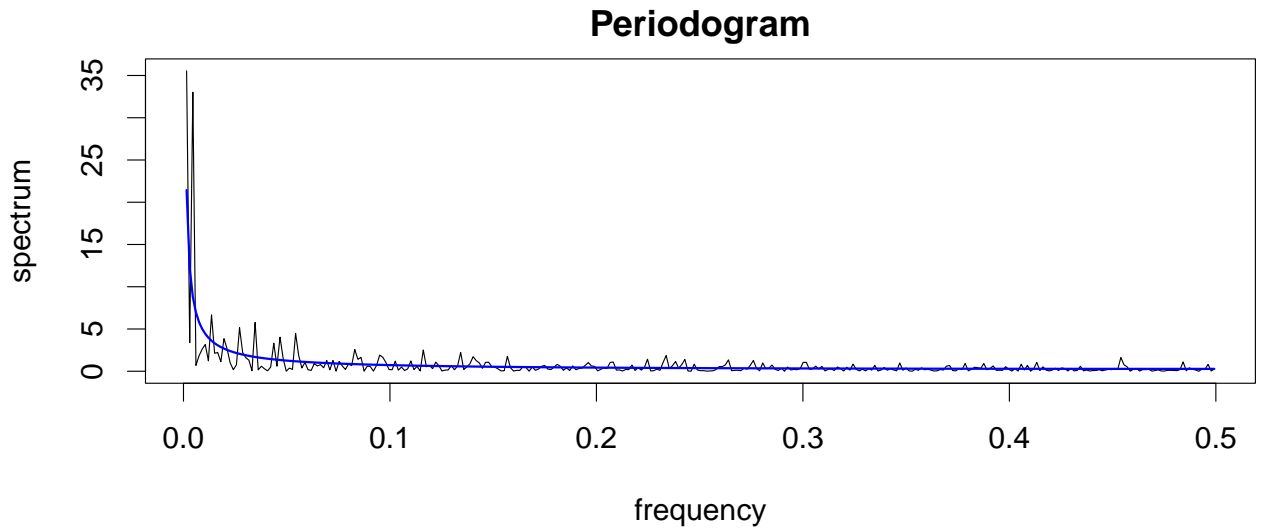


Рис. 2.4. Периодограмма ряда Nile Minima

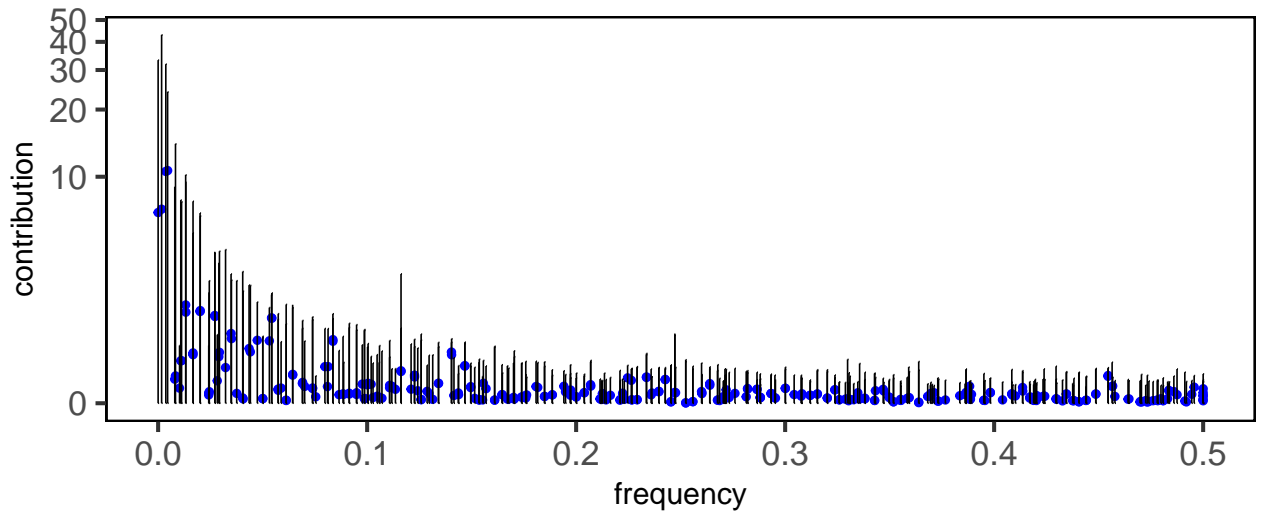


Рис. 2.5. Результат работы MC-SSA для ряда Nile Minima

статистик не является значимой, это означает, что нет оснований полагать, что в этом временном ряде присутствует неслучайный сигнал.

2.3.2. Ireland Wind

На рис. 2.6 изображены среднесуточные данные о скорости ветра (в узлах) за период с 1961 по 1978 год (6574 дней) на станции Roche's Point в Республике Ирландия [5].

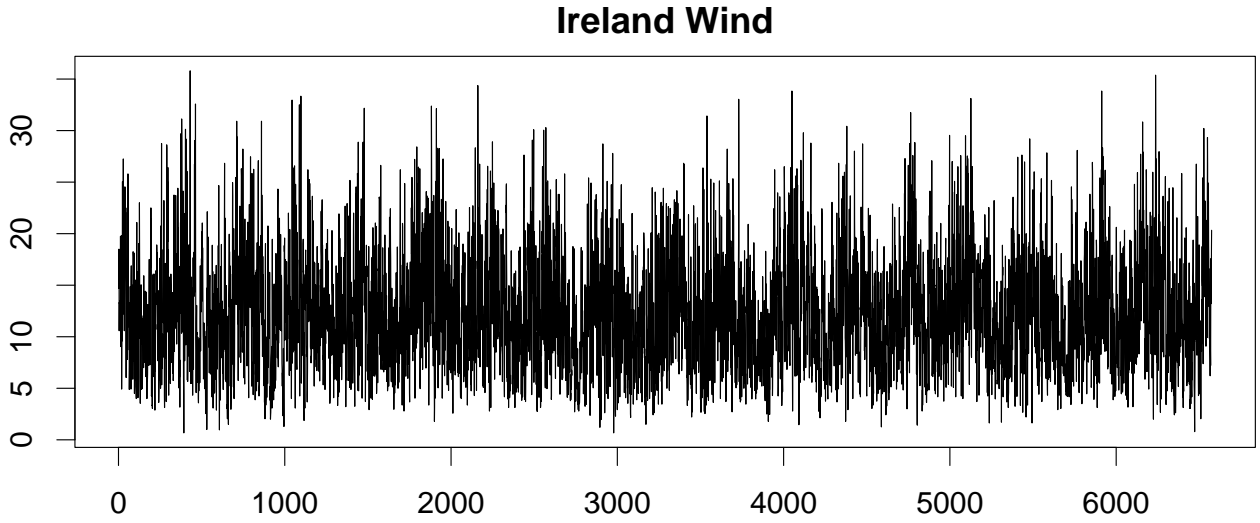


Рис. 2.6. Среднесуточные данные о скорости ветра в Республике Ирландия

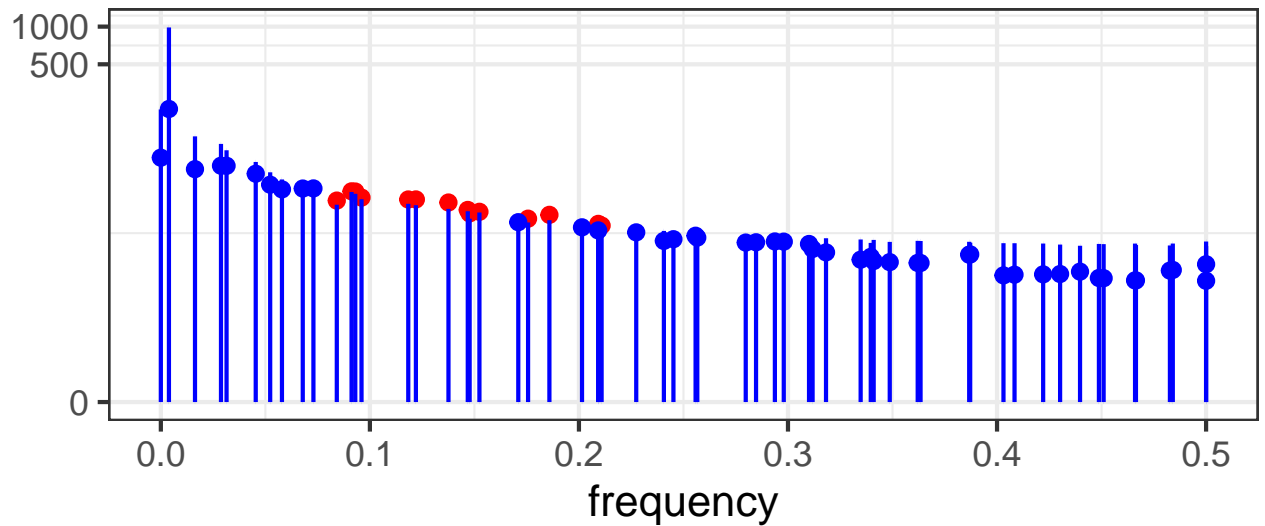
В таблице 2.2 представлены оценки параметров. Полученные оценки примерно одинаковые, но поскольку Whittle дает менее смещенную оценку (см. рис. 1.2 и таблицу 1.2), будем использовать именно ее.

Поскольку ряд достаточно длинный, выберем длину окна не слишком большой, чтобы метод MC-SSA считался за адекватное время, скажем, $L = 60$. На рис. 2.7 представлен результат работы MC-SSA для обеих моделей. Получили следующее: если модель представляет собой чисто процесс с длинной памятью, много векторов оказывается значимыми, в то время как для модели, совмещающей в себе короткую и длинную память, значимых векторов уже нет. Но значимость векторов в первом случае может являться признаком неправильно выбранной модели. Также заметим, что критерий

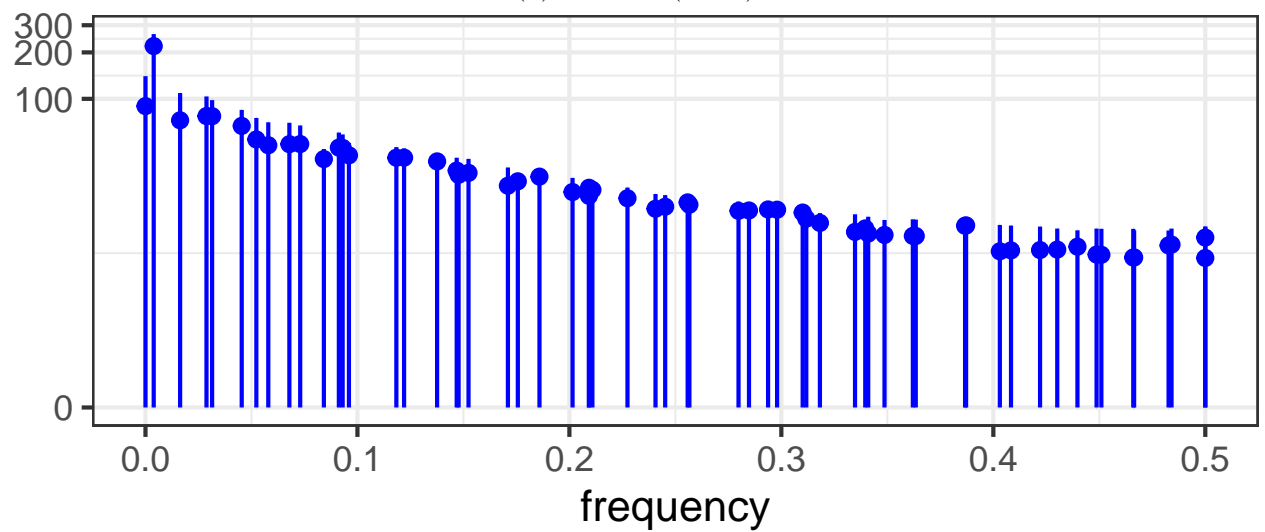
Таблица 2.2. Оценка параметров ряда Ireland Wind

Метод	ARFIMA(0, d , 0)		ARFIMA(1, d , 0)		
	\hat{d}	$\hat{\sigma}^2$	\hat{d}	$\hat{\phi}$	$\hat{\sigma}^2$
MLE(\bar{x})	0.37117	24.39916	0.17306	0.28403	23.7581
H&R	0.36891	24.38116	0.17245	0.28309	23.73458
Whittle	0.37287	24.40285	0.17598	0.28105	23.75983

MC-SSA без поправки является радикальным, и значимые векторы могут на самом деле быть ложно обнаруженными. И поскольку применить поправку для такого длинного ряда не представляется возможным, интерпретировать полученный результат затруднительно.



(a) ARFIMA(0, d , 0)



(б) ARFIMA(1, d , 0)

Рис. 2.7. Результат работы MC-SSA для ряда Ireland Wind

Заключение

TODO

Список литературы

1. Palma Wilfredo. Long-Memory Time Series: Theory and Methods. — Wiley, 2006.
2. Time Series Analysis: Forecasting and Control / Box G., Jenkins G., Reinsel G., and Ljung G. — Fifth ed. — 2016.
3. Hassler Uwe. Time Series Analysis with Long Memory in View. — Wiley, 2018.
4. Hipel Keith W., McLeod Ian. Time series modelling of water resources and environmental systems. — Elsevier, 1994.
5. Haslett John, Raftery Adrian E. Space-Time Modelling with Long-Memory Dependence: Assessing Ireland's Wind Power Resource // Journal of the Royal Statistical Society. Series C (Applied Statistics). — 1989. — Vol. 38, no. 1. — P. 1–50.
6. Long memory effects and forecasting of earthquake and volcano seismic data / Mariani Maria C., Bhuiyan Md Al Masum, Tweneboah Osei K. and Gonzalez-Huizar Hector // Physica A: Statistical Mechanics and its Applications. — 2020. — Vol. 559. — P. 125049.
7. Barkoulas J., Labys W. C., Onochie J. I. Fractional dynamics in international commodity prices // Journal of Futures Markets. — 1997. — Vol. 17. — P. 161–189.
8. Guglielmo Maria Caporale Luis Gil-Alana, Plastun Alex. Long memory and data frequency in financial markets // Journal of Statistical Computation and Simulation. — 2019. — Vol. 89, no. 10. — P. 1763–1779.
9. McLeod A. I., Yu Hao, Krougly Zinovi. Algorithms for Linear Time Series Analysis: With R Package // Journal of Statistical Software. — 2007. — Vol. 23, no. 5. — Access mode: <https://www.jstatsoft.org/v23/i05/>.
10. Whittle P. The Analysis of Multiple Stationary Time Series // Journal of the Royal Statistical Society. Series B (Methodological). — 1953. — P. 125–139.
11. Team R Core. — R: A Language and Environment for Statistical Computing. — R Foundation for Statistical Computing, Vienna, Austria, 2024. — Access mode: <https://www.R-project.org/>.
12. Veenstra J.Q. — arfima: Fractional ARIMA (and Other Long Memory) Time Series Modeling : 2012.
13. Maechler Martin. — fracdiff: Fractionally Differenced ARIMA aka ARFIMA(P,d,q) Models : 1999.

14. Ларин Е. С. Метод SSA для проверки гипотезы о существовании сигнала во временном ряде : квалификационная работа магистра ; СПбГУ. — 2022.
15. Golyandina N. Detection of signals by Monte Carlo singular spectrum analysis: multiple testing // Statistics and Its Interface. — 2023. — Vol. 16, no. 1. — P. 147–157.
16. Beran J. Statistics for Long-Memory Processes. — Chapman & Hall/CRC, 1994.

Приложение А

Графики

А.1. Сравнение `arfima_mle` и `arfima`

На рис. А.1, А.2 и А.3 представлены среднеквадратичное отклонение, смещение и дисперсия оценок параметров ϕ и d модели ARFIMA(1, d , 0). По ним, видно, что при $\phi = 0.1$ на рис. А.1 оценки функцией `arfima` имеют скачок в $d = 0.4$, что может говорить о некоторой вычислительной неустойчивости функции для больших d . Функция `arfima_mle` не только не имеет подобной проблемы, но и дает более точные оценки, например, при $\phi = 0.5$ на рис. А.2.

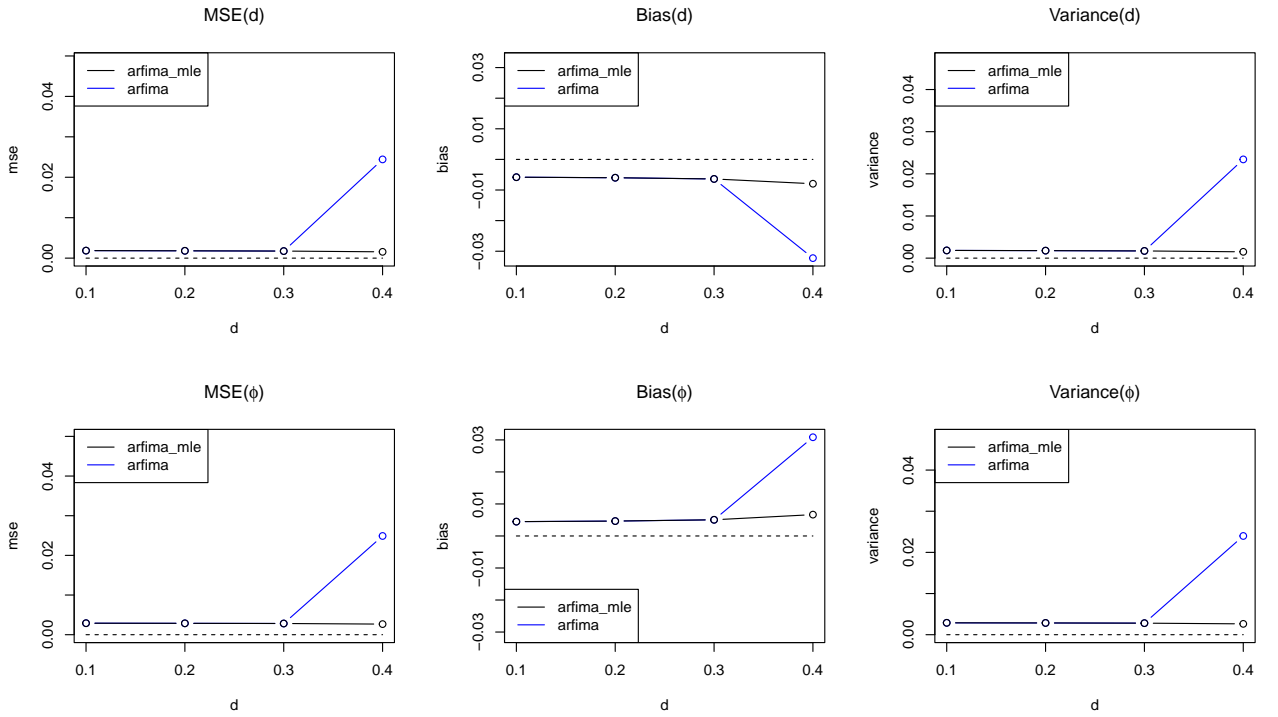


Рис. А.1. Сравнение `arfima_mle` и `arfima` при $\phi = 0.1$ ($n = 1000, 500$ повторений)

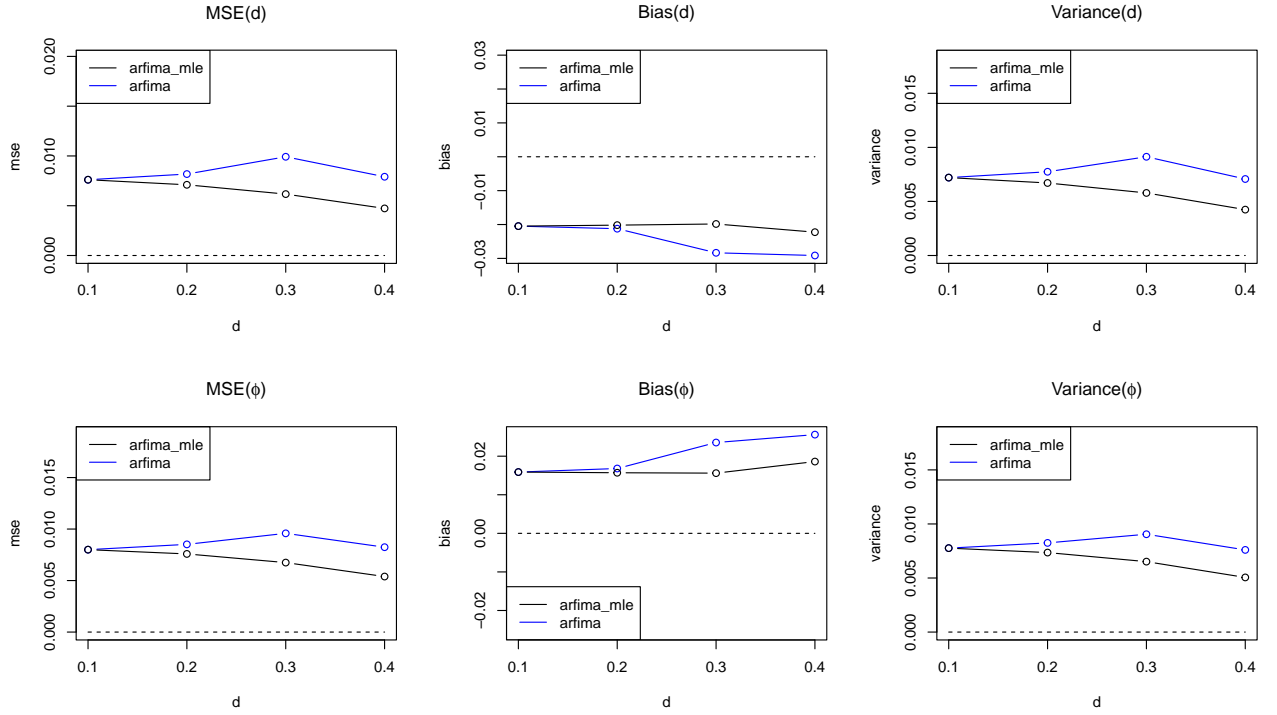


Рис. А.2. Сравнение `arfima_mle` и `arfima` при $\phi = 0.5$ ($n = 1000, 500$ повторений)

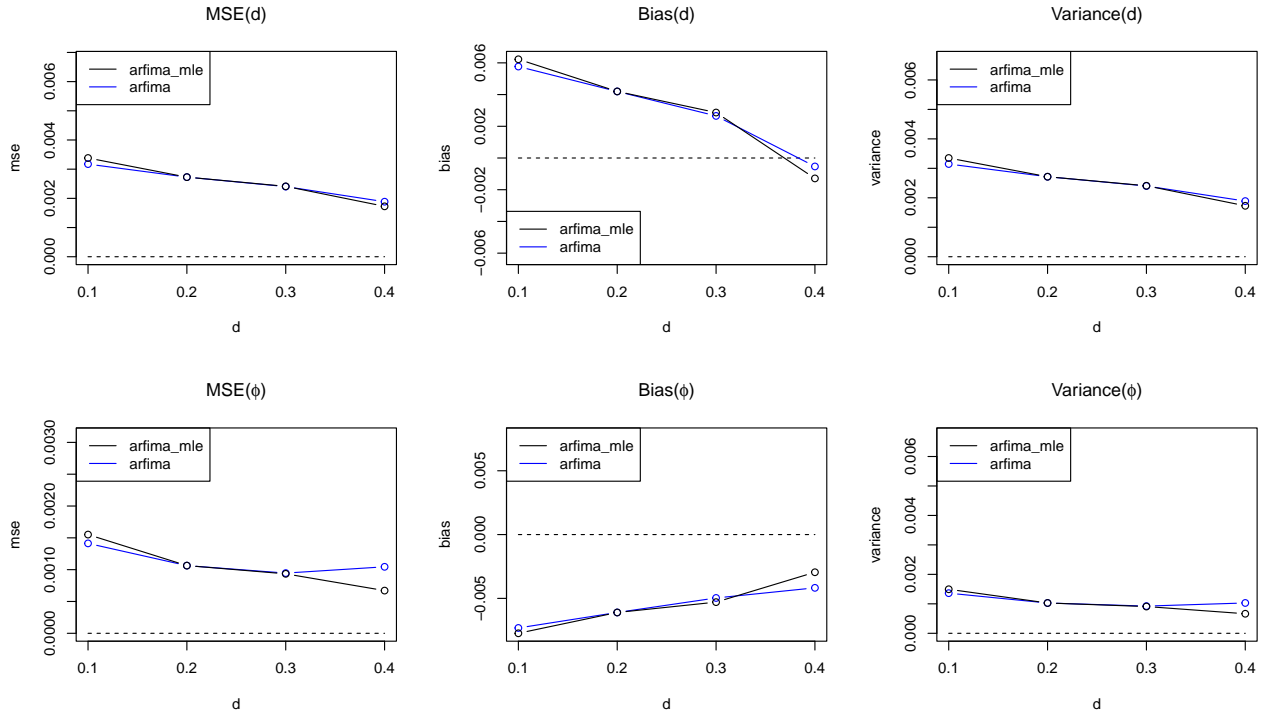


Рис. А.3. Сравнение `arfima_mle` и `arfima` при $\phi = 0.9$ ($n = 1000, 500$ повторений)