Подгонка временных рядов с тяжёлыми хвостами распределений и сильной временной зависимостью

А. Е. МАЗУР
Московский государственный университет
им. М. В. Ломоносова
e-mail: amfolity@gmail.com

УДК 519.217

Ключевые слова: гауссовская последовательность, область максимального притяжения Фреше, эмпирическая квантильная функция.

Аннотация

Ранее была построена модель временного ряда с тяжёлыми хвостами, полученного с помощью преобразования из гауссовского ряда. В настоящей работе решается обратная задача: построена оценка копульной функции, т. е. нелинейного преобразования, применяемого к нормально распределённым случайных величинам и отображающего их в случайные величины с функцией распределения, принадлежащей области притяжения Фреше. Исследованы статистические свойства построенной оценки в случаях действия преобразования на стационарный временной ряд с медленным убыванием корреляции.

Abstract

A. E. Mazur, Fitting time series with heavy tails and strong time dependence, Fundamentalnaya i prikladnaya matematika, vol. 22 (2018), no. 3, pp. 127—144.

Earlier, a model of a time series with heavy tails constructed from a Gaussian time series was developed. In the present paper, the reverse problem is considered: an estimator of the copula function is built; the copula function is a nonlinear function that maps Gaussian variables to the variables from Fréchet maximum domain of attraction. The statistical properties of this estimator are considered for a stationary time series with a low rate of covariance decay.

1. Введение

В работе рассматривается задача построения моделей по данным, предположительно имеющим тяжёлые хвосты, что является нередкой задачей при исследовании финансовых и экономических данных, в теории надёжности, в страховании, а также в некоторых областях физики. Временные ряды с различными хвостами распределений удобно моделировать при помощи гауссовских временных рядов в силу развитой техники работы с ними и простого вида зависимости между наблюдениями. В работе мы рассматриваем гауссовские копульные временные ряды, маргинальные функции распределения которых принадлежат

Фундаментальная и прикладная математика, 2018, том 22, № 3, с. 127—144.

области максимального притяжения Фреше $\mathrm{MDA}(\Phi_\alpha)$. Под гауссовским копульным временным рядом мы понимаем случайную последовательность $X_k = f(\xi_k)$, $k = 1, 2, \ldots$, где ξ_k — гауссовская стационарная последовательность с нулевым средним, единичной дисперсией и ковариационной функцией $R(k)$.

Напомним определение принадлежности функции распределения области притяжения Фреше (см., например, [9]). Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределённых случайных величин, F — функция распределения ξ_i для всех $i \in \mathbb{N}$ и $M_n = \max(\xi_1, \ldots, \xi_n)$. Если для некоторых последовательностей $a_n > 0$, b_n выполнено

$$P\{a_n(M_n - b_n) \leqslant x\} \to G(x),$$

где

$$G(x) = \begin{cases} 0, & x \leqslant 0, \\ \exp(-x^{-\alpha}) \text{ для некоторого } \alpha > 0, & x > 0, \end{cases}$$

то будем говорить, что F принадлежит области притяжения Фреше. Выбирая соответствующим образом копульную функцию $f(x)$, можно получать различные типы хвостов маргинальных распределений временного ряда X_k. В [2] введён и исследован весьма общий вид копульной функции $f(x)$ для получения гауссовского копульного временного ряда, распределение которого принадлежит области максимального притяжения Фреше. В данной модели появляется возможность подгонки модели к данным, имеющим предположительно тяжёлые (степенные) хвосты распределений, принадлежащие области максимального притяжения Фреше.

Для переноса результатов классической теории экстремумов со случая независимых случайных величин на стационарные последовательности используются различные условия на зависимость. Во многих работах используется условие сильного перемешивания Розенблатта при исследовании экстремальных значений (см., например, [1, 4—7, 12], а также литературу этих работ). Довольно часто проверка такого условия на зависимость для далеко отстоящих по времени данных является технически непростой задачей. Если исходные данные удаётся моделировать гауссовским временным рядом, то в качестве меры зависимости можно использовать ковариационную функцию, задача оценивания скорости убывания которой рассмотрена, например, в [3, 11].

Опишем решаемую задачу. Пусть $X_k = f(\xi_k)$, $k \geqslant 1$, где $\xi_1, \xi_2, \ldots$ — стационарный временной ряд стандартных гауссовских случайных величин с ковариационной функцией $\operatorname{cov}(\xi_i, \xi_j) = R(j - i)$, и пусть на эту функцию наложено условие

$$\sum_{k=1}^{\infty} |kR(k)| < \infty. \tag{1}$$

При выполнении этих условий для достаточно больших x построена оценка $\hat{f}(x)$ копульной функции $f(x)$. А именно, пусть y_n — произвольная числовая последовательность, стремящаяся к бесконечности при $n \to \infty$. Тогда

$$\hat{f}(y_n) = X_{n-k_n,n}\left(\frac{k_n}{n\bigl(1-\Phi(y_n)\bigr)}\right)^{\hat{\gamma}}, \tag{2}$$

где $\hat{\gamma}$ — оценка индекса экстремального значения $\gamma = 1/\alpha$, $X_{n-k_n,n}$ — $(n-k_n)$-я порядковая статистика, при этом k_n — *последовательность промежуточного ранга*, т. е. такая последовательность, что выполнено $k_n \to \infty$ и $k_n/n \to 0$ при $n \to \infty$; Φ — функция распределения стандартной нормальной случайной величины. Построенная модель позволяет исследовать экстремальные свойства наблюдаемых данных, оценивая ковариационную функцию с последующим применением асимптотических методов вместо проверки условий перемешивания.

Автор выражает благодарность своему научному руководителю доктору физико-математических наук, профессору Питербаргу Владимиру Ильичу за постановку задачи, постоянную поддержку и внимание к работе. Отдельную благодарность автор приносит Родионову Игорю Владимировичу за полезные обсуждения.

2. Построение оценки для функции f

Пусть F — функция распределения, принадлежащая $\mathrm{MDA}(\Phi_\alpha)$. Возьмём некоторое x_n, превышающее по значению $X_{n-k_n,n}$. Оценим $p_n = 1 - F(x_n) \to 0$ при $n \to \infty$.

Пусть

$$x = \frac{x_n}{X_{n-k_n,n}}.$$

Тогда, пользуясь [9, теорема 1.2.1], получаем

$$p_n := p(x_n) := 1 - F(x_n) \sim \\ \sim \bigl(1 - F(X_{n-k_n,n})\bigr)\left(\frac{x_n}{X_{n-k_n,n}}\right)^{-1/\gamma} \sim \frac{k_n}{n}\left(\frac{x_n}{X_{n-k_n,n}}\right)^{-1/\hat{\gamma}},$$

т. е. оценка $\hat{p}_n$ величины p_n равна

$$\hat{p}_n := p(\hat{x}_n) := \frac{k_n}{n}\left(\frac{x_n}{X_{n-k_n,n}}\right)^{-1/\hat{\gamma}}, \tag{3}$$

и в [9] приведены некоторые её свойства. Здесь $\hat{\gamma}$ — оценка индекса экстремального значения $\gamma = 1/\alpha$.

Из (3) можно получить оценку значения функции F в точке x_n: $\hat{q}(x_n) := 1 - \hat{p}(x_n)$. $\hat{q}(x_n)$ — функция от x_n, и при взятии обратной к ней получаем

$$\hat{q}^{-1}(z) = X_{n-k_n,n}\left((1-z)\frac{n}{k_n}\right)^{-\hat{\gamma}}.$$

Действительно, подставив данное выражение в (3) вместо x_n, получим $1 - z$. Учитывая, что $f(x) = F^{-1}\bigl(\Phi(x)\bigr)$, и подставив в $\hat{q}^{-1}(z)$ аргумент $z = \Phi(y_n)$, где

y_n — произвольная числовая последовательность, стремящаяся к бесконечности при $n \to \infty$, получим оценку $\hat{f}(y_n)$:

$$\hat{f}(y_n) = X_{n-k_n,n} \left(\frac{k_n}{n\big(1 - \Phi(y_n)\big)} \right)^{\hat{\gamma}}. \tag{4}$$

Далее будем обозначать $\tilde{p}_n := 1 - \Phi(y_n)$.

3. Асимптотическая нормальность оценки $\hat{f}$ для стационарного временного ряда

Сформулируем необходимые определения.

Определение 1 [10]. Оценкой Хилла называется следующая статистика:

$$\hat{\gamma}_{\mathrm{H}} := \frac{1}{k_n} \sum_{i=0}^{k_n-1} \log X_{n-i,n} - \log X_{n-k_n,n}.$$

Определение 2 [9]. Оценкой максимального правдоподобия индекса экстремального значения γ, $\hat{\gamma}_{\mathrm{MLE}}$, называется оценка, полученная максимизацией по параметрам γ и σ функции правдоподобия

$$\prod_{i=1}^{k_n} h_{\gamma,\sigma}(z_i),$$

где $z_i = X_{n-i+1,n} - X_{n-k_n,n}$ и $h_{\gamma,\sigma}(x) = \partial H_\gamma(x/\sigma)/\partial x$.

Сформулируем условие второго порядка для функции распределения F. Такое условие определяет скорость сходимости нормализованных экстремальных значений.

Определение 3 (условие второго порядка [9]). Функция распределения F или функция U, связанная с функцией распределения F соотношением

$$U(t) = \left(\frac{1}{1-F} \right)^{-1}(t),$$

удовлетворяют *условию второго порядка*, если существуют $a(t)$ — положительная измеримая функция и функция $A(t)$, не меняющая знак, а также стремящаяся к нулю при $t \to \infty$, такие что выполнено соотношение

$$\lim_{t\to\infty} A(t)^{-1} \left(\frac{U(tx) - U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma} \right) =: H(x), \quad x > 0, \tag{5}$$

где функция $H(x)$ — это функция, которая не представима в виде

$$C \cdot \frac{x^\gamma - 1}{\gamma},$$

где C — некоторая константа. Здесь γ — индекс экстремального значения.

Теперь можем сформулировать основной результат работы.

Теорема 1. *Пусть* $X_1, \ldots, X_n$ *— последовательность одинаково распределённых случайных величин с функцией распределения* $F \in \mathrm{MDA}(\Phi_\alpha)$. *Пусть* $\xi_1, \ldots, \xi_n$ *— стационарный временной ряд стандартных гауссовских случайных величин с ковариацией* $\mathrm{cov}(\xi_i, \xi_j) = R(j-i)$ *с условием* (1). *Пусть функция* f *такая, что* $X_k = f(\xi_k)$, $k = 1, \ldots n$. *Пусть выполнено условие второго порядка* (3) *с* $a(x) = x^\gamma$ *и*

$$H(x) = x^\gamma \frac{x^\rho - 1}{\rho},$$

где $\rho < 0$. *Пусть* y_n *— произвольная числовая последовательность, стремящаяся к бесконечности при* $n \to \infty$. *Пусть для последовательности* k_n *выполнено* $\sqrt{k_n} A(n/k_n) \to \lambda$, *где* $\lambda \in \mathbb{R}$ *при* $n \to \infty$, *и пусть выполнено* $n\tilde{p}_n = o(k_n)$ *и* $\log(n\tilde{p}_n) = o(\sqrt{k_n})$ *при* $n \to \infty$. *Тогда для нормированной оценки* $\hat{f}$ *выполнено*

$$\frac{\sqrt{k_n}}{\log \tilde{d}_n} \left(\frac{\hat{f}(y_n)}{f(y_n)} - 1 \right) \xrightarrow{d} \Gamma, \tag{6}$$

где $\tilde{d}_n = k_n/(n\tilde{p}_n)$ *и* Γ *— нормальная случайная величина, имеющая то же распределение, что и предел последовательности* $\sqrt{k_n}(\hat{\gamma} - \gamma)$.

Здесь и далее будем полагать, что $\hat{\gamma}$ либо оценка Хилла $\hat{\gamma} = \hat{\gamma}_{\mathrm{H}}$, либо оценка максимального правдоподобия $\hat{\gamma} = \hat{\gamma}_{\mathrm{MLE}}$.

4. Формулировка вспомогательных определений и условий

Как и ранее, $X_1, \ldots, X_n$ — стационарный временной ряд, F — функция распределения X_1 и $\bar{F}(x) = 1 - F(x)$. Пусть $\{u_n\}_{n=1}^\infty$ — последовательность уровней, где $u_n \to \infty$ при $n \to \infty$, и $\{\sigma_n > 0\}_{n=1}^\infty$ — нормировочная последовательность.

Нормированным хвостом функции распределения F будем называть

$$T_n(x) = \frac{\bar{F}(u_n + x\sigma_n)}{\bar{F}(u_n)} \quad \text{при} \quad x \geqslant 0, \quad n = 1, 2, \ldots. \tag{7}$$

Нормированный хвост функции распределения F сходится к обобщённому распределению Парето:

$$T_n(x) \to T(x) = \left(1 + \gamma \frac{x}{\sigma}\right)_+^{-1/\gamma} \quad \text{при} \quad x \geqslant 0, \quad n \to \infty, \tag{8}$$

где предельные параметры σ и γ больше 0, а $z_+ := \max(z, 0)$. Из [9, теорема 1.2.5] следует, что можно выбрать $\sigma_n = \gamma u_n$.

Эмпирическим нормированным хвостом функции распределения F будем называть

$$\tilde{T}_n(x) = \frac{1}{n\bar{F}(u_n)} \sum_{i=1}^n I_{\{X_i > u_n + x\sigma_n\}}. \tag{9}$$

Эмпирическим процессом, построенным по хвосту функции распределения F, будем называть

$$e(\tilde{T}_n)(x) = \sqrt{n\bar{F}(u_n)}\big(\tilde{T}_n(x) - T_n(x)\big). \tag{10}$$

Приведём определение коэффициентов перемешивания по Розенблатту. Говорят, что последовательность $\{\xi_n\}$ удовлетворяет *условию сильного перемешивания*, если существует функция $\alpha(k)$, такая что выполнено

$$\alpha(k) = \sup\{|P(AB) - P(A)P(B)|\colon A \in \mathcal{B}_1^l,\ B \in \mathcal{B}_{l+k+1}^{\infty},\ l \geqslant 1\} \to 0 \tag{11}$$

при $k \to \infty$. Здесь $\mathcal{B}_i^j$ — σ-алгебра, порождённая $\xi_i, \ldots, \xi_j$. $\alpha(k)$ будем называть *коэффициентами сильного перемешивания*.

Для изучения поведения случайных величин, превышающих заданный порог, определим σ-алгебру $\mathcal{B}_i^j$, порождённую случайными величинами

$$\xi_i I_{\{\xi_i > f^{-1}(u_n + x\sigma_n)\}}, \ldots, \xi_j I_{\{\xi_j > f^{-1}(u_n + x\sigma_n)\}},$$

где ξ_i, $i \in \mathbb{Z}$, — стационарный временной ряд стандартных гауссовских случайных величин с корреляционной функцией $R(m) = \operatorname{cov}(\xi_1, \xi_{1+m})$. Следуя логике рассуждений [12], разделим числовую прямую на большие блоки длины r_n и маленькие блоки длины l_n, так что

$$r_n = o(n), \quad l_n = o(r_n)$$

при $n \to \infty$. Пусть при этом выполнено, что

$$n\bar{F}(u_n) \to \infty, \quad r_n\bar{F}(u_n) = O(1) \tag{12}$$

при $n \to \infty$. Определим количество попаданий случайных величин в отрезок $[f^{-1}(u_n + x\sigma_n), f^{-1}(u_n + y\sigma_n)]$:

$$N_{n,i}(x, y) = \sum_{k=(i-1)r_n+1}^{ir_n} I_{\{f^{-1}(u_n + x\sigma_n) < \xi_k \leqslant f^{-1}(u_n + y\sigma_n)\}}. \tag{13}$$

Здесь $0 \leqslant x < y < \theta$, где $\theta < \sup\{x\colon T(x) < 1\} = \infty$. Индексация проводится только по случайным величинам, попадающим в блоки длины r_n. Индекс i будем опускать, когда в нём не будет необходимости, и писать $N_n(x, y)$. В этом случае не имеет значения, какой конкретно блок рассматривается.

Приведём условия из [12], проверка выполнимости которых является центральным шагом при доказательстве асимптотической нормальности $\hat{\gamma}$ и оценки нормировочного коэффициента (данная проверка необходима для доказательства асимптотической нормальности $\hat{f}$):

$$\text{C1:}\quad E\{N_n(x,y)^p \mid N_n(x,y)\neq 0\}\leqslant c \ \text{ for some } \ p\geqslant 2, \tag{14}$$

$$\text{C3:}\quad \frac{1}{r_n\bar F(u_n)}\operatorname{cov}\biggl(\sum_{i=1}^{r_n} I_{\{\xi_i>f^{-1}(u_n+x\sigma_n)\}},\sum_{i=1}^{r_n} I_{\{\xi_i>f^{-1}(u_n+y\sigma_n)\}}\biggr)\underset{n\to\infty}{\rightarrow} r(x,y), \tag{15}$$

$$\text{D1:}\quad \frac{1}{n\bar F(u_n)}\sum_{i,j=1}^{[n/r_n]}\bigl|\operatorname{cov}\bigl(N_{n,i}(x,y),N_{n,j}(x,y)\bigr)\bigr|\leqslant K|x-y|^{\mu}, \tag{16}$$

$$\text{D2:}\quad \frac{\alpha(l_n)n}{r_n}\to 0,\quad \frac{n}{r_n}=o\Bigl(\bigl(n\bar F(u_n)\bigr)^{\nu}\Bigr),\quad \alpha(n)=o(n^{\theta}), \tag{17}$$

где c зависит только от θ, $c=c(\theta)$, а константы μ, ν, θ определены в теореме 2 из [12], воспроизведённой ниже.

Во временных рядах экстремумы часто образуют кластеры. Условие C1 ограничивает p-й момент таких кластеров. Условие D1 накладывает ограничение на зависимость между блоками. Условие D2 обеспечивает экспоненциальную скорость убывания по u_n коэффициентов сильного перемешивания.

Для последующего сравнения и обсуждения приведём здесь результат из [12] о сходимости эмпирического процесса к гауссовскому процессу.

Теорема 2. *Предположим, что условие* (8) *и условия* C1, D1, D2, C3 *выполнены, а также* $p/(p-2)<\theta$, $\nu<\theta/2-(\theta+1)/\bigl(2(p-1)\bigr)$ *и* $\mu>2(\theta+p)/\bigl(p(\theta+1)\bigr)$. *Тогда*

$$e(\tilde T_n)(\cdot)\to e(\cdot)\quad \textit{при}\ \ n\to\infty\ \ \textit{в}\ \ D\bigl([0,\infty)\bigr), \tag{18}$$

где $e(\cdot)$ *— непрерывный центрированный гауссовский процесс с ковариационной функцией* $r(x,y)$, *а* $D(I)$ *— пространство действительных функций на конечном или бесконечном интервале* I, *непрерывных справа и имеющих предел слева, т. е. пространство Скорохода, и* $\theta<\sup\{x\colon T(x)<1\}$.

5. Проверка условий C1, C3, D1, D2 при $\lim\limits_{n\to\infty} n\bigl(1-F(u_n)\bigr)=\infty$

Сформулируем условие на зависимость случайных величин, которое будет использовано для доказательства результатов в рамках рассматриваемой модели. Пусть $\delta_{l_n}=\sup\limits_{m\geqslant l_n}|R(m)|<1$, где $R(m)=\operatorname{cov}(\xi_1,\xi_{m+1})$, $\delta=\sup\limits_{m\geqslant 1}|R(m)|<1$, и α — такое число, что $0<\alpha<(1-\delta)/(1+\delta)$. Пусть $l_n=r_n^{\alpha}$, где $r_n=o(n)$.

Будем требовать выполнения следующего условия:

$$\text{D2}'\colon\quad nf^{-1}(u_n+x\sigma_n)\Bigl(\phi\bigl(f^{-1}\bigl(u_n+x\sigma_n\bigr)\bigr)\Bigr)^{(1-\delta_{l_n})/(1+\delta_{l_n})}\sum_{k=l_n}^{\infty}|kR(k)|\to 0 \tag{19}$$

при $n \to \infty$, где ϕ — плотность стандартного нормального распределения. Выбранное обозначение подчёркивает тот факт, что для рассматриваемой модели мы заменяем условие на зависимость D2 на вышеприведённое условие D2$'$.

В этом разделе будем проверять условия, необходимые для доказательства асимптотической нормальности оценки $\hat{f}$ с выполнением условия D2$'$ на зависимость. Перейдём к проверке условий C1, C3, D1 и D2.

Лемма 1. *Условие* C1 *из теоремы* 2 *в выражении* (14) *выполнено в рамках рассматриваемой модели при условии выполнения* D2$'$.

Доказательство. Будем исследовать выполнимость условия

$$\text{C1.1:} \quad \frac{1}{r_n\bar{F}(u_n)}E(N_n(x,y)^p) \leqslant c|x-y| \ \text{ для некоторого } \ p \geqslant 2 \tag{20}$$

при $p = 2$. Согласно [12] при $p \geqslant 2$ условие C1.1 равносильно C1.

Вспомним, что

$$N_{n,i} := N_{n,i}(x,y) := \sum_{k=(i-1)r_n+1}^{ir_n} I_{\{f^{-1}(u_n+x\sigma_n)<\xi_k\leqslant f^{-1}(u_n+y\sigma_n)\}}.$$

Обозначим

$$I_k := I_{\{f^{-1}(u_n+x\sigma_n)<\xi_k\leqslant f^{-1}(u_n+y\sigma_n)\}}.$$

Раскроем левую часть проверяемого условия (20):

$$\begin{aligned} \frac{1}{r_n\bar{F}(u_n)}E\big(N_n(x,y)\big)^2 &= \frac{1}{r_n\bar{F}(u_n)}E\bigg(\sum_{k=(i-1)r_n+1}^{ir_n} I_k\bigg)^2 = \\ &= \frac{1}{r_n\bar{F}(u_n)}\sum_{k=(i-1)r_n+1}^{ir_n} E(I_k) + \frac{1}{r_n\bar{F}(u_n)}\sum_{k=(i-1)r_n+1}^{ir_n}\ \sum_{l=(i-1)r_n+1,\ l\neq k}^{ir_n} E(I_kI_l). \end{aligned} \tag{21}$$

Первое слагаемое выражения (21) можно мажорировать выражением

$$\frac{1}{r_n\bar{F}(u_n)}r_nP\big(\xi_i > f^{-1}(u_n+x\sigma_n)\big) = \frac{\bar{F}(u_n+x\sigma_n)}{\bar{F}(u_n)} = T_n(x) \to T(x)$$

при $n \to \infty$. Последнее стремление выполнено по соотношению (8).

Для любого $k \in [(i-1)r_n+1,\ldots,ir_n]$ разность индексов $k-l$ принимает $r_n - 1$ идущих подряд значений из интервала $\{-(r_n-1),\ldots,(r_n-1)\}$, за исключением нуля. Так как математические ожидания произведений индикаторов неотрицательны, то мы можем ограничить сумму из выражения (21):

$$\frac{1}{r_n\bar F(u_n)}\sum_{k=(i-1)r_n+1}^{ir_n}\ \sum_{l=(i-1)r_n+1,\ l\neq k}^{ir_n} E(I_kI_l)\leqslant$$
$$\leqslant\frac{1}{r_n\bar F(u_n)}\sum_{k=(i-1)r_n+1}^{ir_n}\ \sum_{l-k=-(r_n-1),\ l-k\neq 0}^{r_n-1} E(I_kI_l). \quad (22)$$

Таким образом, переобозначая $m = l - k$ и с учётом симметрии при замене m на $-m$, второе слагаемое в выражении (21) мы можем ограничить следующим выражением, которое может быть мажорировано вероятностью превышения нижнего порога:

$$2\cdot\frac{1}{r_n\bar F(u_n)}\cdot r_n\sum_{m=1}^{r_n-1}E(I_1I_{m+1})\leqslant$$
$$\leqslant 2\cdot\frac{1}{r_n\bar F(u_n)}\cdot r_n\sum_{m=1}^{r_n-1}P\big(\xi_1>f^{-1}(u_n+x\sigma_n),\ \xi_m>f^{-1}(u_n+x\sigma_n)\big). \quad (23)$$

Применяя к (23) предложение 2.4.3 из [3], получаем оценку сверху:

$$\sum_{l=1}^{r_n-1}2\cdot\frac{1}{\bar F(u_n)}\frac{\big(1+R(l)\big)^{3/2}}{2\pi\big(f^{-1}(u_n+x\sigma_n)\big)^2\sqrt{1-R(l)}}\exp\left(-\frac{(f^{-1}(u_n+x\sigma_n)^2)}{1+R(l)}\right). \quad (24)$$

По построению (12) $r_n\big(1-F(u_n)\big)$ при $n\to\infty$ можно мажорировать некоторым конечным K. Также возможен случай $r_n\big(1-F(u_n)\big)\to 0$ при $n\to\infty$. Если $r_n\bar F(u_n)\to 0$ при $n\to\infty$, то $f^{-1}(u_n)\to\infty$ быстрее, чем в случае ограниченного $r_n\bar F(u_n)$. А значит, условия, приведённые выше, также выполнены, так как главный член стремится к нулю быстрее.

Так как $\bar F(u_n+x\sigma_n)/\bar F(u_n)=T_n(x)$, то, пользуясь формулой для асимптотики хвоста нормального распределения

$$1-\Phi(u)\sim\frac{\phi(u)}{u}$$

при $u\to\infty$, где $\phi(\cdot)$ — плотность стандартного нормального распределения, имеем, что при $n\to\infty$

$$\exp\left(-\frac{\big(f^{-1}(u_n+x\sigma_n)\big)^2}{2}\right)=\frac{2\pi\cdot K\cdot\big(f^{-1}(u_n+x\sigma_n)\big)}{r_n}T_n(x)\big(1+o(1)\big). \quad (25)$$

Прологарифмировав левую и правую части выражения (25) и используя то, что $f^{-1}(u_n+x\sigma_n)\to\infty$ при $n\to\infty$, имеем

$$f^{-1}(u_n+x\sigma_n)=(2\ln r_n)^{1/2}+o\big((\ln r_n)^{1/2}\big)\quad\text{при}\quad n\to\infty. \quad (26)$$

Дальнейшее доказательство аналогично доказательству леммы 4.3.2 из [1], где $\delta=\sup\limits_{m\geqslant 1}|R(m)|<1$ и α — такое число, что $0<\alpha<(1-\delta)/(1+\delta)$.

Будем разбивать сумму из выражения (23) на две части: первая часть для $1 \leqslant m \leqslant [r_n^\alpha]$, а вторая часть для $[r_n^\alpha] < m < r_n$, и мажорировать так же, как приведено в доказательстве леммы 4.3.2 из [1]. При этом для исследования второй части будем использовать $\delta_{l_n} = \sup\limits_{m \geqslant l_n} |R(m)|$, где $l_n = r_n^\alpha$. Во вторую часть попадают только те члены, которые находятся на расстоянии не менее l_n друг от друга. Также заметим, что условие Бермана $R(n) \ln n \to 0$ при $n \to \infty$ следует из условий на корреляцию (1). Значит, рассуждения при доказательстве леммы 4.3.2 из [1] полностью применимы и выражение (21) ограниченное. Таким образом, условие C1 выполнено. □

Лемма 2. *Условие* D1 *из теоремы* 2 *в выражении* (16) *выполнено в рамках рассматриваемой модели при условии выполнения* D2′.

Доказательство. Покажем ограниченность суммы из выражения (16). Разобьём данную сумму на две: сумму ковариаций по совпадающим блокам и сумму ковариаций по несовпадающим блокам:

$$\begin{aligned}
\frac{1}{n\bar{F}(u_n)} \sum_{i,j=1}^{[n/r_n]} \bigl|\operatorname{cov}\bigl(N_{n,i}(x,y), N_{n,j}(x,y)\bigr)\bigr| = \\
= \sum_{i=1}^{[n/r_n]} \Biggl| \frac{1}{n\bar{F}(u_n)} E\Biggl(\sum_{k=(i-1)r_n+1}^{ir_n} I_k \sum_{l=(i-1)r_n+1}^{ir_n} I_l \Biggr) - \\
- \frac{1}{n\bar{F}(u_n)} \Biggl(E\Biggl(\sum_{j=(i-1)r_n+1}^{ir_n} I_j \Biggr)\Biggr)^2 \Biggr| + \\
+ \frac{1}{n\bar{F}(u_n)} \sum_{i,j=1, i\neq j}^{[n/r_n]} \bigl|\operatorname{cov}\bigl(N_{n,i}(x,y), N_{n,j}(x,y)\bigr)\bigr|. \qquad (27)
\end{aligned}$$

Перепишем первое слагаемое из первой части (27) как

$$\frac{1}{n\bar{F}(u_n)} \sum_{k=(i-1)r_n+1}^{ir_n} \sum_{l=(i-1)r_n+1}^{ir_n} E(I_k I_l).$$

Как и ранее, обозначим $\delta = \sup\limits_{m \geqslant 1} |R(m)| < 1$, и пусть α — такое число, что $0 < \alpha < (1-\delta)/(1+\delta)$. Аналогично доказательству леммы 1 переходим к одинарной сумме. Разбиваем на сумму по m от 1 до $[r_n^\alpha]$ и сумму от $[r_n^\alpha]$ до r_n. Так как $K = \sup\limits_{n\to\infty} r_n \bar{F}(u_n)$ и так как $\lim\limits_{n\to\infty} T_n(x) = T(x)$, то при проведении аналогичных рассуждений нетрудно показать ограниченность выражения.

Рассмотрим второе слагаемое первой части выражения (27). Вероятность попадания случайной величины в полосу оценим вероятностью превышения нижнего порога:

$$r_n P\big(\xi_i > f^{-1}(u_n + x\sigma_n)\big) = r_n \bar{F}(u_n + x\sigma_n) = r_n T_n(x) \bar{F}(u_n),$$

это выражение ограниченно (см. (12)).

Рассмотрим второе слагаемое из (27). Количество упорядоченных пар блоков равно

$$\left[\frac{n}{r_n}\right] \cdot \left(\left[\frac{n}{r_n}\right] - 1\right).$$

Из этих упорядоченных пар блоков часть пар, а именно $2[n/r_n] - 2$, являются соседними. Расстояние между ближайшими элементами соседних блоков равно $l_n + 1$. Элементы не из соседних блоков находятся на расстоянии не менее чем $r_n + 2l_n$ друг от друга. Разобьём тогда это слагаемое на два по этому признаку, применяя результат теоремы 2.4.3 из [3]. Как и в доказательстве леммы 1, обозначим $\delta_{l_n} = \sup\limits_{m \geqslant l_n} |R(m)|$. Так как случайные величины находятся в соседних блоках, то они находятся на расстоянии не менее $l_n = r_n^\alpha$ друг от друга. Имеем

$$\begin{aligned} &\frac{1}{n\bar{F}(u_n)} \Bigg(\sum_{i,j=1,\ |i-j|=1}^{[n/r_n]} |E(N_{n,i}N_{n,j}) - E(N_{n,i})E(N_{n,j})| + \\ &+ \sum_{i,j=1,\ |i-j|\geqslant 2}^{[n/r_n]} |E(N_{n,i}N_{n,j}) - E(N_{n,i})E(N_{n,j})| \Bigg) \leqslant \frac{2}{n\bar{F}(u_n)} \left[\frac{n}{r_n}\right] r_n \times \\ &\times \sum_{l=r_n^\alpha}^{2r_n + r_n^\alpha} \frac{\big(1 + R(l)\big)^{3/2}}{2\pi\sqrt{1 - R(l)}\big(f^{-1}(u_n + x\sigma_n)\big)^2} \exp\left(-\frac{f^{-1}(u_n + x\sigma_n)^2}{1 + R(l)}\right) - \\ &- \frac{2}{n\bar{F}(u_n)} \sum_{i,j=1,\ |j-i|=1}^{[n/r_n]} \left| \sum_{k=(i-1)r_n+1}^{ir_n} E(I_k) \sum_{l=(j-1)r_n+1}^{jr_n} E(I_l) \right| + \\ &+ \frac{1}{n\bar{F}(u_n)} \sum_{i,j=1,\ |j-i|\geqslant 2}^{[n/r_n]} \Bigg| \sum_{k=(i-1)r_n+1}^{ir_n} \sum_{l=(j-1)r_n+1}^{jr_n} E(I_k I_l) - \\ &- \sum_{k=(i-1)r_n+1}^{ir_n} E(I_k) \sum_{l=(j-1)r_n+1}^{jr_n} E(I_l) \Bigg|. \end{aligned} \tag{28}$$

При таком разбиении на блоки получаем $2[n/r_n]$ соседних блоков, в каждом блоке порядка r_n слагаемых. Для анализа математического ожидания произведения индикаторов случайных величин из соседних блоков применим теорему 2.4.3 из [3].

Итак, для первого слагаемого (28), проводя рассуждения, аналогичные проведенным в доказательстве леммы 1 при $r_n^\alpha = l_n$, получаем оценку сверху $r_n\bar{F}(u_n)T_n^2(x)$ с конечным пределом по построению.

Приведём оценку сверху второго слагаемого выражения (28), применяя рассуждения для оценки выражения $\mathrm{cov}(N_{n,i}, N_{n,i})$ в доказательстве леммы 1:

$$\frac{1}{n\bar{F}(u_n)}\left[\frac{n}{r_n}\right]\big(r_n\bar{F}(u_n)T_n(x)\big)^2 = r_n\bar{F}(u_n)T_n^2(x)\big(1+o(1)\big) \quad \text{при} \quad n\to\infty,$$

выражение ограниченно в пределе по предположению.

Проведём оценку третьего слагаемого выражения (28). Заметим, что каждый член третьего слагаемого ограничивается коэффициентом перемешивания $\alpha(i-j)$ по событиям, отстоящим на расстояние не менее чем $2l_n + r_n$:

$$\frac{1}{n\bar{F}(u_n)}\sum_{i,j=1,\ |j-i|\geqslant 2}^{[n/r_n]}\Bigg|\sum_{k=(i-1)r_n+1}^{ir_n}\ \sum_{l=(j-1)r_n+1}^{jr_n} E(I_kI_l) -$$

$$-\sum_{k=(i-1)r_n+1}^{ir_n} E(I_k)\sum_{l=(j-1)r_n+1}^{jr_n} E(I_l)\Bigg| \leqslant$$

$$\leqslant \frac{1}{n\bar{F}(u_n)}\sum_{i,j=1,\ |j-i|\geqslant 2}^{[n/r_n]}\ \sum_{k=(i-1)r_n+1}^{ir_n}\ \sum_{l=(j-1)r_n+1}^{jr_n}|E(I_kI_l) - E(I_k)E(I_l)|. \quad (29)$$

Так как в рассматриваемом слагаемом расстояние между индексами случайных величин как минимум $r_n + 2l_n$, то каждый такой член $|E(I_kI_l) - E(I_k)E(I_l)|$ можно заменить на коэффициент перемешивания $\alpha(2l_n + r_n)$ по определению коэффициента перемешивания, не накладывая при этом условия сильного перемешивания. Тогда третье слагаемое из выражения (28) ограничивается

$$\frac{1}{n\bar{F}(u_n)}\left[\frac{n}{r_n}\right]^2 r_n^2\alpha(2l_n + r_n) = \frac{n}{\bar{F}(u_n)}\alpha(2l_n + r_n). \quad (30)$$

Для исследования выражения (30) применим оценку сверху для коэффициента перемешивания из теоремы 3.5 из [3]. Введём обозначение для корреляции $\delta_{(r_n+2l_n)} = \sup\limits_{l\geqslant r_n+2l_n} R(l)$. Используя асимптотику для хвоста нормального распределения (см. [1]), а также приведённую оценку сверху для коэффициента перемешивания [3, теорема 3.5], можем оценить сверху (30) следующим выражением:

$$\frac{n}{\bar{F}(u_n)}\alpha(2l_n + r_n) \leqslant K\cdot\frac{nf^{-1}(u_n)}{\exp\big(-\big(f^{-1}(u_n)\big)^2/2\big)}\times$$

$$\times\exp\left(-\frac{\big(f^{-1}(u_n + x\sigma_n)\big)^2}{2}\right)^{2/(1+\delta_{(r_n+2l_n)})}\sum_{k=2l_n+r_n}^{\infty}|kR(k)| =$$

$$= K\cdot nf^{-1}(u_n + x\sigma_n)\Big(\phi\big(f^{-1}(u_n + x\sigma_n)\big)\Big)^{(1-\delta_{(r_n+2l_n)})/(1+\delta_{(r_n+2l_n)})}\times$$

$$\times\sum_{k=2l_n+r_n}^{\infty}|kR(k)|\big(1+o(1)\big), \quad (31)$$

где ϕ — плотность стандартного нормального распределения и $n \to \infty$, K — константа, которая меняется от строки к строке.

Так как $\delta_{(2l_n+r_n)} < \delta_{l_n}$, то $A \to 0$ при $n \to \infty$ при выполнении D2′, что достаточно для верности утверждения леммы 2. □

Лемма 3. *Условие* C3 *из теоремы* 2 *в выражении* (15) *выполнено в рамках рассматриваемой модели при условии выполнения* D2′.

Доказательство. Разложим ковариацию в приведённом условии C3 на две части, пользуясь фактом, что ковариация линейна по каждому из аргументов. Первое из слагаемых является суммой ковариаций индикаторов превышений порогов одной и той же случайной величиной, а второе слагаемое является суммой ковариаций индикаторов превышений порогов различными несовпадающими случайными величинами:

$$\frac{1}{r_n\bar{F}(u_n)}\operatorname{cov}\left(\sum_{i=1}^{r_n} I_{\{\xi_i>f^{-1}(u_n+x\sigma_n)\}},\ \sum_{i=1}^{r_n} I_{\{\xi_i>f^{-1}(u_n+y\sigma_n)\}}\right) =$$
$$= \frac{1}{r_n\bar{F}(u_n)}\left[\sum_{i=1}^{r_n}\operatorname{cov}\bigl(I_{\{\xi_i>f^{-1}(u_n+x\sigma_n)\}},\ I_{\{\xi_i>f^{-1}(u_n+y\sigma_n)\}}\bigr) +\right.$$
$$\left. + 2\sum_{i<j}^{r_n}\operatorname{cov}\bigl(I_{\{\xi_i>f^{-1}(u_n+x\sigma_n)\}},\ I_{\{\xi_j>f^{-1}(u_n+y\sigma_n)\}}\bigr)\right]. \tag{32}$$

Второе слагаемое выражения (32) рассмотрено в доказательстве леммы 1 и является конечным или стремится к нулю, поэтому исследуем асимптотику первого слагаемого:

$$\frac{1}{r_n\bar{F}(u_n)}\sum_{i=1}^{r_n}\operatorname{cov}(I_{\{\xi_i>f^{-1}(u_n+x\sigma_n)\}},\ I_{\{\xi_i>f^{-1}(u_n+y\sigma_n)\}}) =$$
$$= \frac{1}{r_n\bar{F}(u_n)}\sum_{i=1}^{r_n}\Bigl(1 - P\bigl(\xi_i \leqslant f^{-1}(u_n+y\sigma_n)\bigr)\Bigr)P\bigl(\xi_i \leqslant f^{-1}(u_n+x\sigma_n)\bigr) =$$
$$= \frac{1}{r_n\bar{F}(u_n)} r_n\bar{\Phi}\bigl(f^{-1}(u_n+y\sigma_n)\bigr)\Phi\bigl(f^{-1}(u_n+x\sigma_n)\bigr).$$

Так как $y > x$, то $\max(x, y) = y$. Воспользуемся предположением (8) о том, что нормированный хвост функции распределения F $T_n(x)$ сходится к обобщённому распределению Парето при $n \to \infty$. Итак,

$$\frac{1}{r_n\bar{F}(u_n)} r_n\bar{\Phi}\bigl(f^{-1}(u_n+y\sigma_n)\bigr)\Phi\bigl(f^{-1}(u_n+x\sigma_n)\bigr) =$$
$$= T_n(y)F(u_n+x\sigma_n) \underset{n\to\infty}{\to} T(y) = T(x \vee y). \quad \square$$

Лемма 4. *Пусть* $X_1, \ldots, X_n$ — *последовательность одинаково распределённых случайных величин с функцией распределения* $F \in \mathrm{MDA}(\Phi_\alpha)$. *Пусть* $\xi_1, \ldots, \xi_n$ — *стационарный временной ряд стандартных гауссовских случайных* *величин с ковариацией* $\mathrm{cov}(\xi_i, \xi_j) = R(j-i)$ *с условием* (1). *Условие* D2 (*выражение* (17)) *в теореме* 2 *можно заменить на условие* D2′*:*

$$nf^{-1}(u_n + x\sigma_n)\Big(\phi\big(f^{-1}(u_n + x\sigma_n)\big)\Big)^{(1-\delta_{l_n})/(1+\delta_{l_n})} \sum_{k=l_n}^{\infty} |kR(k)| \quad \text{при} \quad n \to \infty,$$

где $\phi(\cdot)$ — *плотность стандартного нормального распределения и* $\delta_{l_n} = \sup\limits_{m \geqslant l_n} R(m)$.

Доказательство. В условии на скорость коэффициента перемешивания из $D2$ (в выражении (17)) $\alpha(l_n)n/r_n \to 0$ при $n \to \infty$. Подставляя оценку для коэффициента $\alpha(\cdot)$ из [3, теорема 3.5] и мажорируя корреляционную функцию $\delta_{l_n} = \sup\limits_{m \geqslant l_n} R(m)$, получаем выражение, которое мы сравним с D2′, и поэтому рассмотрим их частное:

$$\left(\frac{n}{r_n} \exp\left(-\frac{\big(f^{-1}(u_n + x\sigma_n)\big)^2}{1+\delta_{l_n}}\right) \sum_{k=l_n}^{\infty} |kR(k)|\right) \Big/$$

$$\Big/ \left(nf^{-1}(u_n + x\sigma_n)\Big(\phi\big(f^{-1}(u_n + x\sigma_n)\big)\Big)^{(1-\delta_{l_n})/(1+\delta_{l_n})} \sum_{k=l_n}^{\infty} |kR(k)|\right) =$$

$$= \frac{\bar{F}(u_n + x\sigma_n)}{r_n}\big(1 + o(1)\big) \to 0 \quad \text{при} \quad n \to \infty.$$

Последнее всегда верно, так как $\bar{F}(u_n + x\sigma_n) \to 0$, а $r_n \to \infty$ при $n \to \infty$. □

Условия C1, C3, D1 при $\sup\limits_{n \to \infty} n\big(1 - F(u_n)\big) < \infty$ выполнены в рамках рассматриваемой модели, так как из условия на ковариационную функцию следует условие Бермана и соответствующие рассуждения из [12] выполнены.

6. Асимптотическая нормальность оценки f для стационарного временного ряда

Пусть v_n — числовая последовательность, такая что $F(u_n) = 1 - v_n$. Пусть $x^* \in [0,1]$ — такое число, что выполнено равенство

$$x^* = \frac{\bar{F}(u_n + x\sigma_n)}{\bar{F}(u_n)},$$

и тогда

$$0 \leqslant x^* = \frac{\bar{F}(u_n + x\sigma_n)}{\bar{F}(u_n)} = T_n(x) \to T(x) = \Big(1 + \gamma\frac{x}{\sigma}\Big)^{-1/\gamma} \leqslant 1$$

при $n \to \infty$. Тогда эмпирический процесс, построенный по функции распределения F, в выражении (10) можно переписать в виде

$$\bar{e}(\tilde{T}_n)(x) = \frac{1}{\sqrt{nv_n}} \sum_{i=1}^{n} \big(I_{\{U_i > 1 - v_n \cdot x^*\}} - v_n \cdot x^*\big),$$

где $U_k = F(X_k)$, $k = 1, \ldots, n$, равномерно распределены на отрезке $[0,1]$. Иногда $\bar{e}(\tilde{T}_n)(x)$ будем обозначать как $\bar{e}_n(\cdot)$ или $\bar{e}_n(x^*)$. Обозначим предельный процесс при $n \to \infty$ для $\bar{e}(\tilde{T}_n)(\cdot)$ как $\bar{e}(\cdot)$, корреляционную функцию процесса $\bar{e}(\cdot)$ как $\bar{r}(x^*, y^*)$, где $x^*, y^* \in [0,1]$. Здесь y^* определяется аналогично x^* для $y \geqslant 0$.

Сформулируем лемму асимптотической нормальности $\hat{U}(n/k_n) = X_{n-k_n,n}$, результат которой тривиально следует из рассуждений, приведённых в [12], так как доказаны леммы 1—4.

Лемма 5. *Пусть $X_1, \ldots, X_n$ — последовательность одинаково распределённых случайных величин с функцией распределения $F \in \mathrm{MDA}(\Phi_\alpha)$. Пусть $\xi_1, \ldots, \xi_n$ — стационарный временной ряд стандартных гауссовских случайных величин с ковариацией $\mathrm{cov}(\xi_i, \xi_j) = R(j-i)$ с условием* (1). *Тогда оценка $\hat{U}(n/k_n) = X_{n-k_n,n}$ асимптотически нормальна, т. е. выполнено*

$$\sqrt{k_n}\left(\frac{X_{n-k_n,n} - F^{-1}(1-k_n/n)}{a(n/k_n)}\right) \xrightarrow{d} N\big(0, \bar{r}(1,1)\big) \quad \text{при} \quad n \to \infty,$$

где $a(t)$ — нормирующая последовательность, определённая в [9, теорема 1.1.6].

Сформулируем лемму об асимптотической нормальности для оценки максимального правдоподобия и оценки Хилла индекса экстремального значения. Доказательство следует из рассуждений, приведённых в [4—6], так как верны результаты лемм 1—4.

Лемма 6. *Пусть $X_1, \ldots, X_n$ — последовательность одинаково распределённых случайных величин с функцией распределения $F \in \mathrm{MDA}(\Phi_\alpha)$. Пусть $\xi_1, \ldots, \xi_n$ — стационарный временной ряд стандартных гауссовских случайных величин с ковариационной функцией $\mathrm{cov}(\xi_i, \xi_j) = R(j-i)$ с условием* (1). *Тогда оценка максимального правдоподобия индекса экстремального значения $\hat{\gamma} = \hat{\gamma}_{\mathrm{MLE}}$ и оценка Хилла индекса экстремального значения $\hat{\gamma} = \hat{\gamma}_{\mathrm{H}}$ асимптотически нормальны*:

$$\sqrt{k_n}(\hat{\gamma}_{\mathrm{MLE}} - \gamma) \xrightarrow{d} N\big(0, (\gamma+1)^2 \bar{r}(1,1)\big) \quad \text{при} \quad n \to \infty.$$

Пусть

$$\lambda = \lim_{n\to\infty} k_n^{1/2}\left(\frac{F^{-1}(1-k_n/n)}{a(k_n/n)} - \frac{1}{\gamma}\right) \in \mathbb{R}.$$

Тогда

$$\sqrt{k_n}(\hat{\gamma}_{\mathrm{H}} - \gamma) \xrightarrow{d} N\big(\lambda \nu_{H,\gamma}[0,1], \gamma^2 \bar{r}(1,1)\big) \quad \text{при} \quad n \to \infty,$$

где мера $\nu_{H,\gamma}[0,1]$ определена в [6], *$a(t)$ — нормирующая последовательность, определённая в* [9, теорема 1.1.6].

Доказательство теоремы 1. Ход доказательства теоремы 1 повторяет логику рассуждений доказательства теоремы 4.3.8 из [9] при замене p_n на $\tilde{p}_n = 1 - \Phi(y_n)$. При такой замене можно записать

$$U\left(\frac{1}{1-\Phi(y_n)}\right) = U\left(\frac{1}{\tilde{p}_n}\right)$$

и
$$\tilde{d}_n = \frac{k_n}{n\tilde{p}_n}.$$
При доказательстве будем опираться на результат леммы 6 вместо 3.2.5 и 3.4.2 из [9]. А именно, преобразуем (6), подставляя явный вид для оценки $\hat{f}(y_n)$:
$$\begin{aligned}\frac{\sqrt{k_n}}{\log\tilde{d}_n}\left(\frac{\hat{f}(y_n)}{f(y_n)}-1\right) &= \frac{1}{U(1/\tilde{p}_n)}\,\frac{\sqrt{k_n}}{\log\tilde{d}_n}\left(X_{n-k_n,n}\tilde{d}_n^{\hat{\gamma}} - U\left(\frac{1}{\tilde{p}_n}\right)\right) = \\ &= \frac{\tilde{d}_n^{\gamma}U(n/k_n)}{U(1/\tilde{p}_n)}\left(\frac{\sqrt{k_n}}{\log\tilde{d}_n}\left(\frac{X_{n-k_n,n}}{U(n/k_n)}-1\right)\tilde{d}_n^{\hat{\gamma}-\gamma} + \frac{\sqrt{k_n}}{\log\tilde{d}_n}(\tilde{d}_n^{\hat{\gamma}-\gamma}-1) - \right.\\ &\left. - \frac{\sqrt{k_n}A(n/k_n)}{\log\tilde{d}_n}\left(\left(\frac{U(1/\tilde{p}_n)\tilde{d}_n^{-\gamma}}{U(n/k_n)}-1\right)\Big/ A\left(\frac{n}{k_n}\right)\right)\right),\end{aligned} \tag{33}$$
где $A(t)\to 0$ при $t\to 0$ и
$$\lim_{n\to\infty}\sqrt{k_n}A\left(\frac{n}{k_n}\right) = \lambda < \infty.$$

Покажем, что второе слагаемое стремится к нормальной случайной величине по распределению. Для этого покажем, что
$$\frac{\sqrt{k_n}}{\log\tilde{d}_n}(\tilde{d}_n^{\hat{\gamma}-\gamma}-1) \xrightarrow{d} \Gamma,$$
где Γ — нормальная случайная величина с параметрами, определёнными в лемме 6. По лемме 6
$$\sqrt{k_n}(\hat{\gamma}-\gamma) \xrightarrow{d} \Gamma$$
при $n\to\infty$.

Пусть g — некоторая функция, $g(\gamma)=\tilde{d}_n^{\gamma}$ для достаточно большого n. В точке γ существует производная любого порядка. Будем обозначать производную m-го порядка функции $g(\cdot)$ через $g^{(m)}(\cdot)$. Имеем
$$vg^{(m)}(\gamma) = (\tilde{d}_n^{\gamma})^{(m)} = (\log\tilde{d}_n)^m\tilde{d}_n^{\gamma}.$$
Функцию $g(\hat{\gamma})=\tilde{d}_n^{\hat{\gamma}}$ разложим в ряд Тейлора в точке γ:
$$\tilde{d}_n^{\hat{\gamma}} = \tilde{d}_n^{\gamma} + \frac{\partial(\tilde{d}_n^{\gamma})}{\partial\gamma}(\hat{\gamma}-\gamma) + \frac{\partial^2(\tilde{d}_n^{\gamma})}{\partial\gamma^2}\,\frac{(\hat{\gamma}-\gamma)^2}{2}\big(1+o(1)\big) \quad \text{при} \quad n\to\infty.$$
Так как $\log\tilde{d}_n/\sqrt{k_n}\to 0$ при $n\to\infty$, то при условии $\log n\tilde{p}_n = o(\sqrt{k_n})$ выполнено $\tilde{d}_n^{\hat{\gamma}} \xrightarrow{d} \tilde{d}_n^{\gamma}$ при $n\to\infty$.

Рассмотрим первое слагаемое
$$\frac{\sqrt{k_n}}{\log\tilde{d}_n}\left(\frac{X_{n-k_n,n}}{U(n/k_n)}-1\right)\tilde{d}_n^{\hat{\gamma}-\gamma}.$$

Из леммы 5 получаем, что

$$\sqrt{k_n}\left(\frac{X_{n-k_n,n}}{U(n/k_n)}-1\right)\xrightarrow{d}\gamma B$$

при $n\to\infty$, где B — нормальная случайная величина с параметрами, определёнными в лемме 5. При этом

$$\frac{1}{\log\tilde{d}_n}\tilde{d}_n^{\hat{\gamma}-\gamma}\xrightarrow{d}0$$

при $n\to\infty$. Таким образом, первое слагаемое стремится к нулю по распределению.

Рассмотрим третье слагаемое

$$\frac{\sqrt{k_n}A(n/k_n)}{\log\tilde{d}_n}\left(\left(\frac{U\big((1/\tilde{p}_n)\tilde{d}_n^{-\gamma}\big)}{U(n/k_n)}-1\right)\Big/ A\left(\frac{n}{k_n}\right)\right).$$

По [9, теорема 2.3.9] выполнено

$$\lim_{n\to\infty}\left(\frac{U\big((n/k_n)\tilde{d}_n\big)}{U(n/k_n)}\tilde{d}_n^{-\gamma}-1\right)\Big/ A\left(\frac{n}{k_n}\right)=-\frac{1}{\rho}.$$

В условиях доказываемой теоремы

$$\lim_{n\to\infty}\sqrt{k_n}A\left(\frac{n}{k_n}\right)=\lambda$$

и $\log\tilde{d}_n\to\infty$ при $n\to\infty$. Из только что приведённого предельного соотношения также следует, что

$$\frac{1}{U(1/\tilde{p}_n)}\tilde{d}_n^{\gamma}U\left(\frac{n}{k_n}\right)\to 1$$

при $n\to\infty$. Поэтому третье слагаемое выражения (33) стремится к нулю при $n\to\infty$. Таким образом, получаем, что

$$\frac{\sqrt{k_n}}{\log\tilde{d}_n}\left(\frac{\hat{f}(y_n)}{f(y_n)}-1\right)\xrightarrow{d}\Gamma\quad\text{при}\quad n\to\infty.\qquad\square$$

Литература

[1] Лидбеттер М. Р., Линдгрен Г., Ротсен Х. Экстремумы случайных последовательностей и процессов. — М.: Мир, 1989.

[2] Мазур А. Е., Питербарг В. И. Гауссовские копульные временные ряды с тяжёлыми хвостами и сильной временной зависимостью // Вестн. Моск. ун-та. Сер. 1. Математика, механика. — 2015. — Т. 70, № 5. — P. 197—201.

[3] Питербарг В. И. Двадцать лекций о гауссовских процессах. — М.: МЦНМО, 2015.

[4] Drees H. A general class of estimators of the extreme value index // J. Statist. Planning Inference. — 1998. — Vol. 66, no. 1. — P. 95—112.

[5] Drees H. On smooth statistical tail functionals // Scand. J. Statist. — 1998. — Vol. 25, no. 1. — P. 187—210.

[6] Drees H. Weighted approximation of tail processes for β-mixing random variables // Ann. Appl. Probab. — 2000. — Vol. 10, no. 4. — P. 1274—1301.

[7] Drees H. Tail empirical processes under mixing conditions // Empirical Process Techniques for Dependent Data / Dehling H., Mikosch T., Sørensen M., eds. — Boston: Birkhäuser, 2002. — P. 325—342.

[8] Gill R. D. Non- and semi-parametric maximum likelihood estimators and the von Mises method (part 1) // Scand. J. Statist. — 1989. — Vol. 16, no. 2. — P. 97—128.

[9] De Haan L., Ferreira A. Extreme Value Theory: An Introduction. — New York: Springer, 2007.

[10] Hill B. M. A simple general approach to inference about the tail of a distribution // Ann. Statist. — 1975. — Vol. 3, no. 5. — P. 1163—1174.

[11] Piterbarg V. I. Asymptotic Methods in Theory of Gaussian Random Processes and Fields. — Providence: Amer. Math. Soc., 2012. — (Transl. Math. Monogr.; Vol. 148).

[12] Rootzén H. Weak convergence of the tail empirical process for dependent sequences // Stoch. Processes Their Appl. — 2009. — Vol. 119, no. 2. — P. 468—490.