Факторизация натуральных чисел

Михаил Иванов

10 марта 2020 г.

Факторизация и тест простоты

Центральные множества

- $\mathbb{N} = \{1, 2, 3, 4, 5 \ldots\}$
- $\mathbb{P} = \{2, 3, 5, 7, 11, \ldots\}$

Факторизация и тест простоты

Основные задачи

- $n \in \mathbb{N} \setminus \{1\}$
- Тест непростоты: $n \notin \mathbb{P}$
- Переформулировка:
  $\exists n_1, n_2 \in \{2, \ldots, n-1\}\colon n = n_1 n_2$
- Факторизация:
  найти $p_1, \ldots, p_k \in \mathbb{P}\colon n = \prod_{i=1}^{k} p_i$

Факторизация и тест простоты

Сведение

- тест простоты $\rightarrow$ факторизация
- $n \in \mathbb{P} \iff$
  $\iff \left(p_1, \ldots, p_k \in \mathbb{P} \colon n = \prod_{i=1}^{k} p_i \Rightarrow k = 1 \right)$
- Обязательно ли факторизовывать?

Факторизация и тест простоты

Отсутствие необходимости факторизации

- Малая теорема Ферма: если $n \in \mathbb{P}$, $a \not\equiv_n 0$, то $a^{n-1} \equiv_n 1$
- Возьмём много случайных a и проверим это
- a^{n-1} вычисляется двоичным возведением в степень
- Победа?

Факторизация и тест простоты

Тест Ферма не работает

- Не победа

Факторизация и тест простоты

Числа Кармайкла

- Описанная процедура называется *тестом Ферма*
- Число Кармайкла — n, такое что $n \notin \mathbb{P} \wedge (a, n) \equiv 1 \iff a^{n-1} \underset{n}{\equiv} 1$
- На числах Кармайкла тест Ферма не работает

Факторизация и тест простоты

Тест Миллера-Рабина

- Действуем аккуратнее
- $x^2 \underset{n}{\equiv} 1 \Rightarrow x \in \{-1_n, 1_n\}$ только при $n \in \{1\} \cup \mathbb{P}$
- Пусть $n - 1 = 2^k m$
- Малая теорема Ферма: если $n \in \mathbb{P}$, $a \not\underset{n}{\equiv} 0$, то $a^{n-1} \underset{n}{\equiv} 1$
- Малая теорема Ферма: если $n \in \mathbb{P}$, $a \not\underset{n}{\equiv} 0$, то $\left(a^m\right)^{2^k} \underset{n}{\equiv} 1$

Факторизация и тест простоты

Тест Миллера-Рабина

- Найдём $a^m, (a^m)^2, (a^m)^4, (a^m)^8, \ldots, (a^m)^{2^k}$
- Если в конце не получилось 1, то $n \notin \mathbb{P}$
- Рассмотрим первый момент, когда получилось 1
- Если это $(a^m)^{2^t}$, $t > 0$, посмотрим на шаг назад
- Если $(a^m)^{2^{t-1}} \not\equiv_n -1$, то $n \notin \mathbb{P}$

Факторизация и тест простоты

Тест простоты лежит в классе P

- Для составных n случайное a обнаружит непростоту с вероятностью 75%
- Одна итерация работает за $\mathcal{O}\left(\log^2 n \log\log n \log\log\log n\right)$
- Рандомизированно $\mathcal{O}\left(\log^3 n\right)$
- Тест Агравала — Каяла — Саксены (2004), модификация Ленстры и Померанса (2005) — детерминированно $\mathcal{O}\left(\log^6\right)$
- Тест простоты имеет полиномиальную сложность (от длины числа на входе)

Факторизация

Нахождение нетривиального делителя

- Вернёмся к факторизации
- Ищем $p_1, \ldots, p_k \in \mathbb{P}$: $n = \prod_{i=1}^{k} p_i$
- Пусть $n \notin \mathbb{P}$ умеем представлять в виде $n_1 n_2$, $n_1, n_2 \in \{2, \ldots, n-1\}$
- Процедура факторизации рекурсивна:
  - Если n простое, вернуть n
  - Иначе представить $n = n_1 n_2$
  - Рекурсивно представить $n_1 = \prod_{i=1}^{k} p_i$, $n_2 = \prod_{i=k+1}^{\ell} p_i$
  - Склеить: $n = \prod_{i=1}^{\ell} p_i$

Факторизация

Тривиальные алгоритмы

- Достаточно научиться находить нетривиальный делитель у составного числа
- $\mathcal{O}(n)$: перебрать $n_1 \in \{2, 3, \ldots, n-1\}$, проверить $n_2 = \frac{n}{n_1} \in \mathbb{N}$
- $\mathcal{O}\left(\sqrt{n}\right)$: так как $\min\left(n_1, n_2\right) \leqslant \sqrt{n}$, перебрать $n_1 \in \{2, 3, \ldots, \left[\sqrt{n}\right]\}$, проверить $n_2 = \frac{n}{n_1} \in \mathbb{N}$
- $\mathcal{O}\left(\sqrt[4]{n}\log n\right)$ — ρ-алгоритм Полларда

ρ-алгоритм Полларда

Парадокс дней рождения

- C детей в классе, D дней в году, $C \ll D$
- D^C способов
- $D(D-1)\dots(D-C+1)$ способов с разными днями
- $\mathbb{P}\{\text{разных дней рождения}\} = \frac{D(D-1)\dots(D-C+1)}{D^C}$
- $1\left(1-\frac{1}{D}\right)\dots\left(1-\frac{C-1}{D}\right) < \varepsilon$

ρ-алгоритм Полларда

Продолжение парадокса дней рождения

- $\ln 1 \left(1 - \frac{1}{D}\right) \ldots \left(1 - \frac{C-1}{D}\right) < \ln \varepsilon$
- $\sum_{i=0}^{C-1} \ln \left(1 - \frac{i}{D}\right) < -E$
  - $\ln(1 + x) \approx x$
- $\sum_{i=0}^{C-1} -\frac{i}{D} < -E$
  - $\sum_{i=0}^{n-1} i \approx \frac{n^2}{2}$
- $-\frac{C^2}{2D} < -E$
- $C > \sqrt{2DE}$

ρ-алгоритм Полларда

Применение к факторизации, метод за $\mathcal{O}\left(\sqrt{n}\log n\right)$

- $n = n_1 n_2$, $n_1 \leqslant \sqrt{n} \leqslant n_2$
- Рассмотрим случайные $a_1, \ldots, a_C$
- Пусть $D = n_1$
- $a_1, \ldots, a_C$ — «дети» с днями рождения $a_i \mod D$ среди $\{0, 1, \ldots, D-1\}$
- Если $C \approx \sqrt{\mathrm{Const} D} \leqslant \sqrt[4]{n}$, то есть $a_i \mod D = a_j \mod D$
- $(a_i - a_j, n)$ кратно D.

ρ-алгоритм Полларда

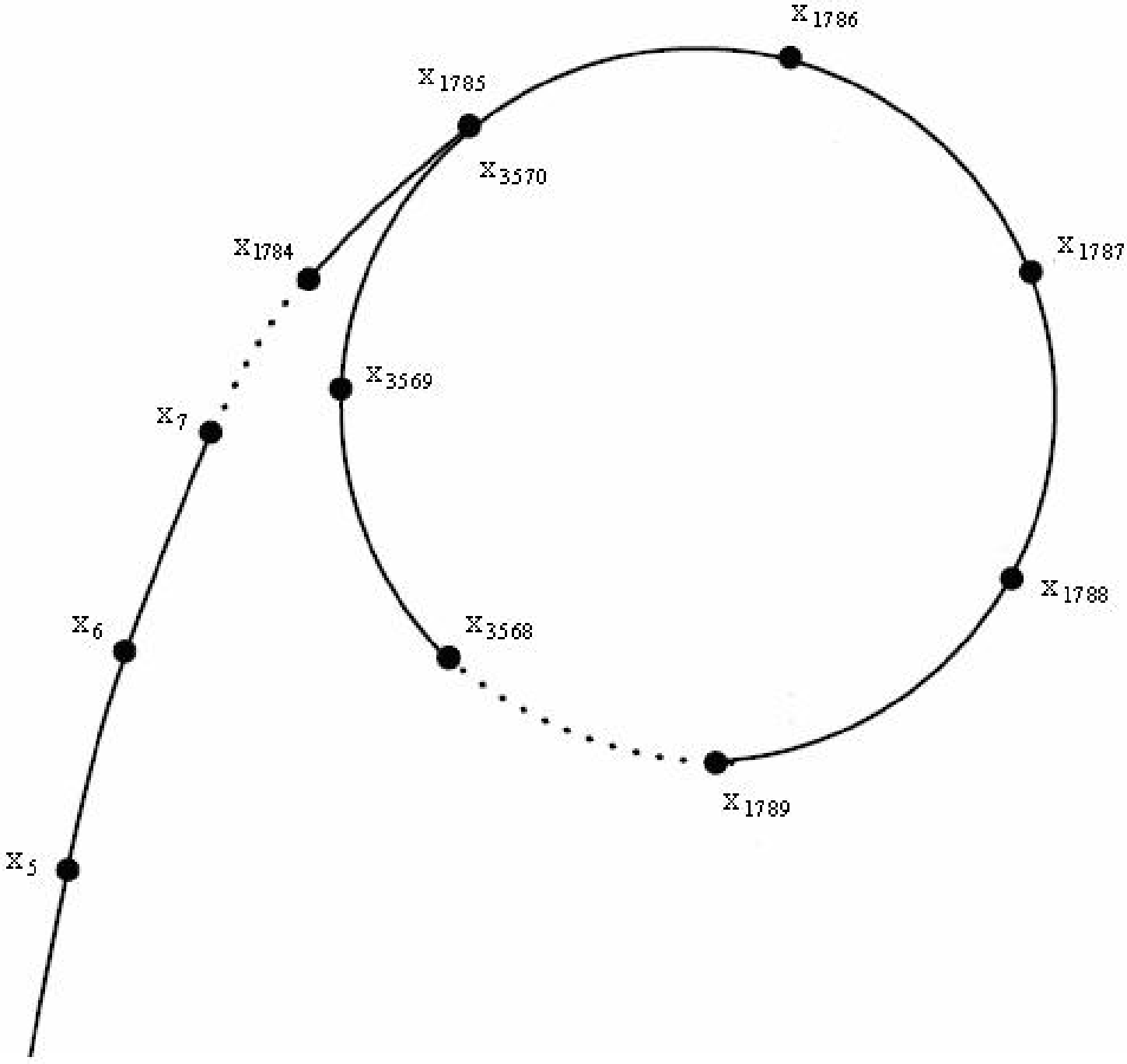
Применение к факторизации, метод за $\mathcal{O}\left(\sqrt{n}\log n\right)$

- $\mathbb{P}\left\{(a_i - a_j, n) = n\right\} = \frac{1}{n_2} < \frac{1}{\sqrt{n}}$
- Итого, при достаточно большом Const число $(a_i - a_j, n)$ с большой вероятностью — нетривиальный делитель n
- Алгоритм Евклида — $\mathcal{O}(\log n)$
- $\forall i, j$ сравнить $(a_i - a_j, n)$ с 1 и n
- $\mathcal{O}\left(\sqrt{n}\log n\right)$ — медленно

ρ-алгоритм Полларда

Применение к факторизации, метод за $\mathcal{O}\left(\sqrt[4]{n}\log n\right)$

- Не любая последовательность $\{a_i\}_{i\in\mathbb{N}}$ хороша
- Свойство остатков: $a_i \underset{n_1}{\equiv} a_j \Rightarrow a_{i+1} \underset{n_1}{\equiv} a_{j+1}$
- Например, $a_{i+1} = f\left(a_i\right)$, $f(x) = x^2 + 1$
- Эвристика: $\{a_i\}_{i\in\mathbb{N}}$ всё ещё достаточно случайна для парадокса дней рождения
- Почему ρ-алгоритм Полларда?


X1786
X1785
X3570
X1784
X1787
X3569
X7
X1788
X6
X3568
X1789
X5

ρ-алгоритм Полларда

Метод за $\mathcal{O}\left(\sqrt[4]{n}\log n\right)$

- $a_i \underset{n_1}{\equiv} a_j$ с наименьшим $j > i$
- В среднем $j \approx \mathrm{Const}\sqrt[4]{n}$
- Предпериод $P = i$, период $Q = j - i$
- $P + Q$ порядка $\mathrm{Const}\sqrt[4]{n}$
- Хотим два элемента a_x, a_y, где $x < y$, $y - x \vdots Q$, $x \geqslant P$

ρ-алгоритм Полларда

Метод за $\mathcal{O}\left(\sqrt[4]{n}\log n\right)$

- Будем брать (a_i, a_{2i})
- $a_{i+1} = f\left(a_i\right)$, $a_{2(i+1)} = f\left(f\left(a_{2i}\right)\right)$
- Условия для $x = i, y = 2i$:
  - $x < y$: $i < 2i$
  - $y - x \vdots Q$: надо $i \vdots Q$
  - $x \geqslant P$: $i \geqslant P$
- Достаточно $i \in \{1, 2, \ldots, P + Q\}$

ρ-алгоритм Полларда

Итоговый алгоритм

- $f(x) = x^2 + 1$ (можно $x^2 + c$, где c случайное)
- Инициализируем a, b одинаковым случайным числом
- Повторяем $\mathrm{Const}\sqrt[4]{n}$ раз:
  - $a := f(a), b := f(f(b)), g := (a, b)$
  - Если $g \notin \{1, n\}$, вернуть нетривиальный делитель g
- Повторить с другими случайными $a = b$ и c

$p-1$-алгоритм Полларда

Идея

- Опять вспомним малую теорему Ферма (сегодня без неё никуда): если $p \in \mathbb{P}$, $(a,p)=1$, то $a^{p-1} \underset{p}{\equiv} 1$.
- Следствие: если ещё и $M \vdots p-1$, то $a^{M} \underset{p}{\equiv} 1$.
- Пусть $M(i) = \operatorname{lcm}(1,2,\ldots,i)$
- Возьмём некоторое B и $M = M(B)$
- Если $p-1$ кратно маленьким степеням простых, то $a^{M} \underset{p}{\equiv} 1$.

$p-1$-алгоритм Полларда

Алгоритм

- Выбрать планку B
- Решетом Эратосфена найти все простые $p_1 \ldots, p_m \leqslant B$ в максимальных степенях $p_i^{\alpha_i} \leqslant B$
- Для нескольких случайных c повторять:
  - Для каждого i заменить c на $c^{p_i^{\alpha_i}}$ (α_i раз возвести c в степень p_i)
  - $g := (c-1, n)$
  - Если $g \notin \{1, n\}$, вернуть нетривиальный делитель g

$p-1$-алгоритм Полларда

Хитрости

- Если для всех p, делящих n, $p-1$ кратно маленьким степеням простых, то g будет равно n
- Брать нечётное ядро $M(B)$

$p-1$-алгоритм Полларда

Вторая фаза

- Если для всех p, делящих n, $p-1$ кратно большим степеням простых, то g будет равно 1
- Нужна вторая фаза
- Рассмотрим $B' \approx B \ln B$
- Пусть $Q_1 < Q_2 < \ldots < Q_t$ — все простые в $(B; B')$
- Пробуем $M_i = Q_i M(B)$

$p-1$-алгоритм Полларда

Перебор всех M_i

- Чтобы перейти от a^{M_i} к $a^{M_{i+1}}$, нужно домножить на $a^{M_{i+1}-M_i}$
- $a^{Q_{i+1}M(B)-Q_iM(B)}$
- $\left(a^{M(B)}\right)^{Q_{i+1}-Q_i}$
- $Q_{i+1}-Q_i$ малы
- Предподсчитаем $\left(a^{M(B)}\right)^e$ для малых e
- Пересчитываем $a^{M_{i+1}}$ через a^{M_i} за $\mathcal{O}(1)$
- При $B' \approx B \ln B$ первая и вторая фаза займут равное число времени

Эллиптические кривые

Быстрая факторизация

- Всё предыдущее работало за $\mathcal{O}\left(n^{\alpha}\right)$
- Перепишем: за $\mathcal{O}\left(e^{c \ln n}\right)$

Эллиптические кривые

L-нотация

- Обозначим $L_n[\alpha, c] = \exp\left((c + o(1)) \ln^{\alpha} n \ln^{1-\alpha} \ln n\right)$ при $n \to +\infty$, $c \in (0; +\infty)$, $\alpha \in [0; 1]$
- Тривиальная факторизация: $L_n\left[1, \frac{1}{2}\right]$
- ρ-алгоритм Полларда: $L_n\left[1, \frac{1}{4}\right]$
- Метод Ленстры факторизации с помощью эллиптических кривых: $L_p\left[\frac{1}{2}, \sqrt{2}\right]$, где p — наименьший простой делитель n
- С помощью оценки $p \leqslant \sqrt{n}$: $L_n\left[\frac{1}{2}, 1\right]$

Эллиптические кривые

Эллиптические кривые

- Пусть F — кубический многочлен от двух переменных (формально, кубический однородный от трёх, не кратный z)
- Проективная плоскость — множество (x, y, z), $x^2 + y^2 + z^2 \neq 0$, с точностью до домножения
- $F(x, y) = 0$ — кривая Γ на проективной плоскости
- Предположим, что класса C^∞ (без самопересечений и точек с нулевой производной)
- Тогда кривая называется эллиптической

Эллиптические кривые

Эллиптические кривые

- После проективного преобразования:
- $F(x, y) = y^2 - \left(x^3 + ax + b\right)$
- Заведём операцию $+$ на кривой:
  - коммутативная
  - ассоциативная
  - $a + b + c = 0$ для коллинеарных точек
- Заведём операцию $-$, противоположную $+$
- Получилась абелева группа

Эллиптические кривые

Умножение

- Заведём операцию $[n]a$, $n \in \mathbb{N}_0$, $a \in \Gamma$ — умножение на целое
- Бинарное умножение:
  - $n \vdots 2 \Rightarrow [n]a = \left[\frac{n}{2}\right] a + \left[\frac{n}{2}\right] a$
  - иначе $[n]a = \left[\frac{n-1}{2}\right] a + \left[\frac{n-1}{2}\right] a + a$

Эллиптические кривые

Эллиптическая псевдокривая

- Определим всё сказанное в кольце $\mathbb{Z}_n$
- $(n, 6) = 1$, $\left(4a^3 + 27b^2, n\right) = 1$, $E_{a,b}(\mathbb{Z}_n) = \left\{(x, y) \in \mathbb{Z}_n \times \mathbb{Z}_n \mid y^2 = x^3 + ax + b\right\} \cup \{\infty\}$
- Будем выполнять некоторые операции
- Если произойдёт что-то странное (пытаемся попасть в бесконечно удалённую точку относительно нетривиального делителя n, но не относительно n), то только во время деления на некоторое a
- $(a, n) \notin \{1, n\}$
- Можно сразу прекращать и возвращать делитель

Эллиптические кривые

Метод Ленстры. Начало

- Проверить, не является ли n точной степенью
- Выбрать планку B_1
- Выбрать случайную точку $(x, y) \in \{(x, y) \in \mathbb{Z}_n \times \mathbb{Z}_n\}$
- Выбрать случайный параметр $a \in \mathbb{Z}_n$
- Получилась случайная эллиптическая кривая: $b = y^2 - x^3 - ax \mod n$

Эллиптические кривые

Валидация кривой

- $g = \left(4a^3 + 27b^2, n\right)$
- Если $g = n$, то вернуться на предыдущий слайд
- Если $g \in \{2, \ldots, n-1\}$, нашли нетривиальный делитель
- Если $g = 1$, зафиксируем кривую $E = E_{a,b}(Z_n)$ и точку $P = (x, y)$

Эллиптические кривые

Попытка сломать эллиптическую арифметику

- Перебираем $p \in \mathbb{P}$, $p \leqslant B_1$
  - Пусть α — наибольшая степень, что $p^\alpha \leqslant B_1$
  - Заменим P на $[p^\alpha]\,P$ (α раз умножим P на p)
  - Возвращаем нетривиальный делитель, если какое-то эллиптическое сложение не удалось
- Если ничего не нашлось, ищем другую случайную кривую (возможно, с увеличением планки)

Эллиптические кривые

Оценка сложности

- Пусть $p < q$ — два наименьших простых делителя n
- Хотим: $[k]P = \infty$ в $E_{a,b}(\mathbb{Z}_p)$, но $[k]P \neq \infty$ в $E_{a,b}(\mathbb{Z}_q)$
- Второе событие гораздо вероятнее, чем первое
- Не будем обращать внимание на второе событие

Эллиптические кривые

Гладкость

- $x \in \mathbb{N}$ — y-гладкое, если все простые в x не превосходят y
- $\psi(x, y)$ — число y-гладких чисел в $\{1, \ldots, x\}$
- $\#E_{a,b}(\mathbb{Z}_p)$ — порядок эллиптической группы
- Если порядок B_1-гладкий, разумно ожидать, что случайный элемент обнулится от умножения на $\mathrm{lcm}(1, 2, \ldots, B_1)$

- Теорема Хассе:
  $\#E_{a,b}(\mathbb{Z}_p) \in \left(p+1-2\sqrt{p}; p+1+2\sqrt{p}\right)$
- Количество B_1-гладких чисел в промежутке:
  $\psi\left(p+1+2\sqrt{p}, B_1\right) - \psi\left(p+1-2\sqrt{p}, B_1\right)$
- Плотность B_1-гладких чисел в промежутке:
  $\frac{\psi\left(p+1+2\sqrt{p}, B_1\right) - \psi\left(p+1-2\sqrt{p}, B_1\right)}{4\sqrt{p}}$
- Теорема Ленстры: вероятность $\mathbb{P}(B_1)$, что $\#E_{a,b}(\mathbb{Z}_p)$ B_1-гладкое, не меньше
  $c\frac{\psi\left(p+1+2\sqrt{p}, B_1\right) - \psi\left(p+1-2\sqrt{p}, B_1\right)}{\sqrt{p}\ln p}$

- Одна попытка тратит $\mathcal{O}(B_1)$ арифметических операций и срабатывает с вероятностью $\mathbb{P}(B_1) \approx c\frac{\psi\left(p+1+2\sqrt{p},B_1\right)-\psi\left(p+1-2\sqrt{p},B_1\right)}{\sqrt{p}\ln p}$
- Минимизируем $\frac{B_1}{\mathbb{P}(B_1)}$
- Минимальное число операций достигается при $B_1 = \exp\left(\left(\frac{1}{\sqrt{2}}+o(1)\right)\sqrt{\ln p \ln\ln p}\right)$
- Соответствующая асимптотика $\exp\left(\left(\sqrt{2}+o(1)\right)\sqrt{\ln p \ln\ln p}\right)$
- $L_p\left[\frac{1}{2},\sqrt{2}\right]$

Спасибо!