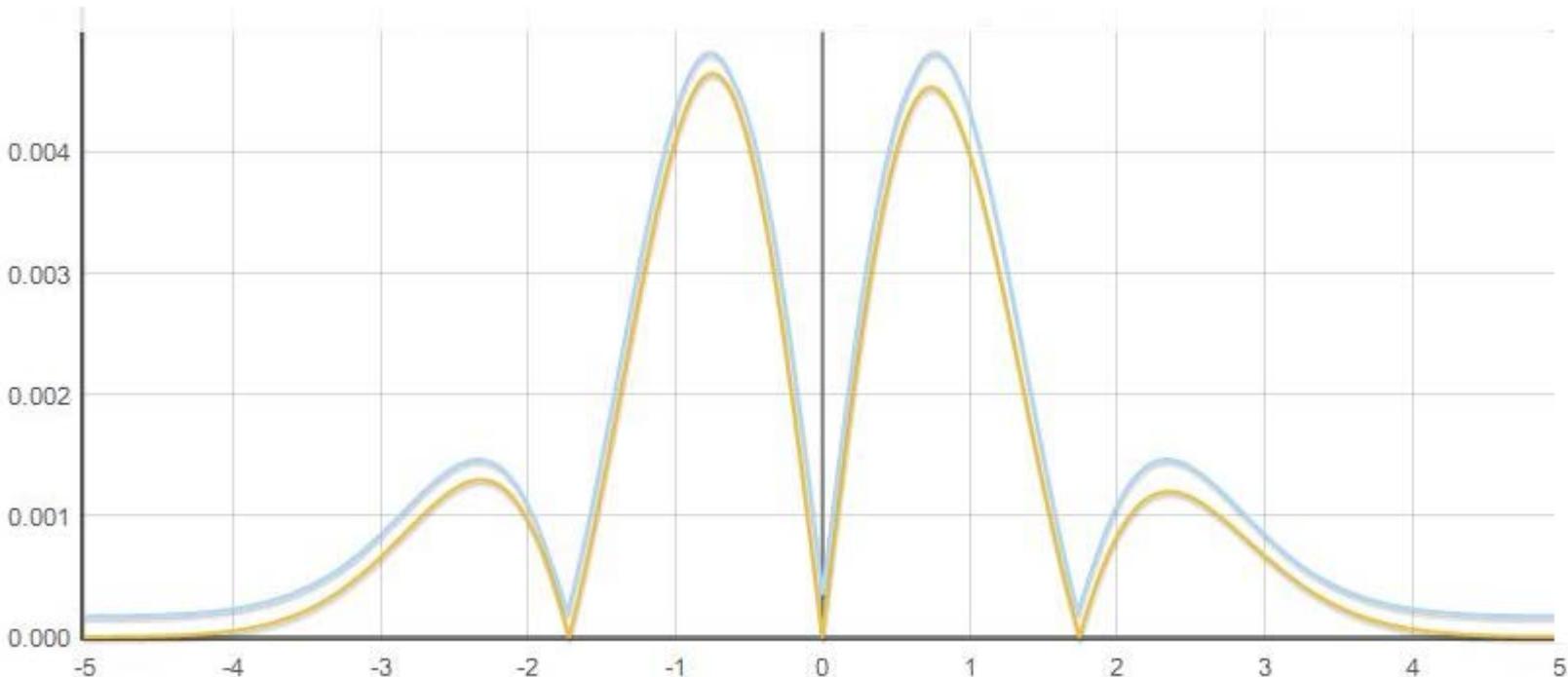


# О новом типе оценок точности аппроксимации в ЦПТ

В.В.Сенатов



Пусть  $X_1, X_2, \dots$  – независимые одинаково распределённые случайные величины с нулевым средним, единичной дисперсией и общим распределением  $P$ . Обозначим  $P_n$  распределение нормированной суммы  $\frac{(X_1 + \dots + X_n)}{\sqrt{n}}$  и  $\Phi$  – нормальный закон с плотностью  $\varphi(x) = e^{-x^2/2}/\sqrt{2\pi}$ ,  $F_n(x)$  и  $G(x)$  – функции распределения мер  $P_n$  и  $\Phi$ . ЦПТ утверждает, что при  $n \rightarrow \infty$

$$F_n(x) \rightarrow G(x) \text{ равномерно по } -\infty < x < \infty,$$

т. е.

$$\rho(F_n, G) = \sup_{-\infty < x < \infty} |F_n(x) - G(x)| \rightarrow 0,$$

поэтому для оценки точности нормальной аппроксимации можно оценивать  $\rho(F_n, G)$ .

**Можно не означает нужно.**

Далее

$$\alpha_k = \int_{-\infty}^{\infty} x^k P(dx), \quad k = 3, 4, \dots, \text{ и } \beta_s = \int_{-\infty}^{\infty} |x|^s P(dx), \quad s > 2.$$

Теорема Берри – Эссеена (А. Берри, 1941, К.-Г. Эссеен, 1942)

$$\rho(F_n, G) \leq c \frac{\beta_3}{\sqrt{n}},$$

где  $c > 0$  – постоянная.

К.-Г. Эссеен утверждал, что  $c \leq 7.59$ , А. Берри утверждал, что  $c \leq 1.88$ , но его вычисления содержали ошибку, ..... В 1956 г. К.-Г. Эссеен получил нижнюю оценку

$$c \geq c_E = \frac{3 + \sqrt{10}}{6\sqrt{2\pi}} = 0.4097 \dots .$$

Нижняя оценка доказывается с привлечением распределения Эссеена, которое является частным случаем распределения Бернулли. Распределения Бернулли с нулевым средним и единичной дисперсией образуют однопараметрическое семейство, для них

$$P\left(\left\{-\frac{p}{\sqrt{pq}}\right\}\right) = q, \quad P\left(\left\{\frac{q}{\sqrt{pq}}\right\}\right) = p, \quad q = 1 - p, \quad 0 < p < 1.$$

Это распределение является решетчатым с шагом  $h = \frac{1}{\sqrt{pq}} \geq 2$ . Распределение нормированной суммы  $\frac{X_1 + \dots + X_n}{\sqrt{n}}$  независимых случайных величин с этим распределением также является решетчатым с шагом  $h/\sqrt{n}$ , оно сосредоточено на решетке

$$D_n = \left\{ -p \sqrt{\frac{n}{pq}} + k \frac{h}{\sqrt{n}}, k = 0, \pm 1, \pm 2, \dots \right\}.$$

Для распределения Эссеена  $p = \frac{4-\sqrt{10}}{2}$  и для него

$$\rho(F_n, G) \sim c_E \frac{\beta_3}{\sqrt{n}}.$$

В 2001 г. Г. П. Чистяков получил оценку

$$\rho(F_n, G) \leq c_E \frac{\beta_3}{\sqrt{n}} + C \left( \frac{\beta_3}{\sqrt{n}} \right)^{40/39} \left| \ln \frac{\beta_3}{\sqrt{n}} \right|^{7/6},$$

где  $C$  – некоторая постоянная, а в 2012 г. И. Г. Шевцова получила оценку

$$\rho(F_n, G) \leq c_E \frac{\beta_3}{\sqrt{n}} + 2.58 \left( \frac{\beta_3}{\sqrt{n}} \right)^2.$$

Последний результат делает дальнейшие попытки уточнения оценки теоремы Берри – Эссеена бессмысленными.

В 1937 г. были получены два следующих результата. Г. Крамер установил, что при выполнении условия  $\overline{\lim}_{t \rightarrow \infty} |f(t)| < 1$ , где  $f(t)$  – характеристическая функция распределения  $P$ , справедливо равенство

$$F_n(x) - G(x) = \frac{1}{6\sqrt{2\pi}} \frac{\alpha_3}{\sqrt{n}} (1 - x^2) e^{-x^2/2} + o\left(\frac{1}{\sqrt{n}}\right), \quad -\infty < x < \infty, \quad n \rightarrow \infty.$$

Я. В. Успенский получил результат, из которого следует, что если  $P$  – распределение Бернулли, то при  $npq \geq 25$

$$F_n(x) - G(x) = \frac{1}{6\sqrt{2\pi}} \frac{\alpha_3}{\sqrt{n}} (1 - x^2) e^{-x^2/2} + O\left(\frac{1}{n}\right), \quad x \in D_n + \frac{h}{2\sqrt{n}},$$

$$\left| O\left(\frac{1}{n}\right) \right| \leq \frac{0.13 + 0.18|p - q|}{npq} + e^{-3\sqrt{npq}/2}.$$

Для распределения Эссеена из результата Я. В. Успенского следует, что

$$\max \left\{ |F_n(x) - G(x)| : x \in D_n + \frac{h}{2\sqrt{n}} \right\} \leq \frac{0.022}{\sqrt{n}} + \frac{0.655}{n} + e^{-0.74\sqrt{n}}.$$

Это неравенство можно записать в виде

$$\rho(F_n, G_n^{a,h}) \leq \frac{0.022}{\sqrt{n}} + \frac{0.655}{n} + e^{-0.74\sqrt{n}},$$

где  $G_n^{a,h}(x)$  – решетчатая функция распределения, постоянная на интервалах постоянства  $F_n(x)$ , и совпадающая с  $G(x)$  в серединах этих интервалов.

В то же время, для распределения Эссеена  $\beta_3 = 1.04 \dots$  и правая часть оценки Берри – Эссеена для распределения Эссеена больше

$$c_E \frac{\beta_3}{\sqrt{n}} > 0.4097 \frac{1.04}{\sqrt{n}} = \frac{0.426}{\sqrt{n}}.$$

Для распределения Бернулли  $\alpha_3 = (q - p)/\sqrt{pq}$  и мы можем считать, что  $0 < p \leq 0.5$ . Рассмотрим интервал  $(x'_n, x''_n)$  постоянства  $F_n$ , середина которого  $x_n^* = (x'_n + x''_n)/2$  ближе всего к нулю. Из результата Я. В. Успенского следует, что

$$F_n(x_n^*) - G(x_n^*) = \frac{q - p}{6\sqrt{2\pi prq}} + o\left(\frac{1}{\sqrt{n}}\right) > 0.$$

Очевидно, что

$$\begin{aligned} F_n(x'_n + 0) - G(x'_n + 0) &= \frac{q - p}{6\sqrt{npq}} \frac{1}{\sqrt{2\pi}} + \frac{h}{2\sqrt{n}} \frac{1}{\sqrt{2\pi}} + o\left(\frac{1}{\sqrt{n}}\right) = \\ &= \frac{q - p + 3}{6\sqrt{2\pi prq}} + o\left(\frac{1}{\sqrt{n}}\right) \leq c \frac{\beta_3}{\sqrt{n}}, \end{aligned}$$

где  $c$  – постоянная из оценки Берри – Эссеена.

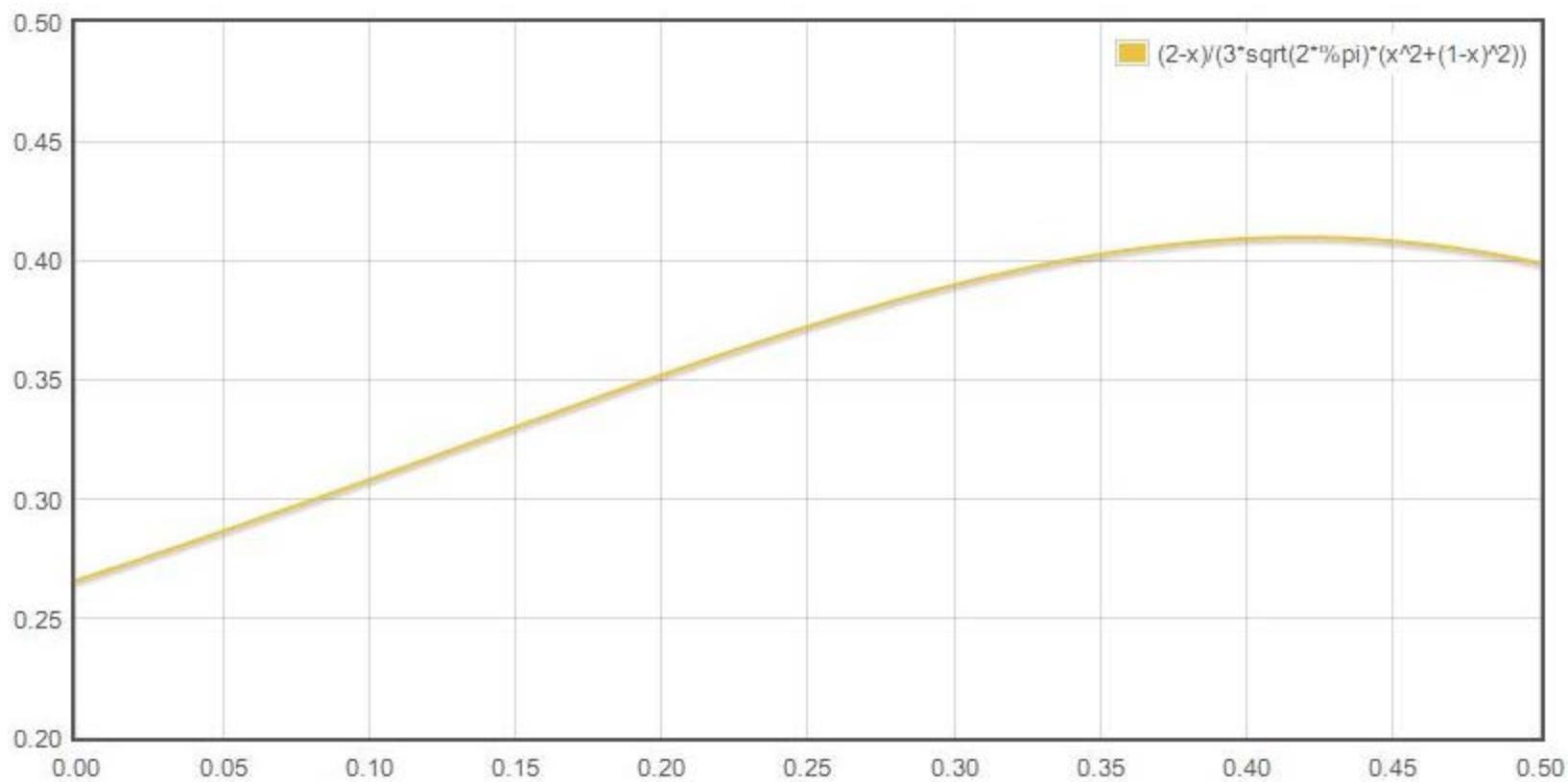
Отсюда следует, что

$$\frac{q - p + 3}{6\beta_3 \sqrt{2\pi pq}} = c_p \leq C.$$

Для распределения Бернулли  $\beta_3 = \frac{p^2 + q^2}{\sqrt{pq}}$ , откуда следует, что

$$c_p = \frac{q - p + 3}{6\sqrt{2\pi} (p^2 + q^2)} = \frac{2 - p}{3\sqrt{2\pi} (p^2 + (1 - p)^2)}.$$

Это – нижняя оценка постоянной  $C$  в теореме Берри – Эссеена для распределения Бернулли с параметром  $p$ . Её максимум достигается при  $p = \frac{4 - \sqrt{10}}{2}$ .



Вернёмся к равенству

$$\begin{aligned} F_n(x'_n + 0) - G(x'_n + 0) &= \frac{q-p}{6\sqrt{npq}} \frac{1}{\sqrt{2\pi}} + \frac{h}{2\sqrt{n}} \frac{1}{\sqrt{2\pi}} + o\left(\frac{1}{\sqrt{n}}\right) = \\ &= \frac{1}{6\sqrt{2\pi}} \frac{|\alpha_3|}{\sqrt{n}} + \frac{h}{2\sqrt{n}} \frac{1}{\sqrt{2\pi}} + o\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Из него видно, как формируется оценка  $\rho(F_n, G)$  для распределений Бернулли: оценка равномерного расстояния есть сумма оценки

$$\max \left\{ |F_n(x) - G(x)| : x \in D_n + \frac{h}{2\sqrt{n}} \right\}$$

и половины максимального скачка функции  $F_n(x)$ . Для распределения Эссеена вторая величина в 18.47 раз больше первой.

Результаты Г. Крамера и Я. В. Успенского были улучшены К.-Г. Эссееном в 1945 г., который для результата Г. Крамера ослабил условие гладкости до его естественной границы, до условия нерешетчатости распределения  $P$ . Он показал также, что результат Я. В. Успенского остаётся справедливым для всех действительных  $x$ , если в правую часть равенства добавить слагаемое, которое компенсирует возрастание нормальной функции распределения  $G(x)$  на интервалах постоянства функции  $F_n(x)$ . Явных оценок остаточных частей разложений у К.-Г. Эссеена не было.

«... задача выделения достаточно широкого класса распределений, для которых гарантированная точность нормальной аппроксимации и ее уточнений соответствует фактически наблюдаемой, все еще остается нерешенной.»

Ю. В. Прохоров, Ю. А. Розанов, Теория вероятностей. Основные понятия. Предельные теоремы. Случайные процессы., М., 1967, стр. 232.

Далее нас интересуют оценки в локальной форме ЦПТ для плотностей. Мы будем использовать условие гладкости распределения  $P$

$$\int\limits_{-\infty}^{\infty} |f(t)|^v dt < \infty,$$

где  $f(t)$  – характеристическая функция распределения  $P$ ,  $v$  – некоторое положительное число. Выполнение этого условия гарантирует существование непрерывных и ограниченных плотностей  $p_n(x) = F'_n(x)$  для  $n \geq v$ , и справедливость соотношения

$$p_n(x) \rightarrow \varphi(x), n \rightarrow \infty, -\infty < x < \infty.$$

Нам понадобятся многочлены Чебышева – Эрмита  $H_k(x) = (-1)^k \varphi^{(k)}(x)/\varphi(x)$ ,  $k = 0, 1, \dots$ , в частности,  $H_0(x) \equiv 1, H_1(x) = x, H_2(x) = x^2 - 1, H_3(x) = x^3 - 3x, H_4(x) = x^4 - 6x^2 + 3, H_5(x) = x^5 - 10x^3 + 15x, H_6(x) = x^6 - 15x^4 + 45x^2 - 15$ .

Для распределений  $P$  с конечным шестым моментом

$$p_n(x) - \varphi(x) = \frac{A_1(x)}{\sqrt{n}} \varphi(x) + \frac{A_2(x)}{n} \varphi(x) + \frac{A_3(x)}{n^{3/2}} \varphi(x) + O\left(\frac{1}{n^2}\right), n \rightarrow \infty,$$

где

$$A_1(x) = \frac{\theta_3}{3!} H_3(x), \quad A_2(x) = \frac{\theta_4}{4!} H_4(x) + \frac{1}{2} \left( \frac{\theta_3}{3!} \right)^2 H_6(x),$$

$$A_3(x) = \frac{\theta_5}{5!} H_5(x) + \frac{\theta_3 \theta_4}{3! 4!} H_7(x) + \frac{1}{6} \left( \frac{\theta_3}{3!} \right)^3 H_9(x).$$

Здесь  $\theta_l = \int_{-\infty}^{\infty} H_l(x) P(dx)$  – числа, которые мы называем моментами Чебышева – Эрмита распределения  $P$ , в частности,  $\theta_3 = \alpha_3, \theta_4 = \alpha_4 - 3, \theta_5 = \alpha_5 - 10\alpha_3$ .

Отсюда следует, что

$$p_n(x) - \varphi(x) = \frac{A_1(x)}{\sqrt{n}} \varphi(x) + O\left(\frac{1}{n}\right), n \rightarrow \infty, \quad (1)$$

$$p_n(x) - \varphi(x) - \frac{A_1(x)}{\sqrt{n}} \varphi(x) = \frac{A_2(x)}{n} \varphi(x) + O\left(\frac{1}{n^{3/2}}\right), n \rightarrow \infty, \quad (2)$$

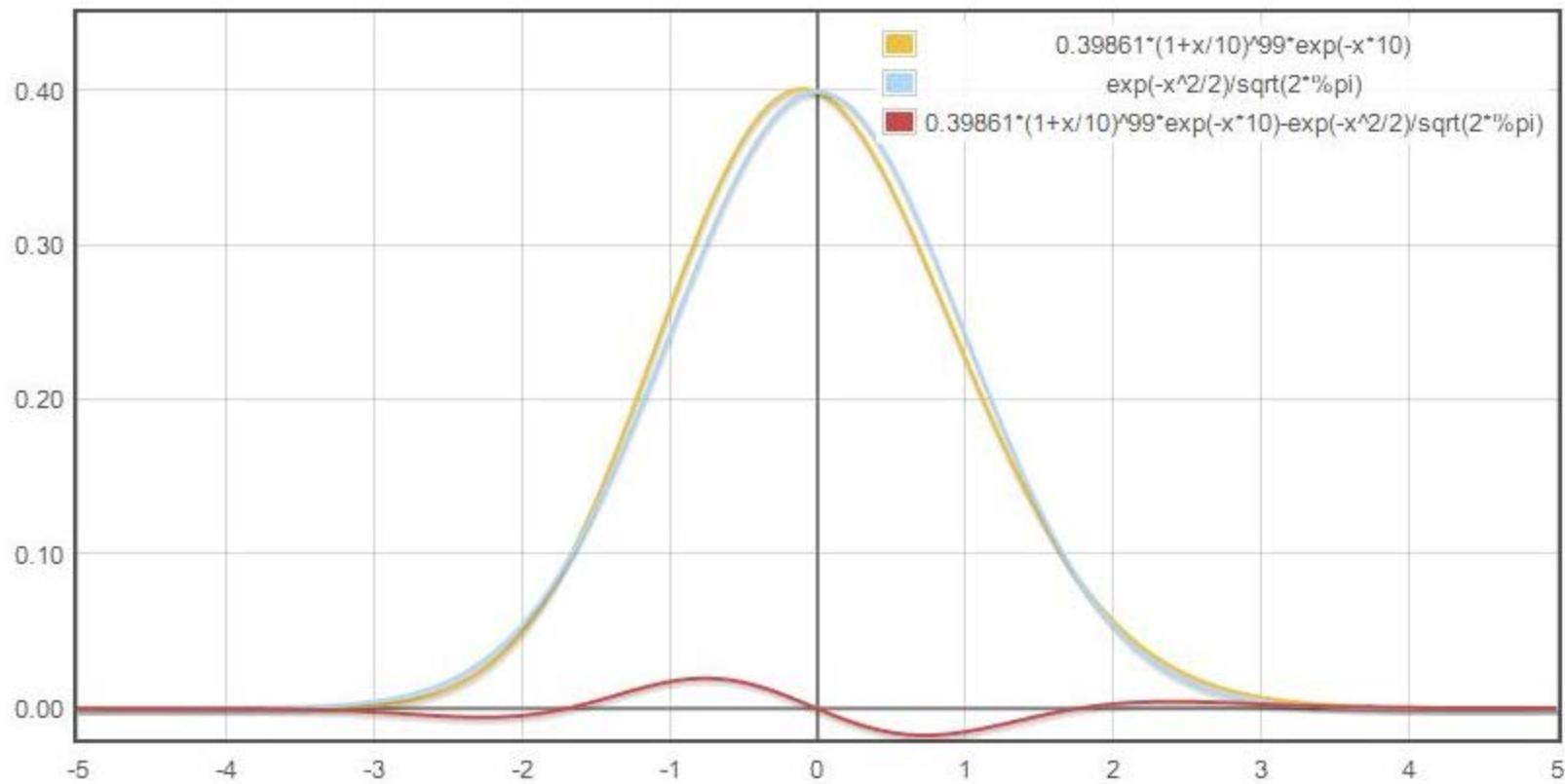
$$p_n(x) - \varphi(x) - \left( \frac{A_1(x)}{\sqrt{n}} \varphi(x) + \frac{A_2(x)}{n} \varphi(x) \right) = \frac{A_3(x)}{n^{3/2}} \varphi(x) + O\left(\frac{1}{n^2}\right), n \rightarrow \infty. \quad (3)$$

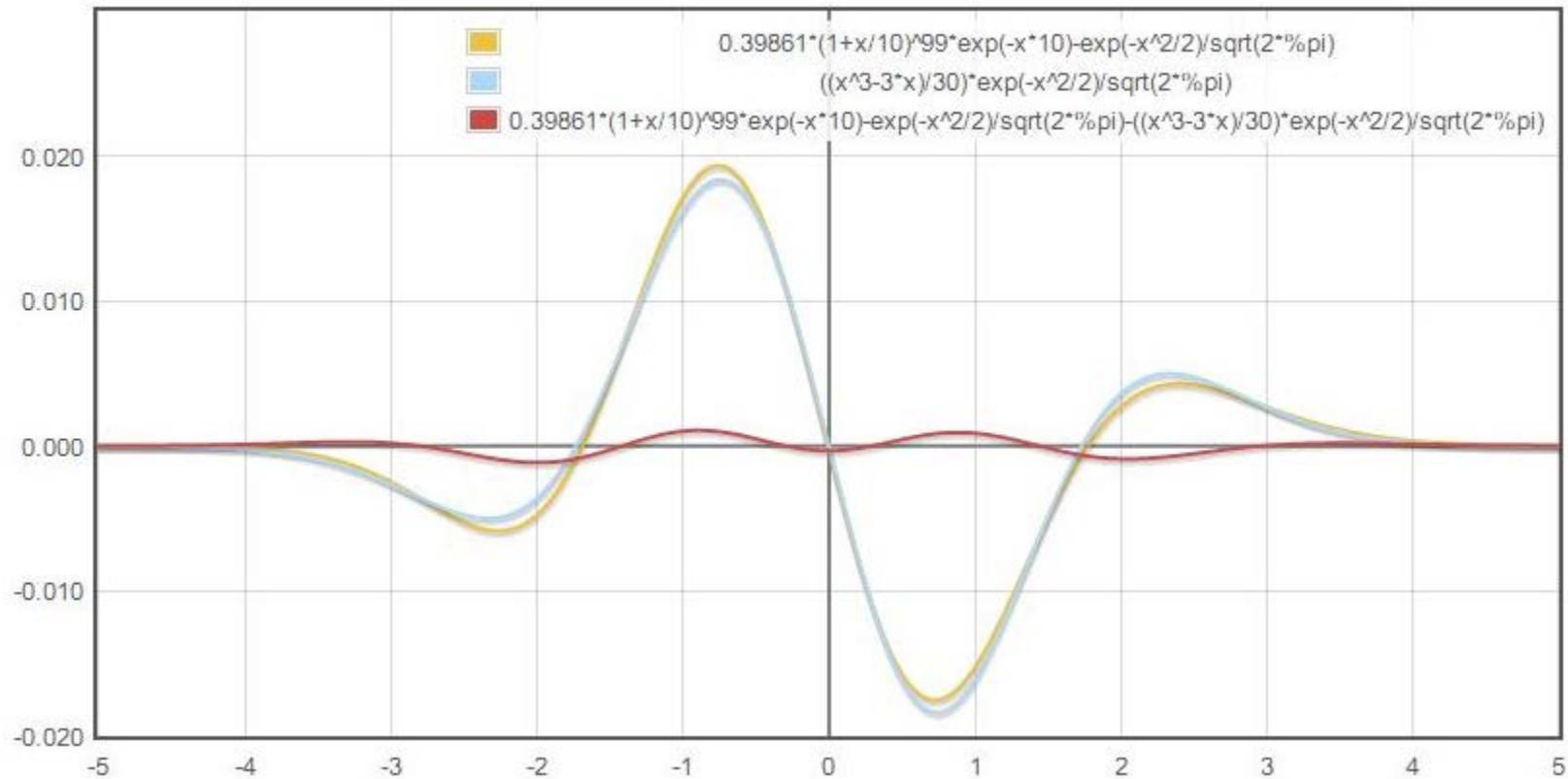
Для иллюстраций мы будем использовать распределение  $P$ , которое является центрированным экспоненциальным распределением с параметром 1, т. е. распределением с плотностью

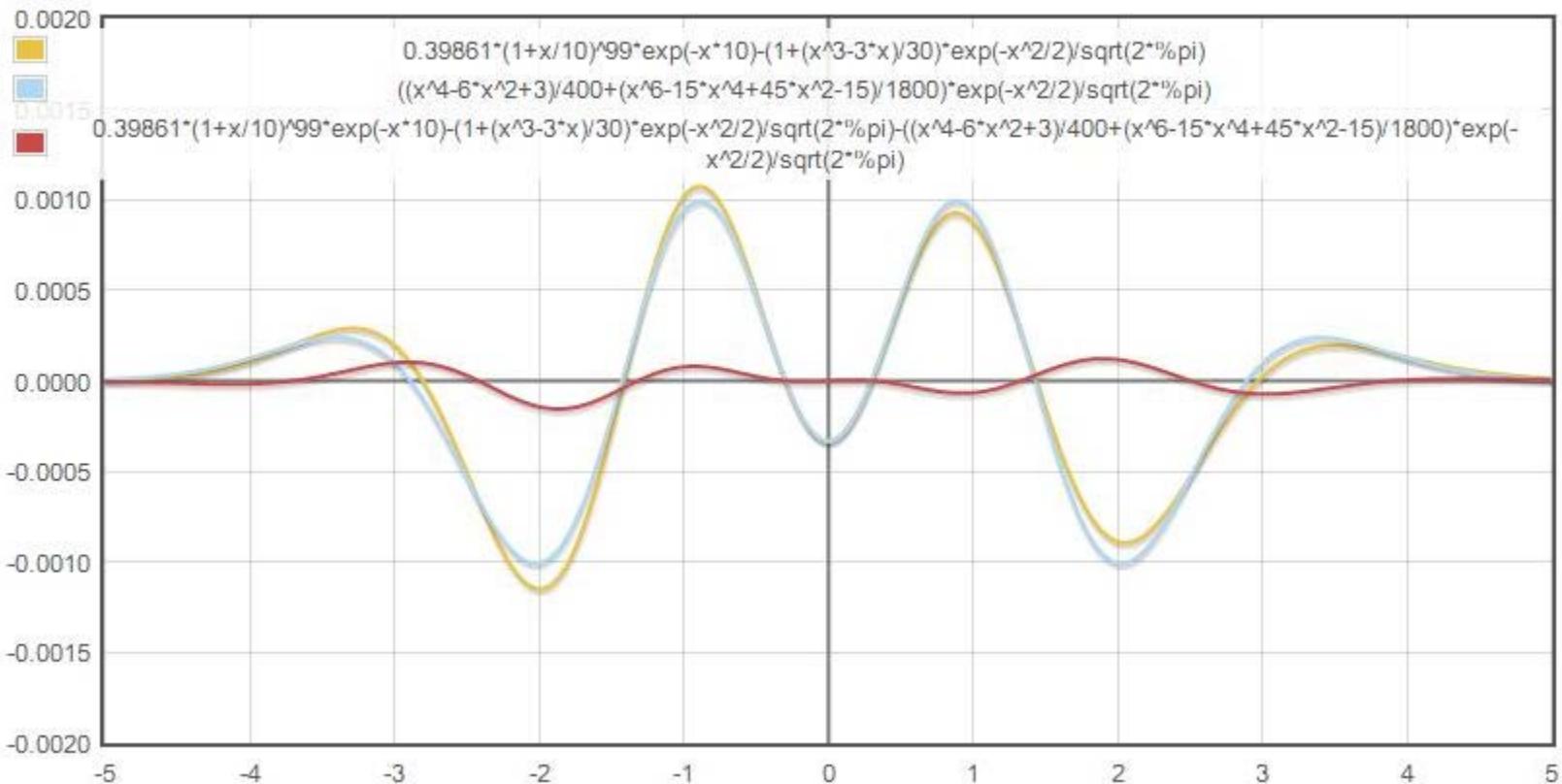
$$p(x) = 0, x < -1, \text{ и } p(x) = e^{-(x+1)}, x \geq -1.$$

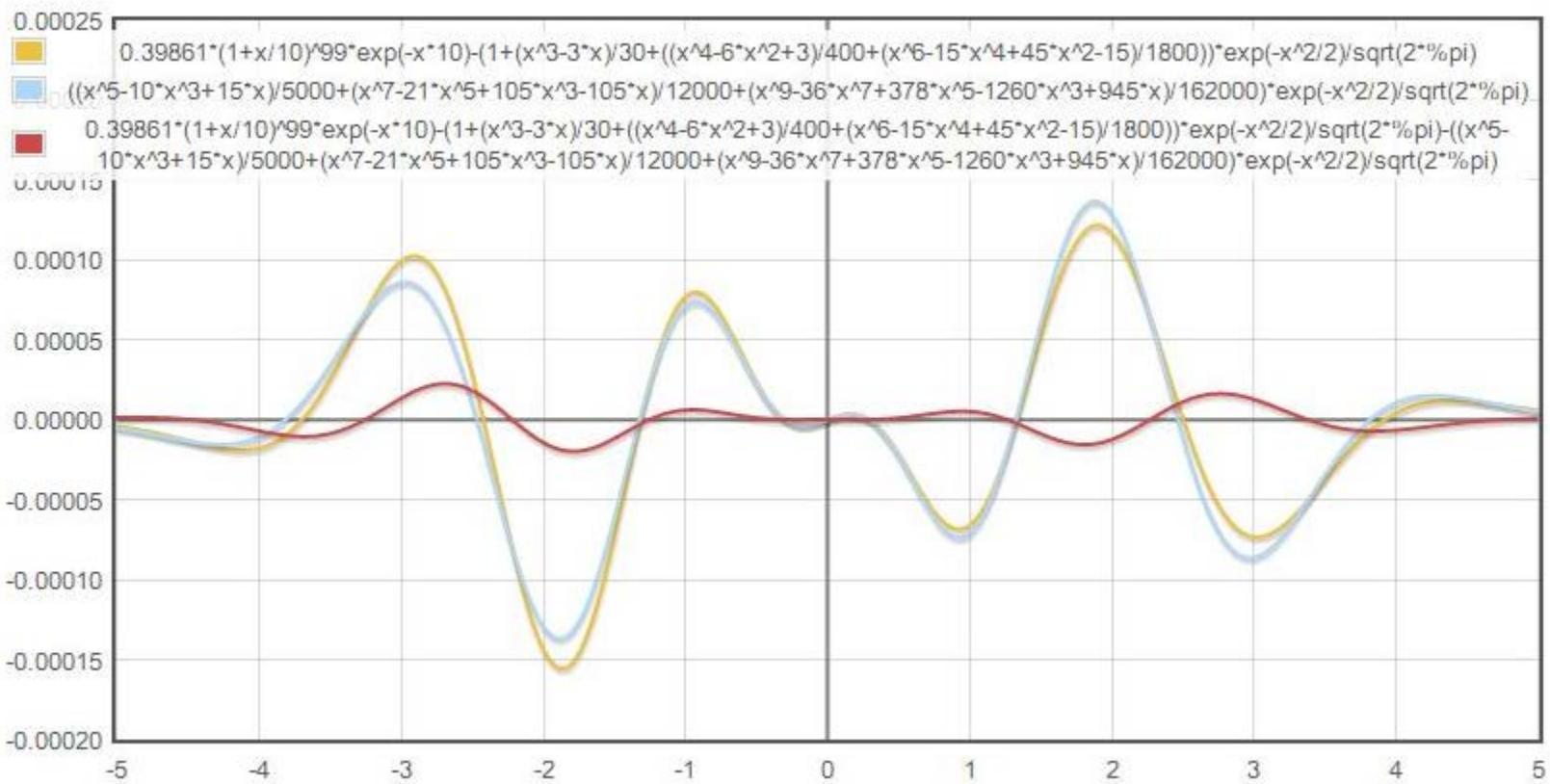
Оно имеет нулевое среднее, единичную дисперсию, для него  $\alpha_3 = 2, \alpha_4 = 9, \alpha_5 = 44, \alpha_6 = 265$ . Его моменты  $\alpha_k \sim k!/e$  очень быстро возрастают при росте  $k$ , кроме того оно очень асимметрично в том смысле, что  $\alpha_k \sim \beta_k$  для нечётных  $k$ . Для этого распределения

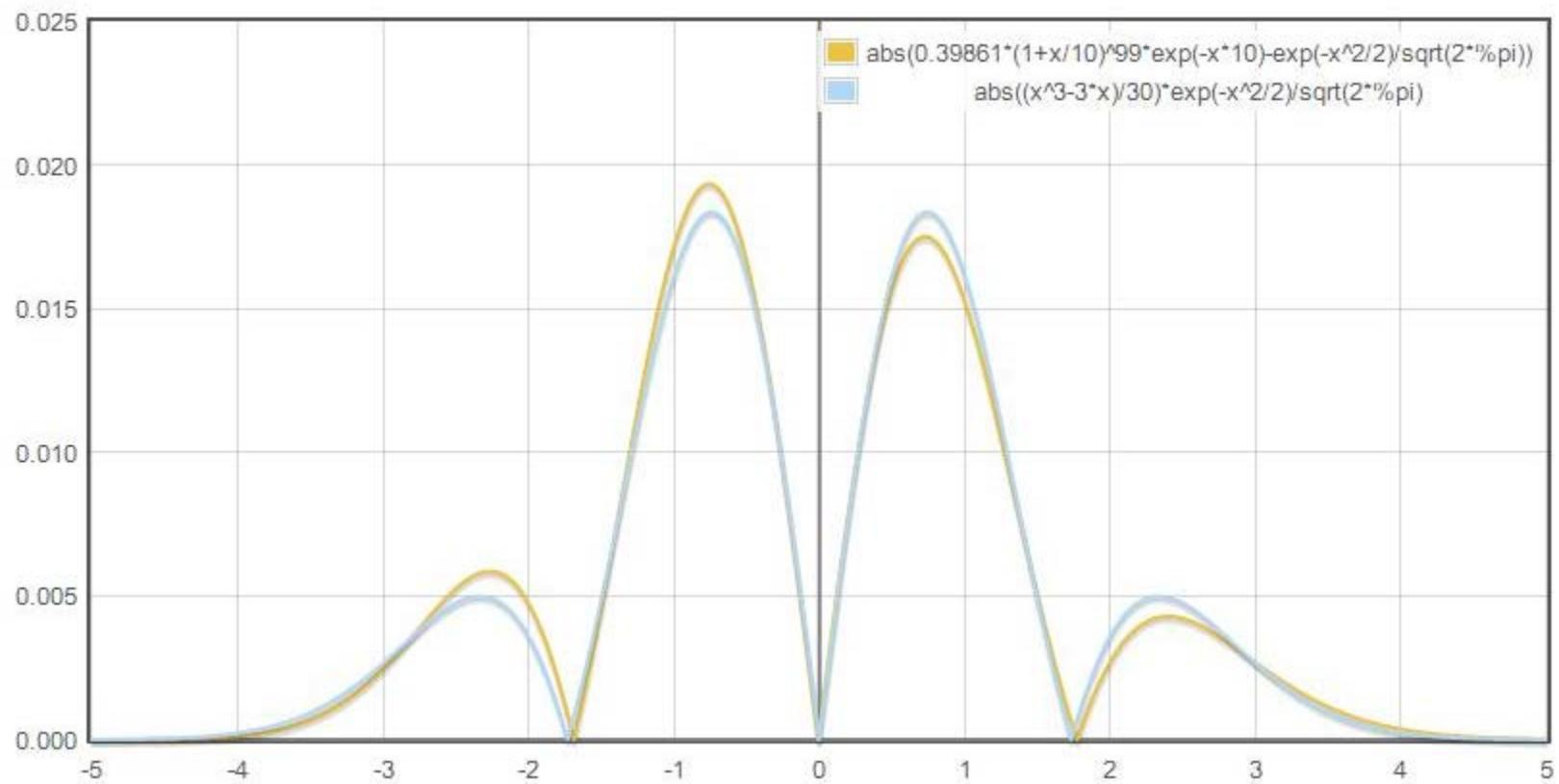
$$p_n(x) = \sqrt{n} \frac{n^n}{n! e^n} \left(1 + \frac{x}{\sqrt{n}}\right)^{n-1} e^{-x\sqrt{n}}, \quad x \geq -\sqrt{n}.$$

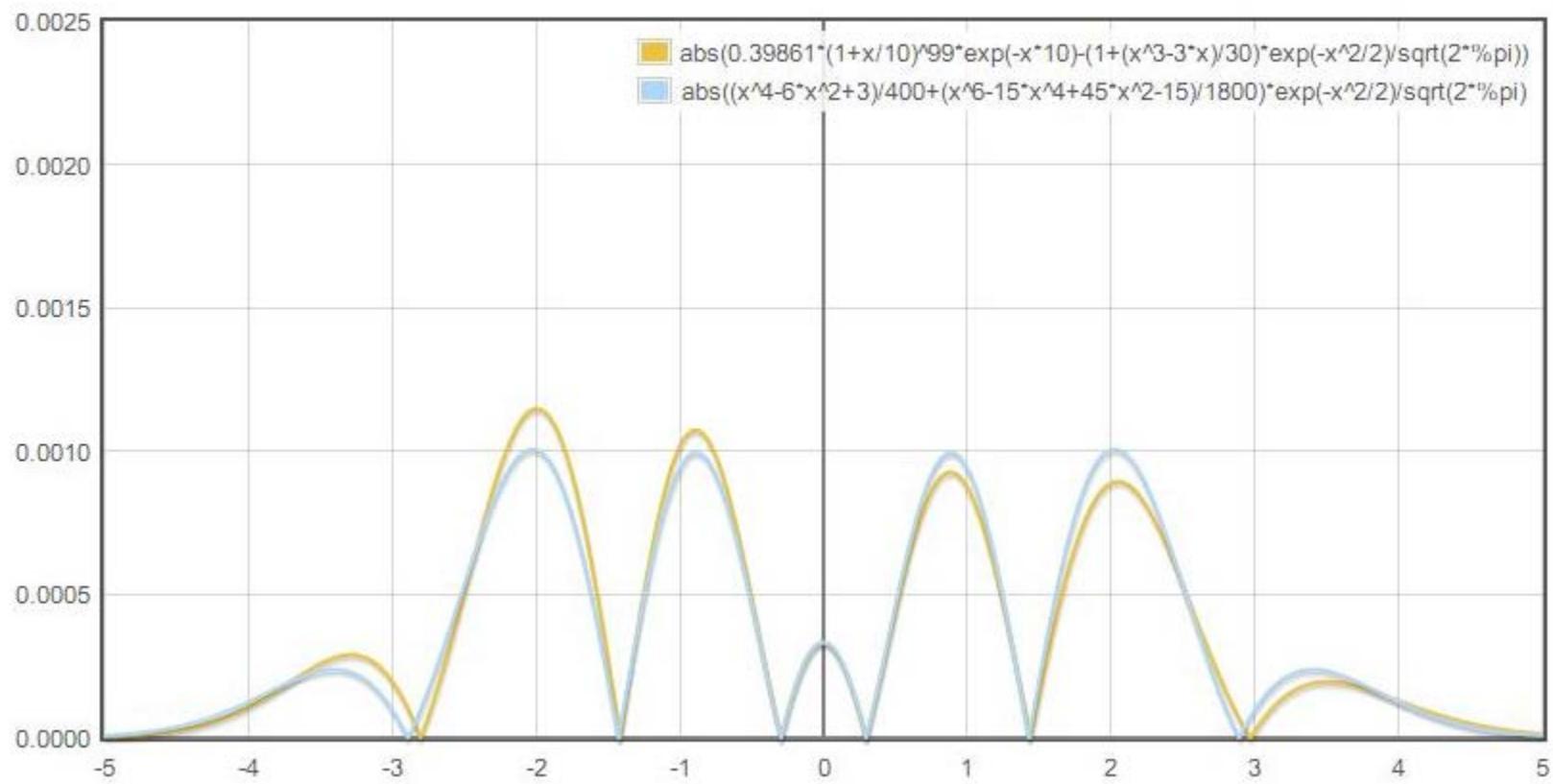


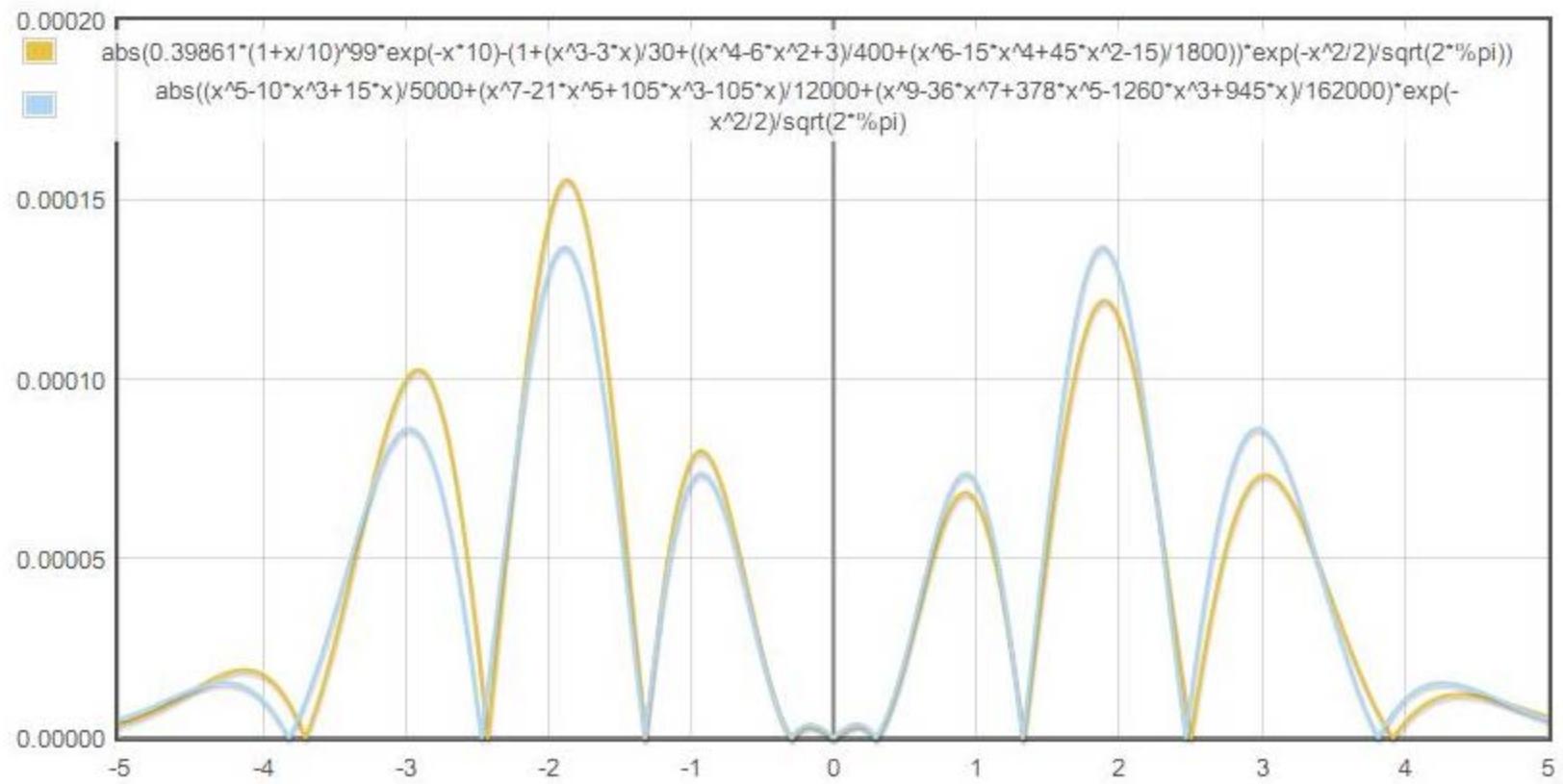












Справедливы оценки: для распределений  $P$  с конечным четвёртым моментом

$$|p_n(x) - \varphi(x)| \leq \frac{|A_1(x)|}{\sqrt{n}} \varphi(x) + O_1\left(\frac{1}{n}\right), n \rightarrow \infty,$$

для распределений  $P$  с конечным пятым моментом

$$\left| p_n(x) - \varphi(x) - \frac{A_1(x)}{\sqrt{n}} \varphi(x) \right| \leq \frac{|A_2(x)|}{n} \varphi(x) + O_2\left(\frac{1}{n^{3/2}}\right), n \rightarrow \infty,$$

для распределений  $P$  с конечным шестым моментом

$$\left| p_n(x) - \varphi(x) - \left( \frac{A_1(x)}{\sqrt{n}} \varphi(x) + \frac{A_2(x)}{n} \varphi(x) \right) \right| \leq \frac{|A_3(x)|}{n^{3/2}} \varphi(x) + O_3\left(\frac{1}{n^2}\right), n \rightarrow \infty.$$

$$O_1\left(\frac{1}{n}\right) \leq 0.03 \frac{\beta_4}{n} + \frac{|A_2^*(x)|\varphi(x)}{n} + O\left(\frac{1}{n^{3/2}}\right), \quad n \rightarrow \infty,$$

$$O_2\left(\frac{1}{n^{3/2}}\right) \leq 0.0124 \frac{\beta_5}{n^{3/2}} + \frac{|A_3^*(x)|\varphi(x)}{n^{3/2}} + O\left(\frac{1}{n^2}\right), \quad n \rightarrow \infty,$$

$$O_3\left(\frac{1}{n^2}\right) \leq 0.0048 \frac{\beta_6}{n^2} + \frac{|A_4^*(x)|\varphi(x)}{n^2} + O\left(\frac{1}{n^{5/2}}\right), \quad n \rightarrow \infty,$$

где

$$A_2^*(x) = H_4(x)/40 + H_6(x)/18, \quad A_3^*(x) = -H_5(x)/72 + H_7(x)/12 + H_9(x)/162,$$

$$A_4^*(x) = -0.043651H_6(x) + (1/15 + 1/32)H_8(x) + H_{10}(x)/72 + H_{12}(x)/1944.$$

Обсудим оценку

$$|p_n(x) - \varphi(x)| \leq \frac{|A_1(x)|}{\sqrt{n}} \varphi(x) + O_1\left(\frac{1}{n}\right), n \rightarrow \infty,$$

и приведём явный вид слагаемых, которые убывают в точности как  $1/n$ .

Для распределений  $P$  с конечным четвёртым моментом

$$p_n(x) - \varphi(x) = \frac{\alpha_3}{3! \sqrt{n}} H_3 \varphi(x) + \frac{\alpha_4 - 3}{4! n} H_4(x) \varphi(x) + \frac{1}{2} \left( \frac{\alpha_3}{3! \sqrt{n}} \right)^2 H_6(x) \varphi(x) + o\left(\frac{1}{n}\right).$$

Аналог этого разложения

$$p_n(x) - \varphi(x) = \frac{\alpha_3}{3! \sqrt{n}} H_3(x) \varphi(x) + \frac{\lambda \alpha_4 - 3}{4! n} H_4(x) \varphi(x) + \frac{1}{2} \left( \frac{\alpha_3}{3! \sqrt{n}} \right)^2 H_6(x) \varphi(x) + R,$$

где  $0 \leq \lambda \leq 1$  – параметр, выбор которого находится в нашем распоряжении, а

$$|R| \leq q(\lambda) \frac{1}{8\sqrt{2\pi}} \frac{\beta_4}{n} + O\left(\frac{1}{n^{3/2}}\right), \quad n \rightarrow \infty,$$

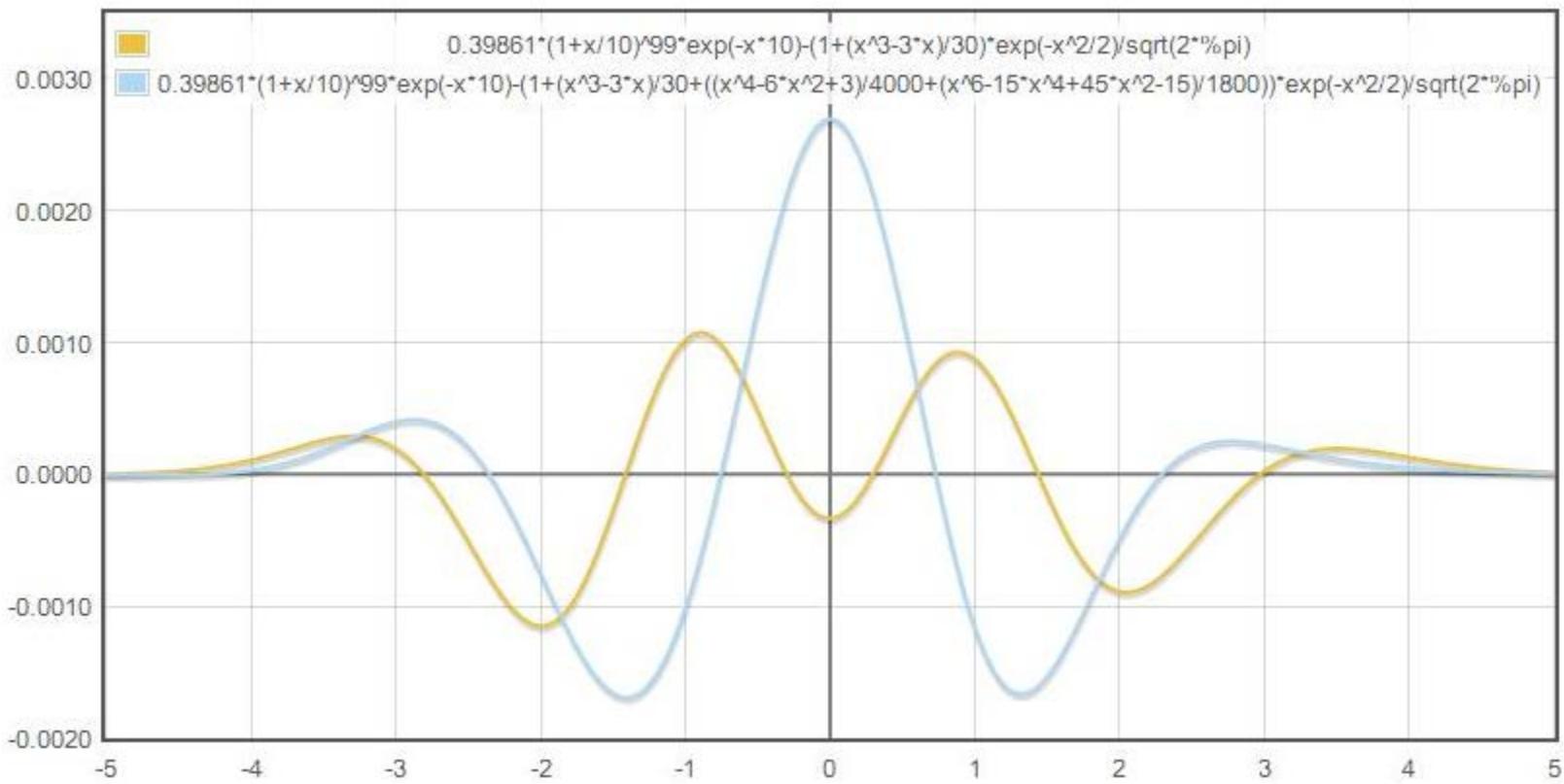
$$q(\lambda) = \sup_{u>0} \frac{\left| e^{iu} - \left( 1 + iu + \frac{(iu)^2}{2} + \frac{(iu)^3}{3!} + \lambda \frac{(iu)^4}{4!} \right) \right|}{u^4/4!}.$$

При  $0 \leq \lambda \leq 0.4$  справедливо равенство  $\lambda + q_4(\lambda) = 1$  и из последнего разложения следует, что при этих  $\lambda$

$$\left| p_n(x) - \varphi(x) - \left( \frac{\alpha_3}{3! \sqrt{n}} H_3(x) \varphi(x) + \frac{\lambda \alpha_4 - 3}{4! n} H_4(x) \varphi(x) + \frac{1}{2} \left( \frac{\alpha_3}{3! \sqrt{n}} \right)^2 H_6(x) \varphi(x) \right) \right| \leq$$

$$\leq \frac{1 - \lambda}{8\sqrt{2\pi}} \frac{\beta_4}{n} + O\left(\frac{1}{n^{3/2}}\right).$$

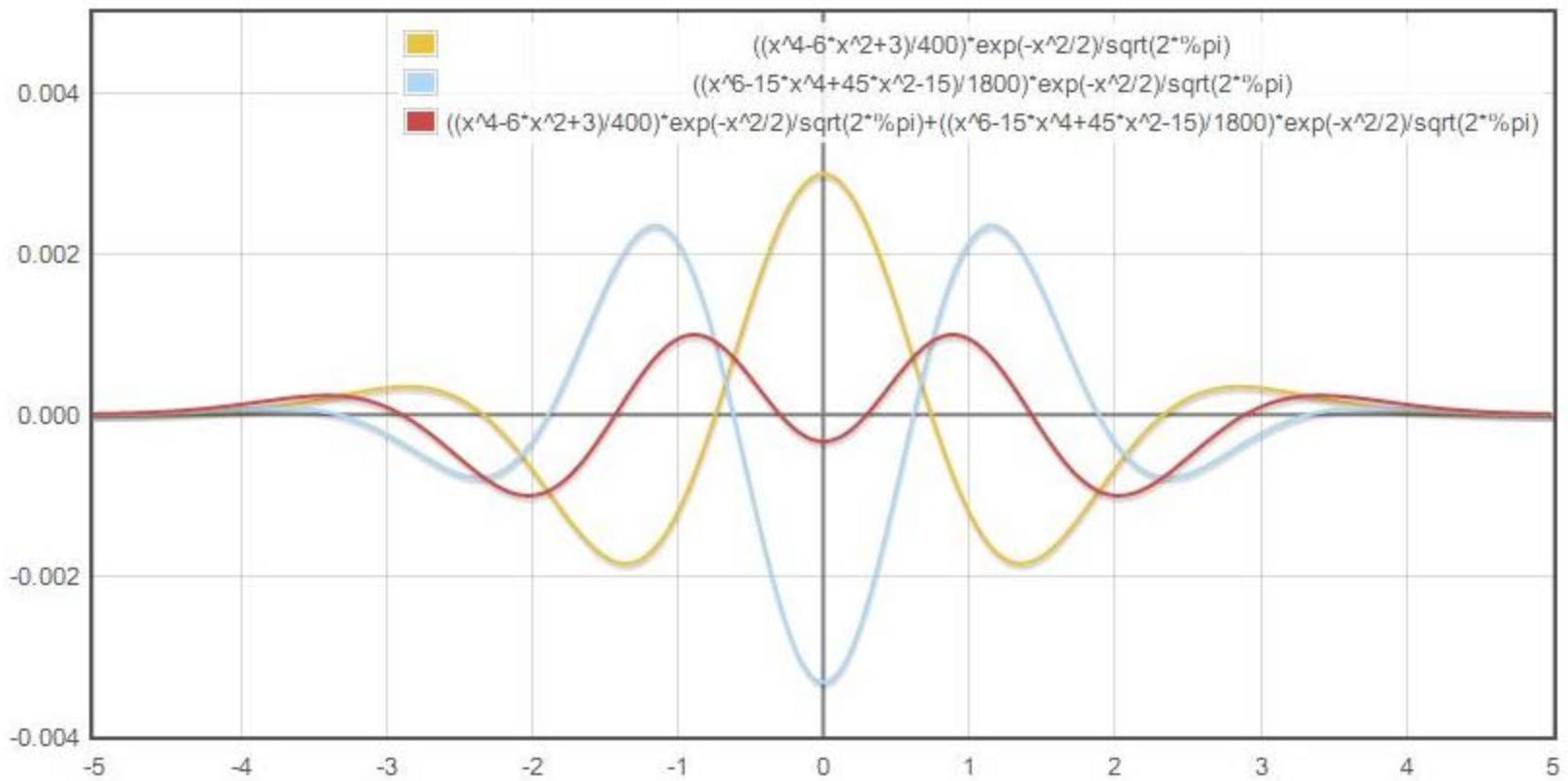
Для любого распределения  $P$  с конечным четвёртым моментом правая часть этой оценки эквивалентна левой части при  $x = 0$ . Возьмём  $\lambda = 0.4$ . Оказывается, что сумма трёх слагаемых под знаком модуля аппроксимирует  $p_n(x) - \varphi(x)$  хуже, чем первое слагаемое.



«Нерегулярность» асимптотических разложений связана с тем, что все члены разложений, начиная со второго, являются суммами нескольких слагаемых. Добавляя к предыдущему разложению новую сумму «целиком», мы (как правило) улучшаем точность аппроксимации. Добавляя к предыдущему разложению слагаемые из следующего члена разложения поодиночке, мы можем столкнуться с ухудшением точности аппроксимации. Рассмотрим сумму

$$\frac{\alpha_4 - 3}{4!n} H_4(x)\varphi(x) + \frac{1}{2} \left( \frac{\alpha_3}{3!\sqrt{n}} \right)^2 H_6(x)\varphi(x).$$

Для ЭР при  $n = 100$  и  $x = 0$  первое слагаемое равно  $2.992 \cdot 10^{-3}$ , второе равно  $-3.3245 \cdot 10^{-3}$ , их сумма равна  $-3.325 \cdot 10^{-4}$ .



Перепишем разложение

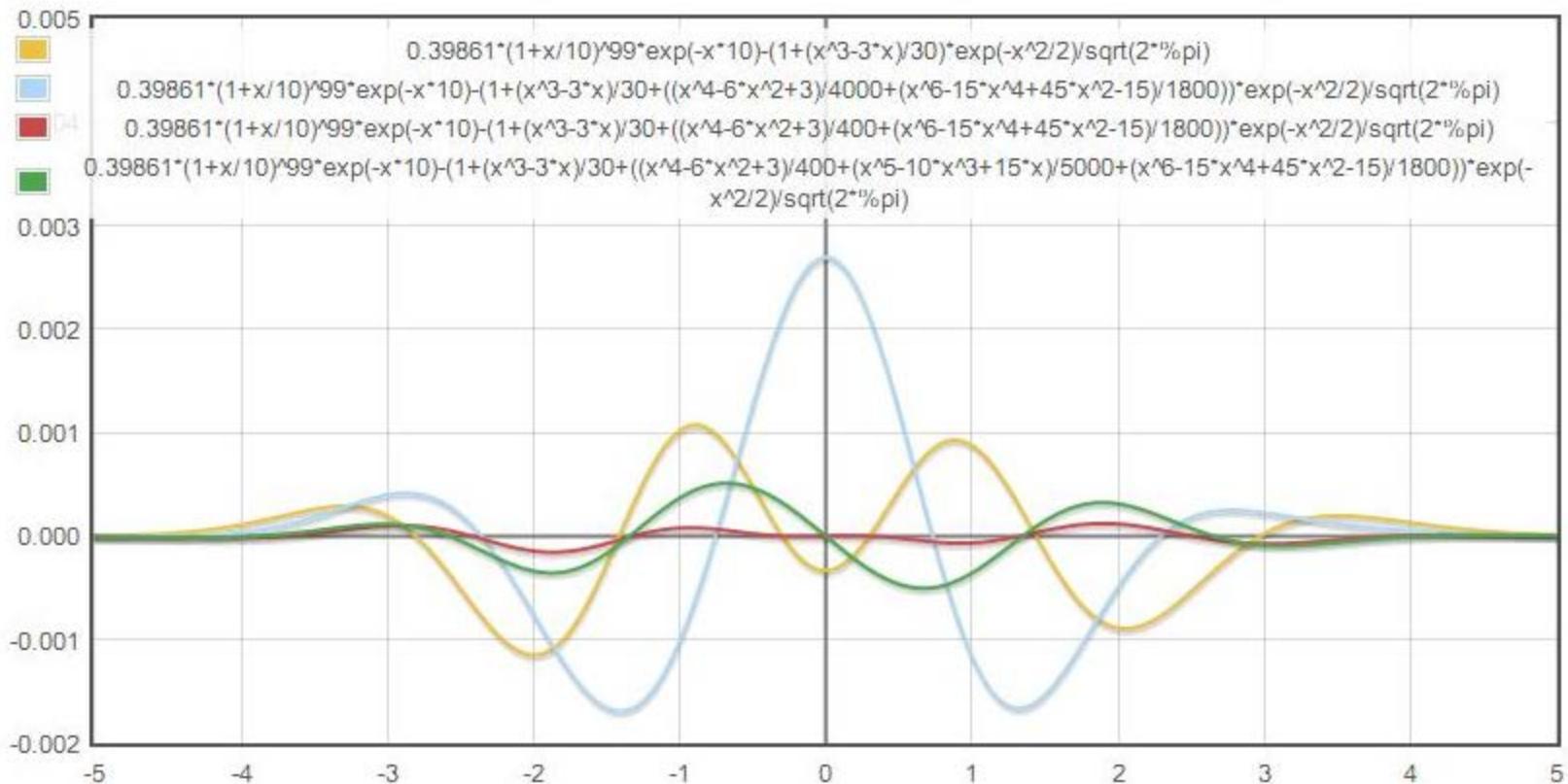
$$p_n(x) - \varphi(x) = \frac{A_1(x)}{\sqrt{n}} \varphi(x) + \frac{A_2(x)}{n} \varphi(x) + \frac{A_3(x)}{n^{3/2}} \varphi(x) + O\left(\frac{1}{n^2}\right), n \rightarrow \infty,$$

в виде

$$p_n(x) - \varphi(x) = \frac{\theta_3}{3! \sqrt{n}} H_3(x) \varphi(x) + \frac{\theta_4}{4! n} H_4(x) \varphi(x) + \frac{\theta_5}{5! n^{3/2}} H_5(x) \varphi(x) +$$

$$+ \frac{1}{2} \left( \frac{\theta_3}{3! \sqrt{n}} \right)^2 H_6(x) \varphi(x) + \frac{\theta_3}{3! \sqrt{n}} \frac{\theta_4}{4! n} H_7(x) \varphi(x) + \frac{1}{6} \left( \frac{\theta_3}{3! \sqrt{n}} \right)^3 H_9(x) \varphi(x) + O\left(\frac{1}{n^2}\right), n \rightarrow \infty,$$

и попробуем вычесть из левой части слагаемые правой части по одиночке.



Понятно, как можно обойти «нерегулярность»: новые члены разложения нужно добавлять «целиком». Но, если мы имеем информацию только о четвёртом моменте, добавить второй член «целиком» мы сможем, только сильно загрублляя оценку остаточной части. Вернёмся к неравенству

$$\left| p_n(x) - \varphi(x) - \left( \frac{\alpha_3}{3! \sqrt{n}} H_3(x) \varphi(x) + \frac{\lambda \alpha_4 - 3}{4! n} H_4(x) \varphi(x) + \frac{1}{2} \left( \frac{\alpha_3}{3! \sqrt{n}} \right)^2 H_6(x) \varphi(x) \right) \right| \leq \\ \leq \frac{1 - \lambda}{8\sqrt{2\pi}} \frac{\beta_4}{n} + O\left(\frac{1}{n^{3/2}}\right),$$

где  $0 \leq \lambda \leq 0.4$ . Следствием этого неравенства при  $\lambda = 0.4$  является неравенство

$$|p_n(x) - \varphi(x)| \leq \frac{|A_1(x)|}{\sqrt{n}} \varphi(x) + 0.03 \frac{\beta_4}{n} + \frac{1}{n} \left| \frac{0.4 \alpha_4 - 3}{4!} H_4(x) + \frac{1}{2} \left( \frac{\alpha_3}{3!} \right)^2 H_6(x) \right| \varphi(x) \\ + O\left(\frac{1}{n^{3/2}}\right)$$

