Две теоремы Бернштейна

В. ТИХОМИРОВ

Среди многих выдающихся достижений Сергея Натановича Бернштейна два его результата произвели особое впечатление на современников. Они вызвали множество восхищенных отзывов, дискуссий, обобщений и комментариев. Этими результатами были «вероятностное» доказательство теоремы Вейерштрасса и замечательное неравенство, известное ныне всем математикам как неравенство Бернштейна.

В 1885 году знаменитый немецкий математик — патриарх математического анализа Карл Вейерштрасс — доказал, что любую непрерывную на отрезке функцию можно с любой точностью приблизить алгебраическими многочленами.

Вот что это значит. Пусть $y = f(x)$ — функция, непрерывная на отрезке $[a;b]$, а ε — любое наперед заданное число (например, $\varepsilon = 1/10^m$). Тогда существует такой многочлен $p(x) = a_0x^n + \ldots + a_n$, где $a_0, a_1, \ldots, a_n$ — действительные числа, что

$$|f(x) - p(x)| < \varepsilon$$

при всех x из отрезка $[a;b]$. Таким образом, имеет место приближенное равенство $f(x) \approx p(x)$ с точностью до ε.

Рассмотрим теперь график функции $y = f(x)$ (рис.1) и сдвинем его на

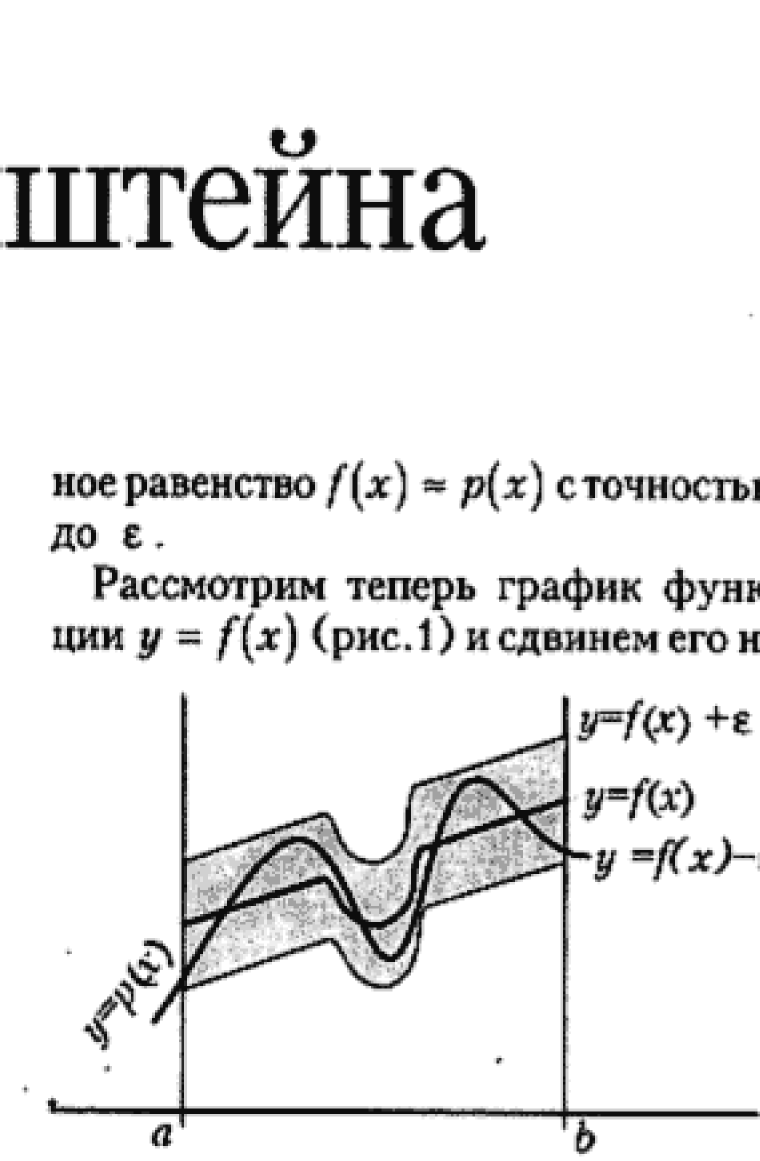

Рис. 1

ε вверх и на ε вниз, т.е. построим графики $y = f(x) + \varepsilon$ и $y = f(x) - \varepsilon$. На плоскости образовался «коридор», верхней и нижней границами которого служат построенные графики. Теорема Вейерштрасса утверждает, что внутри этого коридора содержится график некоторого многочлена.

Как и большинство выдающихся математических результатов, теорема Вейерштрасса допускала множество подходов к ее осмыслению. Появилось большое число доказательств этой теоремы.

Однако, когда в 1912 году вышла из печати полуторастраничная заметка С.Н.Бернштейна «Доказательство теоремы Вейерштрасса, основанное на теории вероятностей», оно произвело огромное впечатление неожиданностью подхода и красотой.

Вот это доказательство.

Все знают, что

$$(a+b)^2 = a^2 + 2ab + b^2,$$
$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3.$$

Известна и общая формула для $(a+b)^n$:

$$(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \ldots + \binom{n}{n-1}ab^{n-1} + \binom{n}{n}b^n = \sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k. \quad (1)$$

Она получила название формулы бинома Ньютона. Числа $\binom{n}{1}, \binom{n}{2}, \ldots, \binom{n}{k}$ называются биномиальными коэффициентами. Вот их выражения:

$$\binom{n}{0} = 1,\ \binom{n}{1} = n,$$
$$\binom{n}{2} = \frac{n(n-1)}{1\cdot 2},\ \ldots,\ \binom{n}{n} = 1,$$
$$\binom{n}{k} = \frac{n(n-1)\ldots(n-k+1)}{1\cdot 2\cdot\ldots\cdot k} = \frac{n!}{k!(n-k)!}.$$

Формула (1) без особого труда доказывается по индукции. Попробуйте это проделать самостоятельно.

В первых строках своей заметки С.Н.Бернштейн пишет:

«Мы укажем здесь очень простое доказательство следующей теоремы Вейерштрасса:

Если $F(x)$ — произвольная непрерывная функция на отрезке $[0;1]$, то сколь бы мало ни было ε, всегда можно определить многочлен $E_n(x) = a_0x^n + a_1x^{n-1} + \ldots + a_n$ достаточно высокой степени n, для которого имеет место неравенство

$$|F(x) - E_n(x)| < \varepsilon$$

в каждой точке рассматриваемого отрезка».

И Бернштейн выражает этот многочлен явной формулой:

$$E_n(x) = F(0)\binom{n}{0}x^n + F\left(\frac{1}{n}\right)\binom{n}{1}x^{n-1}(1-x) + \ldots + F\left(\frac{k}{n}\right)\binom{n}{k}x^{n-k}(1-x)^k + \ldots + F(1)\binom{n}{n}(1-x)^n = \sum_{k=0}^{n}F\left(\frac{k}{n}\right)\binom{n}{k}x^{n-k}(1-x)^k. \quad (2)$$

Полином $E_n(x)$ получил название полинома Бернштейна; обычно его обозначают в честь Бернштейна $B_n(x)$.

Доказательство теоремы Вейерштрасса основывается на двух формулах:

$$\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k} = (x + (1-x))^n = 1, \quad (3)$$

$$\sum_{k=0}^{n}(k-nx)^2\binom{n}{k}x^k(1-x)^{n-k} = nx(1-x) \quad (4)$$

и на таких двух утверждениях из классического анализа:

1) если $F(x)$ — непрерывная на отрезке $[0;1]$ функция, то для любого $\varepsilon > 0$ найдется $\delta > 0$ такое, что как только $|x - x'| < \delta$, так $|F(x) - F(x')| < \varepsilon/2$ (теорема Кантора),

2) непрерывная на конечном отрезке функция ограничена по модулю (теорема Вейерштрасса).

Выведем из (3), (4) и теоремы Кантора теорему Вейерштрасса, а затем выведем (4) и прокомментируем доказательство с вероятностной точки зрения. Имеем:

$$E_n(x) - F(x) = \sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}\left(F\left(\frac{k}{n}\right) - F(x)\right) = \sum_{\left|\frac{k}{n}-x\right|\le\delta} + \sum_{\left|\frac{k}{n}-x\right|>\delta} = S_1 + S_2;$$

При этом, если δ выбрано в соответствии с теоремой Кантора, то

$$|S_1| = \left|\sum_{\left|\frac{k}{n}-x\right|\le\delta}\binom{n}{k}x^k(1-x)^{n-k}\left(F\left(\frac{k}{n}\right) - F(x)\right)\right| \le \frac{\varepsilon}{2}\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k} = \frac{\varepsilon}{2}.$$

Обозначим через C число, ограничивающее максимальное по модулю значение функции. Тогда из определений, простейших свойств неравенств и (4) мы получим

$$|S_2| = \left|\sum_{\left|\frac{k}{n}-x\right|>\delta}\binom{n}{k}x^k(1-x)^{n-k}\left(F\left(\frac{k}{n}\right) - F(x)\right)\right| \le 2C\sum_{\left|\frac{k}{n}-x\right|>\delta}\binom{n}{k}x^k(1-x)^{n-k} \le$$
$$\le 2C\frac{1}{n^2\delta}\sum_{|k-nx|>n\delta}(k-nx)^2\binom{n}{k}x^k(1-x)^{n-k} \le 2C\frac{1}{n^2\delta}\sum_{k=0}^{n}(k-nx)^2\binom{n}{k}x^k(1-x)^{n-k} \le \frac{2Cx(1-x)}{n\delta} \le \frac{2C}{4n\delta}$$

(ибо $x(1-x) \le \frac{1}{4}$ при $0 \le x \le 1$).

Выбрав n столь большим, что $2C/(4n\delta) < \varepsilon/2$, получим, что $|E_n(x) - F(x)| < \varepsilon$ для любого x из $[0;1]$. Теорема доказана.

Формула (4) получается двукратным дифференцированием формулы

$$\sum_{k=0}^{n}\binom{n}{k}x^kb^{n-k} = (x+b)^n.$$

Дифференцируя ее один раз и умножая затем на x, получим

$$\sum_{k=0}^{n}k\binom{n}{k}x^k(x+b)^{n-k} = nx(x+b)^{n-1}. \quad (5)$$

Повторив еще раз ту же операцию, получим

$$\sum_{k=0}^{n}k^2\binom{n}{k}x^k(x+b)^{n-k} = nx(nx+b)(x+b)^{n-2}. \quad (6)$$

Подставив в (5), (6) вместо b величину $(1-x)$, получим

$$\sum_{k=0}^{n}k\binom{n}{k}x^k(1-x)^{n-k} = nx,$$

$$\sum_{k=0}^{n}k^2\binom{n}{k}x^k(1-x)^{n-k} = nx(nx+1-x),$$

и из этих двух формул читатель легко получит (4), используя то, что $(k-nx)^2 = k^2 - 2knx + n^2x^2$.

На самом деле мы по ходу дела вывели замечательный результат теории вероятностей — так называемый закон больших чисел Бернулли. Предоставим снова слово Бернштейну: «Рассмотрим событие A, вероятность которого равна x. Предположим, что произведено n испытаний, и мы условились платить некоторому игроку сумму $F\left(\frac{m}{n}\right)$, если событие A произойдет m раз. В этом случае математическое ожидание выигрыша (т.е. «средний выигрыш») будет иметь значение E_n». В силу теоремы Бернулли (а мы ее доказали),

$$\left|\sum_{\left|\frac{m}{n}-x\right|>\delta} \binom{n}{k} x^k (1-x)^{n-k} - F(x)\right| \le \frac{2C}{4n\delta},$$

откуда Бернштейн и вывел свою теорему.

В том же 1912 году Бернштейн доказал одно замечательное неравенство.

Пусть

$$p(x) = a_0 + a_1 \cos x + b_1 \sin x + \dots$$

$$\dots + a_n \cos nx + b_n \sin nx.$$

Такая функция называется тригонометрическим многочленом степени n. Пусть также $M = \max|p(x)|$, а $M' = \max|p'(x)|$. Тогда $M \le nM'$.

Иначе говоря, максимум модуля производной тригонометрического многочлена степени n не больше чем в n раз превосходит максимум самого многочлена.

Простейший пример тригонометрического многочлена степени n

$$q(x) = a\cos nx + b\sin nx = \sqrt{a^2+b^2}\cos(nx+\varphi)$$

показывает, что оценка эта достигается, так как в этом случае

$$M = \max|q(x)| = \sqrt{a^2+b^2},$$

а

$$M' = \max|q'(x)| = \max\left|n\sqrt{a^2+b^2}\sin(nx+\varphi)\right| = n\sqrt{a^2+b^2},$$

т.е. $M' \le nM$.

В дальнейшем все функции будут рассматриваться на отрезке $[-\pi;\pi]$, что естественно ввиду периодичности тригонометрических многочленов с периодом 2π.

Этот результат снова вызвал оживленную дискуссию, появилось множество его доказательств и истолкований. Здесь я приведу доказательство, принадлежащее знаменитому бельгийскому математику Валле Пуссену (1866 – 1962).

В доказательстве мы используем четыре элементарных факта математического анализа. Эти факты наглядно очевидны, да и доказываются достаточно просто при более подробном знакомстве с действительными числами. Мы их иллюстрируем рисунками.

1) Если на отрезке непрерывная функция принимает в двух точках значения разных знаков, то между этими точками она имеет нуль (теорема Коши) (рис. 2).

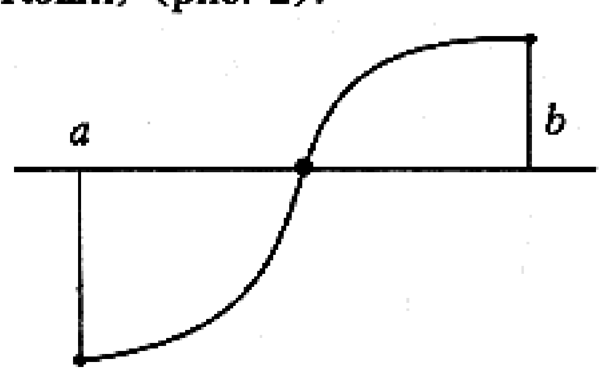

Рис. 2

2) Если дифференцируемая функция принимает в двух точках нулевое значение, то между ними есть точка, где производная функции равна нулю (теорема Ролля) (рис. 3).

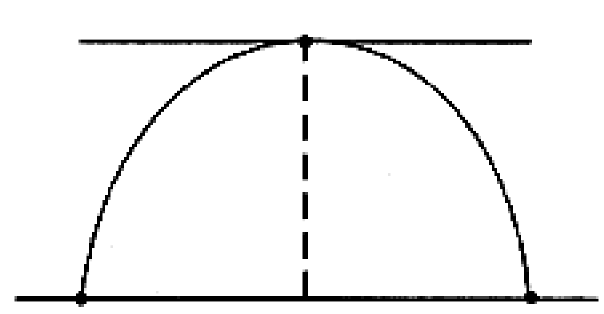

Рис. 3

3) Если дифференцируемая функция достигает максимума в некоторой точке, то в этой точке производная равна нулю (теорема Ферма) (см. рис. 3).

4) Тригонометрический полином степени n не может иметь больше $2n$ нулей на промежутке $[0;2\pi)$.

Итак, приступим к доказательству неравенства Бернштейна. Достаточно доказать (убедитесь в этом), что если $M \le 1$, то $M' \le n$.

Пусть существует тригонометрический многочлен $\bar{p}(x)$ степени n такой, что в любой точке x справедлива оценка $|\bar{p}(x)| \le 1$, в то время как в некоторой точке $\bar{x}$ его производная принимает максимальное значение, равное M' и превосходящее n. Не ограничивая общности, можно считать, что $\bar{x} = 0$ (иначе мы рассмотрели бы многочлен $\tilde{p} = \bar{p}(\bar{x} - x)$, для которого максимум модуля производной достигается при $x = 0$, а значения M и M' — те же самые).

Рассмотрим многочлен

$$q(x) = \bar{p}(x) - \left(\frac{M'}{n}\right)\sin nx.$$

Функция $\left(\frac{M'}{n}\right)\sin nx$ колеблется $2n$ раз между значениями M'/n и $-M'/n$ и при этом M'/n больше единицы. Полином же $\bar{p}(x)$ в каждой точке x по модулю не превосходит единицы.

Поэтому в точках $x_k = x_0 + \frac{\pi k}{n}$, где $x_0 = \frac{\pi}{2n}$ $(k = 0, 1, \dots, 2n)$, многочлен $q(x)$ принимает попеременно положительные и отрицательные значения, так что график функции q имеет вид, показанный на рисунке 4, и, в силу утверждения 1, уравнение $q(x) = 0$

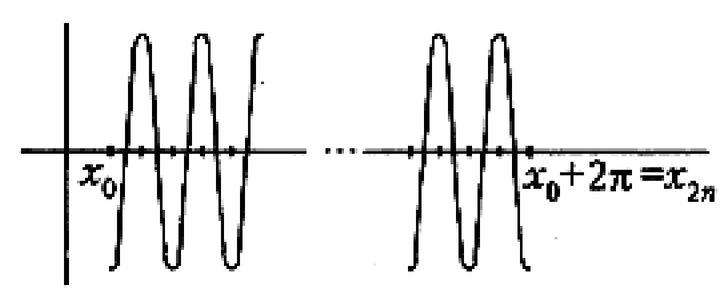

Рис. 4

имеет не меньше $2n$ корней на интервале $(x_0; x_0 + 2\pi)$, а следовательно, и на интервале $(0;2\pi)$. По утверждению 2, многочлен $q'(x)$ имеет не менее $2n-1$ нулей на $(0;2\pi)$. Кроме того,

$$q'(0) = \bar{p}'(0) - \frac{M'}{n}\cdot n = M' - M' = 0,$$

и следовательно, на полуинтервале $[0;2\pi)$ фунция $q'(x)$ имеет $2n$ нулей, а на отрезке $[0;2\pi]$ — $2n+1$ нуль(поскольку $q'(2\pi) = q'(0) = 0$). Возьмем производную еще раз. Многочлен $q''(x)$ также имеет степень n, обращается в нуль в $2n$ точках интервала $(0;2\pi)$ и, кроме того $q''(0) = 0$, так как

$$q''(x) = \bar{p}''(x) + M'\sin nx$$

и $\bar{p}'(x)$ достигает в нуле своего максимума, т.е. $\bar{p}''(0) = 0$.

Значит, $q''(x)$ имеет больше $2n$ нулей на полуинтервале $[0;2\pi)$. Но, в силу утверждения 4, это невозможно! Получилось противоречие, и тем самым теорема доказана.