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 но 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

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

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