The Bernstein polynomial $B_n(f)$ of a function $f$ defined on $[0, 1]$ is defined as

$$B_n(f)(x) = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} f\left(\frac{k}{n}\right)$$

**Approximation theorem**

Let $f$ be a function defined on $[0, 1]$. For each point $x$ of continuity of $f$, $B_n(f)(x) \to f(x)$ as $n \to \infty$. If $f$ is continuous on $[0, 1]$, then the Bernstein polynomial $B_n(f)$ converges to $f$ uniformly i.e., $\max_{x \in [0,1]} |f(x) - B_n(f)| \to 0$. Moreover for $x$ a point of differentiability of $f$, $B_n'(f)(x) \to f'(x)$ If $f$ is continuously differentiable on $[0, 1]$, then $B_n'(f)(x) \to f'(x)$ uniformly.
We have the following formulae

\[ B_n(1) = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} = 1 \]

\[ B_n(x) = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \left( \frac{k}{n} \right) = x \]

\[ B_n(x^2) = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \left( \frac{k}{n} \right)^2 = \frac{n-1}{n} x^2 + \frac{x}{n} \]

\[ \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \left( \frac{k}{n} \right)^2 = \frac{x(1-x)}{n} \]

\[ \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \left( \frac{k}{n} \right)^4 = \frac{x(1-x)(1+x(1-x)(3n-6))}{n^3} \]
Probabilistic idea:
\( \binom{n}{k} x^k (1 - x)^{n-k} \) is the probability of getting \( k \) heads in \( n \) throws if the probability of getting head is \( x \).— See Bernoulli distribution.

Hence the above expression can be interpreted as the expectation of the random variable \( f(K/n) \), where \( K \) is Bernoulli variable with probability parameter \( x \)
The expected value of \( K \) is \( nx \) and by law of large numbers, the probability is mostly concentrated around \( k \approx nx \). So \( f(k/n) \) is almost likely \( f(x) \). Hence the expected value which is \( B_n(f) \) behaves like \( f(x) \) for large \( n \).
Proof of the theorem:

\[ B_n(f)(x) = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} f\left(\frac{k}{n}\right) \]

\[ f(x) = f(x) \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \]

\[ B_n(f)(x) - f(x) = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} f\left(\frac{k}{n}\right) - \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} f(x) \]

\[ |B_n(f)(x) - f(x)| = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} |f\left(\frac{k}{n}\right) - f(x)| \]
We consider terms with $|\frac{k}{n} - x| < \delta$ and those with $|\frac{k}{n} - x| \geq \delta$ separately. As discussed in the proof, the former contributes the most to the sum. $\delta$ is chosen as follows

Given $\varepsilon > 0$, for $x$ a point of continuity we have $|f(x) - f(y)| < \varepsilon$ if $|x - y| < \delta$.

\[
\sum_{|\frac{k}{n} - x| < \delta} \binom{n}{k} x^k (1 - x)^{n-k} |f\left(\frac{k}{n}\right) - f(x)| < \sum_{|\frac{k}{n} - x| < \delta} \binom{n}{k} x^k (1 - x)^{n-k} \varepsilon
\]

\[
< \sum_{k=0}^{n} \binom{n}{k} x^k (1 - x)^{n-k} \varepsilon = \varepsilon
\]
\[
\sum_{|\frac{k}{n} - x| \geq \delta} \binom{n}{k} x^k (1 - x)^{n-k} |f(\frac{k}{n}) - f(x)| < \sum_{|\frac{k}{n} - x| \geq \delta} \binom{n}{k} x^k (1 - x)^{n-k} (2M)
\]

\[
< 2M \sum_{|\frac{k}{n} - x| \geq \delta} \frac{(x - \frac{k}{n})^2}{\delta^2} \binom{n}{k} x^k (1 - x)^{n-k}
\]

\[
< 2M \sum_{k=0}^{n} \frac{(x - \frac{k}{n})^2}{\delta^2} \binom{n}{k} x^k (1 - x)^{n-k} = \frac{2Mx(1-x)}{n\delta^2} < \frac{2M}{n\delta^2}
\]

Choosing \( n \) large enough so that \( \frac{2M}{n\delta^2} < \varepsilon \) we have

\[
|f(x) - B_n(f)| < 2\varepsilon
\]

\( \square \)