---

We give a sketch of the proof of de Moivre's theorem in the symmetric case $p = q = 1/2$.

Reading this proof is not required. It will not help you apply the CLT for binomial random variable. However, it is a very good read if you are interested in an elementary but sophisticated mathematical exercise.

For a fixed integer $k$, let us try to understand the quantity $P(S = k)$ when $S$ is a binomial $\mathrm{Bin}(n, 1/2)$ random variable. We know that

$$P(S = k) = 2^{-n} \frac{n!}{k!(n-k)!}.$$

We also know Stirling's formula which says that, for large $n$,

$$n! \sim \sqrt{2\pi n}\, n^n\, e^{-n}.$$

We are interested in values of $k$ that are not very far from the mean of $S$, $E(S) = n/2$. So we will set

$$k - n/2 = x, \quad k = n/2 + x$$

where $x$ is much smaller than $n$, namely, $|x| \le A\sqrt{n}$, and observe that this implies that both $k \sim n/2$ and $n - k \sim n/2$ are large when $n$ is large. Using Stirling's formula three times, we see that

$$\begin{aligned} P(S = k) \quad &\sim \quad 2^{-n} \frac{\sqrt{2\pi n}\, n^n\, e^{-n}}{\sqrt{2\pi k}\, k^k\, e^{-k}\, \sqrt{2\pi(n-k)}\, (n-k)^{n-k}\, e^{-(n-k)}} \\ &\sim \quad \frac{1}{\sqrt{2\pi}} \frac{\sqrt{n}}{\sqrt{k(n-k)}} \left(\frac{n}{2k}\right)^k \left(\frac{n}{2(n-k)}\right)^{n-k}. \end{aligned}$$

Now, we observe that

$$\frac{2k}{n} = \frac{n+2x}{n} = \left(1 + \frac{2x}{n}\right), \quad \frac{2(n-k)}{n} = \frac{n-2x}{n} = \left(1 - \frac{2x}{n}\right)$$

and (because $x$ is much smaller than $n$, we have $k \sim n - k \sim n/2$)

$$\sqrt{\frac{n}{k(n-k)}} \sim \frac{1}{\sqrt{n/4}}$$

and report this in the previous computation to obtain

$$P(S = k) \quad \sim \quad \frac{1}{\sqrt{\pi n/2}} \left(1 + \frac{2x}{n}\right)^{-k} \left(1 - \frac{2x}{n}\right)^{-(n-k)}.$$

Because $x/n$ is small and $k$ and $n - k$ are close to $n/2$ (we allow $x$ to depends on $n$ but $x/n$ is assumed to go to 0 as $n$ tends to infinity), we can use the approximations

$$\left(1 + \frac{2x}{n}\right)^{-k} \sim e^{-k\left(\frac{2x}{n} - \frac{1}{2}\left(\frac{2x}{n}\right)^2\right)}, \quad \left(1 - \frac{2x}{n}\right)^{-(n-k)} \sim e^{(n-k)\left(\frac{2x}{n} + \frac{1}{2}\left(\frac{2x}{n}\right)^2\right)}.$$

The educated way to prove these statement is to write, for the first term,

$$\left(1 + \frac{2x}{n}\right)^{-k} = e^{-k \log(1 + \frac{2x}{n})} = e^{-k\left(\frac{2x}{n} - \frac{1}{2}\left(\frac{2x}{n}\right)^2 + \frac{1}{3}\left(\frac{2x}{n}\right)^3 - \frac{1}{4}\left(\frac{2x}{n}\right)^4 + \dots\right)}.$$

Similarly, for the second term,

$$\left(1-\frac{2x}{n}\right)^{-(n-k)} = e^{(k-n)\log(1-\frac{2x}{n})} = e^{-(k-n)\left(\frac{2x}{n}+\frac{1}{2}\left(\frac{2x}{n}\right)^2+\frac{1}{3}\left(\frac{2x}{n}\right)^3+\frac{1}{4}\left(\frac{2x}{n}\right)^4+\dots\right)}.$$

Recall that $k\sim n/2$ and that $|x|\le A\sqrt{n}$. This tells us that that the term $\frac{1}{3}\left(\frac{2x}{n}\right)^3 k$ and all the terms after it in the expansion tend to 0. This is the basic reason why we can ignore these terms and why we must keep the first two terms.

Now, remember that $2k-n=2x$ and write

$$\begin{aligned} P(S=k) &\sim \frac{1}{\sqrt{\pi n/2}}e^{-\frac{2x}{n}(2k-n)+\frac{1}{2}\frac{(2x)^2}{n}} \\ &= \frac{1}{\sqrt{\pi n/2}}e^{-\frac{2x^2}{n}} = \frac{1}{\sqrt{\pi n/2}}e^{-\frac{x^2}{n/2}} = \frac{1}{\sqrt{\pi n/2}}e^{-\frac{(k-n/2)^2}{n/2}} \end{aligned}$$

To finish the proof, we need to show that, for any fixed $a<b$,

$$P(a\le \frac{S-n/2}{\sqrt{n/2}}\le b) \quad \text{is close to} \quad \int_a^b \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx.$$

To see how the previous computation leads to this result, write

$$\begin{aligned} P\left(a\le \frac{S-n/2}{\sqrt{n/2}}\le b\right) &= = \sum_{n/2+a\sqrt{n/2}\le k\le n/2+b\sqrt{n/2}} P(S=k) \\ &\approx \sum_{n/2+a\sqrt{n/2}\le k\le n/2+b\sqrt{n/2}} \frac{1}{\sqrt{\pi n/2}}e^{-\frac{(k-n/2)^2}{n/2}} \\ &= \frac{1}{\sqrt{n/4}}\sum_{x_k} f(x_k) \end{aligned}$$

where $f(t)=\frac{1}{\sqrt{2\pi}}e^{-t^2/2}$ and the points $x_k=(k-n/2)/\sqrt{n/4}$ (almost) form an equipartition of the interval $[a,b]$ into subintervals of length $\sqrt{n/4}$. This tells us that the last sum above is a Riemann sum for the function $f$ over $[a,b]$ and it follows that

$$P\left(a\le \frac{S-n/2}{\sqrt{n/2}}\le b\right)\to \int_a^b \frac{1}{\sqrt{2\pi}}e^{-t^2/2}$$

as desired.