Moments of the Binomial Distribution:

Theorem 5.3.3 *Moment generating function of the binomial distribution is*

$$M_X(t) = (1 + \theta(e^t - 1))^n.$$

Proof 5.3.4

$$\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= \sum_{x=0}^{n} (e^{tx}) \cdot \binom{n}{x} \theta^x (1-\theta)^{n-x} \\
&= \sum_{x=0}^{n} \binom{n}{x} (e^t\theta)^x (1-\theta)^{n-x} \\
&= (e^t\theta + (1-\theta))^n \quad \textit{(by the binomial theorem)} \\
&= (1 + \theta(e^t - 1))^n.
\end{aligned}$$

From the moment generating function, we can find the mean.

Finding the **mean** from the moment generating function:

$$\mu = \frac{d}{dt} M_X(t) \bigg|_{t=0}$$

$$= \frac{d}{dt} (1 + \theta(e^t - 1))^n \bigg|_{t=0}$$

$$= n(1 + \theta(e^t - 1))^{n-1} \cdot (\theta e^t) \big|_{t=0}$$

$$= n(1 + \theta(e^0 - 1))^{n-1} \cdot (\theta e^0)$$

$$= n\theta.$$

Next we want to find the **variance**.
First, we need the second moment about the origin:

$$
\begin{aligned}
E(X^2) &= \frac{d^2}{dt^2} M_X(t) \Big|_{t=0} \\
&= \frac{d^2}{dt^2} (1 + \theta(e^t - 1))^n \Big|_{t=0} \\
&= \frac{d}{dt} n\theta e^t (1 + \theta(e^t - 1))^{n-1} \Big|_{t=0} \\
&= n\theta e^t (1 + \theta(e^t - 1))^{n-1} \\
&\quad + n(n-1)\theta e^t (1 + \theta(e^t - 1))^{n-2} \cdot (\theta e^t) \big|_{t=0} \\
&= n\theta + n(n-1)\theta^2
\end{aligned}
$$

Finally, we can use the formula for the variance:

$$
\sigma^2 = E(X^2) - \mu^2 = n\theta + n(n-1)\theta^2 - (n\theta)^2 = n\theta - n\theta^2 = n\theta(1-\theta).
$$

We obtained the following theorem:

Theorem 5.3.5 *The mean and variance of the binomial distribution:*

$$
\mu = n\theta, \quad \sigma^2 = n\theta(1-\theta)
$$

Consider a random variable Y that denotes the proportion of successes in n trials.

So, $Y = \frac{X}{n}$, where X is the binomial random variable.
Then, the following holds.

Theorem 5.3.6 *Let X be a binomial random variable and let $Y = \frac{X}{n}$. Then*

$$E(Y) = \theta, \quad \sigma_Y^2 = \frac{\theta(1-\theta)}{n}.$$

By Chebyshev's Theorem, with $C = k\sigma$ ($k = \frac{C}{\sigma}$) we have

$$P(|Y - \theta| < C) \geq 1 - \frac{1}{k^2} = 1 - \frac{1}{(\frac{C^2}{\sigma^2})} = 1 - \frac{\theta(1-\theta)}{C^2 n}$$

Thus for any value of $C > 0$ we have

$$P\left(\left|\frac{X}{n} - \theta\right| < C\right) \geq 1 - \frac{\theta(1-\theta)}{C^2 n}.$$

When n is large, the fraction on the right side gets small, and so

$$\lim_{n\to\infty} P\left(\left|\frac{X}{n}-\theta\right|<C\right)=1.$$

This holds for any $C>0$, no matter how small.

Explanation: The more trials we perform, the more likely it is that the proportion of successes will be close to the probability of a success θ.

Example: In repeatedly flipping a balanced coin, the more flips we perform (n), the more likely that the proportion of heads obtained ($\frac{X}{n}$) will be 0.5 (θ).

The Binomial distribution gives the probability of getting x successes in n trials.

Suppose we want to know the probability that the kth success occurs precisely on trial n.

For the kth success to occur on the nth trial, there must be exactly $k-1$ successes on the first $n-1$ trials, and consequently $n-1-(k-1) = n-k$ failures.

If θ is the probability of a success on a given trial, then the probability of getting $k-1$ successes in $n-1$ trails is

$$b(k-1; n-1, \theta) = \binom{n-1}{k-1} \theta^{k-1} (1-\theta)^{n-k}.$$

Then the probability that the kth success is on trial n is

$$b^*(k; n, \theta) = \theta \cdot b(k-1; n-1, \theta) = \binom{n-1}{k-1} \theta^{k} (1-\theta)^{n-k}.$$

The special case when $k = 1$ (first success appears in trial n) is called the **geometric distribution**:

$$g(n; \theta) = b^*(1; n, \theta) = \theta(1 - \theta)^{n-1}$$

For a geometric distribution, we have the following:

$$\mu = \frac{1}{\theta}, \quad \sigma^2 = \frac{1}{\theta}\left(\frac{1}{\theta} - 1\right) = \frac{1 - \theta}{\theta^2}$$