STAT241/251 Lecture Notes
Chapter 6 Part 3

Yew-Wei Lim

Next topic: Normal approximation to the Binomial distribution (Ch7.2)

Picture this:

$$X \sim Binomial(\underbrace{36}_{n}, \overbrace{\frac{1}{3}}^{p})$$

Find $P(X > 13)$.

You'll realize it is tedious:

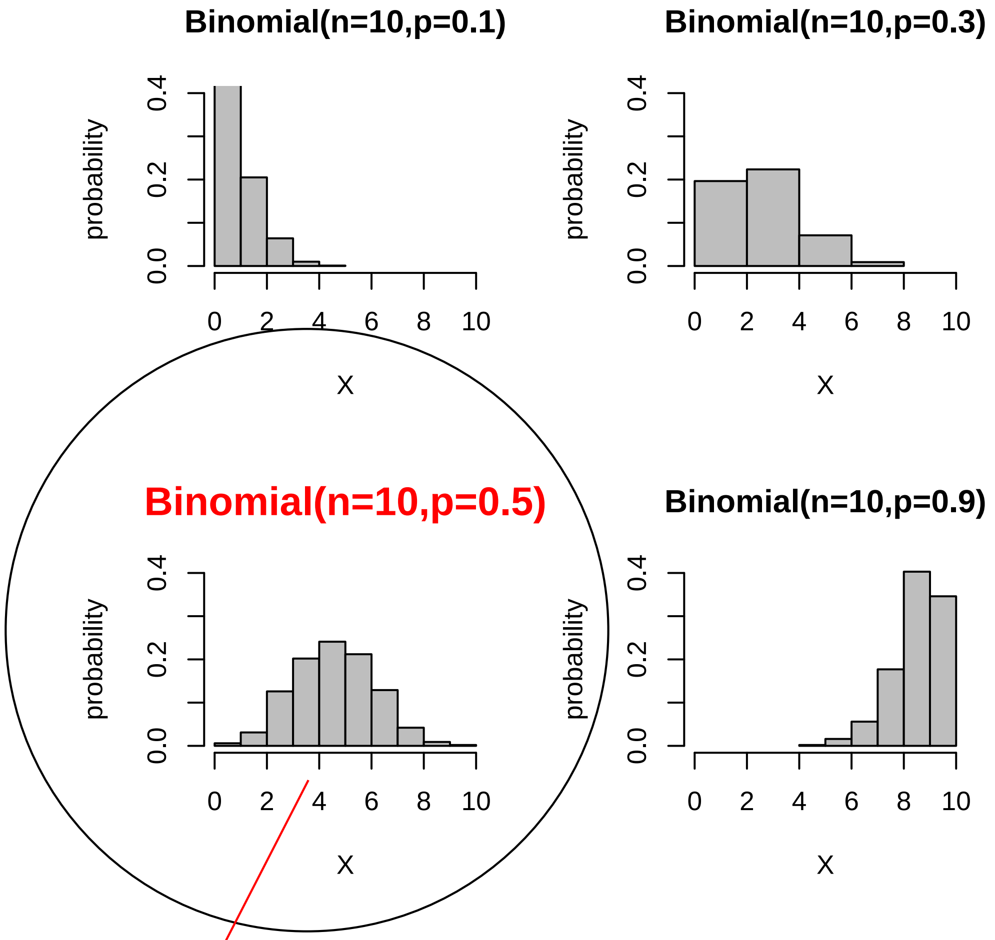
$$P(X > 13) = P(X = 14) + P(X = 15) + \cdots + P(X = 36).$$

Shorter, but still tedious:

$$P(X > 13) = 1 - P(X \leq 13) = 1 - [P(X = 0) + P(X = 1) + \cdots + P(X = 13)]$$

When n is large, and $np \geq 5$ and $nq \geq 5$, we can use the normal distribution to get an approximate answer.

Important:	Check n is large by np$\geq$ 5 and nq $\geq$ 5 before using normal approximation.



For instance, when n=10 and p = 0.5 $\Rightarrow$

$$np \geq 5, nq \geq 5$$

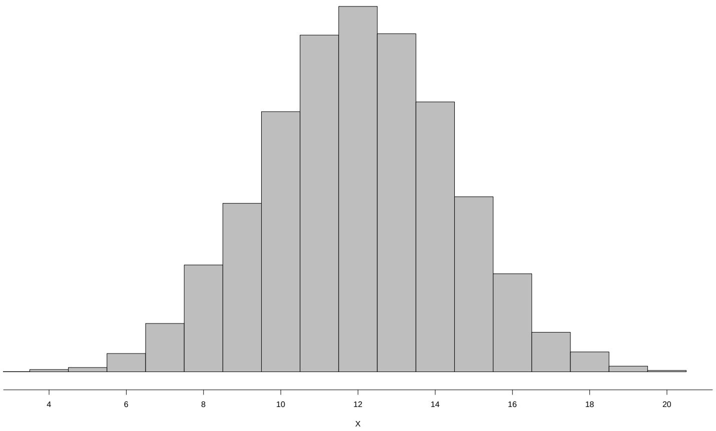
and we can use normal approximation.

When $np \geq 5, nq \geq 5$ condition is not satisfied, do not use normal approximation to binomial.

Remember to check $np \geq 5$, and $nq \geq 5$ before using normal approximation.

Idea:

When n is large and $np \geq 5$, and $nq \geq 5$, the shape of the binomial distribution is approximately symmetrical.



When using continuous normal distribution to approximate discrete r.v., use continuity correction!

Steps: when using continuity correction:

Example:

Because normal is continuous

– To find $P(X \leq 10)$, use $P(X \leq 10.5)$

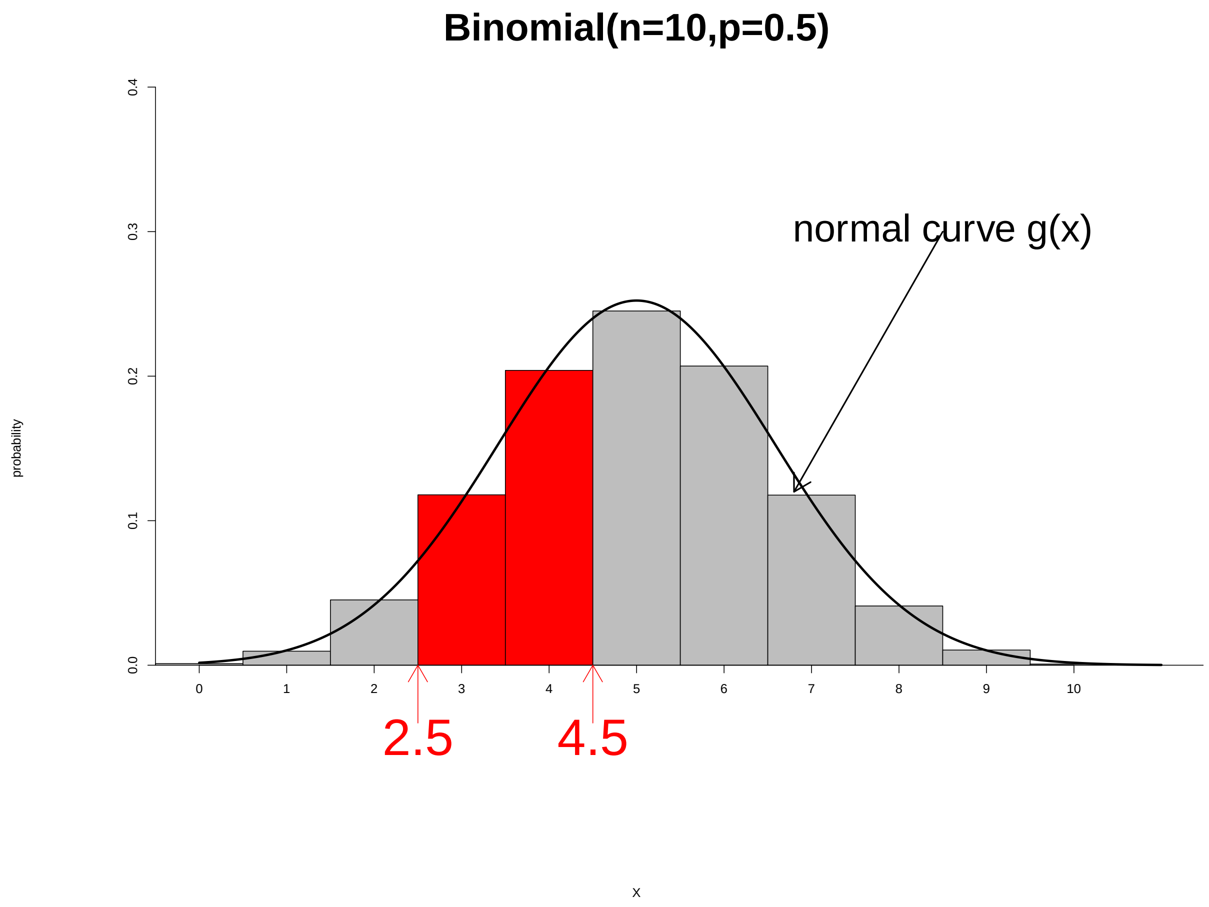
– To find $P(X \geq 14)$, use $P(X \geq 13.5)$

What if we need to find $P(X > 13)$? First, rewrite using equality sign, eg $P(X > 13) = P(X \geq 14) = P(X \geq 13.5)$, [*continuity correction*]

What about

$$P(10 \leq X \leq 14)\text{? Use } P(9.5 \leq X \leq 14.5)$$

Why? Best explained using a graph (see next page for explanation)



Note that binomial is a discrete r.v. but normal is a contiuous r.v. Consider $X \sim Binomial(10, 0.5)$ and say our goal is to estimate $P(3 \leq X \leq 4)$.

$$P(3 \leq X \leq 4) = \text{red region}$$
$$= \int_{\boxed{2.5}}^{\boxed{4.5}} g(x)dx$$

Use $\int_{2.5}^{4.5} g(x)dx$ rather than $\int_{3}^{4} g(x)dx$ when finding area under normal curve.

Let's see how to apply normal approximation with continuity correction to approximate binomial distribution.

Question: $X \sim Binomial(36, \frac{1}{3})$, Find $P(X > 13)$

<u>Sol:</u>

Since

$$np = 36 \times \frac{1}{3} = 12 \geq 5$$

and

$$nq = 36 \times \frac{2}{3} = 24 \geq 5,$$

we can use normal approximation to binomial distribution.

$$\begin{aligned} E(X) =& np = 36 \times \frac{1}{3} = 12 \\ Var(X) =& npq = 36 \times \frac{2}{3} \times \frac{1}{3} = 8. \end{aligned}$$

use np to get μ use npq to get σ^2

$$\rightarrow X \overset{approx.}{\sim} Normal(12, 8)$$

still treating X as discrete

$$\begin{aligned} P(X > 13) =& P(X \geq 14) && \text{[First, write as equality sign]} \\ =& P(X \geq 13.5) && \text{[figure out continuity correction]} \\ =& 1 - P(X \leq 13.5) \\ =& 1 - P(Z \leq \frac{13.5 - 12}{\sqrt{8}}) \\ =& 1 - P(Z \leq 0.530) \\ =& 1 - 0.7019 \\ =& 0.298 \end{aligned}$$

treat X as continuous now

Note: You are expected to use continuity correction in assignments and exams.

Example:

It may be assumed that dates of birth in a large population are distributed throughout the year so that the probability of a randomly chosen person's date of birth being in any particular month may be taken as $\frac{1}{12}$.

(a) Find the probability that 6 people chosen at random, exactly two will have birthdays in January.

(b) Find the probability that 8 people chosen at random, at least one will have a birthday in January.

(c) N people are chosen at random. Find the least value of N so that the probability that at least one will have a birthday in January exceeds 0.9.

(d) Find the probability that of 100 people chosen at random, at least 40 will have birthdays in May, June, July or August.

<u>Solution:</u>

Let X be the number of people with birthday in January.

(a) $X \sim Binomial(6, \frac{1}{12})$

$$\begin{aligned} P(X=2) =& {}^6C_2(\frac{1}{12})^2(\frac{11}{12})^4 \\ =& 0.0735 \end{aligned}$$

(b) $X \sim Binomial(8, \frac{1}{12})$

$$\begin{aligned} P(X \geq 1) =& 1 - P(X=0) \\ =& 1 - \underbrace{(\frac{11}{12})^8}_{{}^8C_0(\frac{1}{12})^0(\frac{11}{12})^8} \\ =& 0.5015 \end{aligned}$$

(c) $X \sim Binomial(N, \frac{1}{12})$

$$\begin{aligned}
P(X \geq 1) &> 0.9 \\
1 - P(X = 0) &> 0.9 \\
P(X = 0) &< 0.1 \\
(\frac{11}{12})^N &< 0.1 \\
N \ln(\frac{11}{12}) &< \ln(0.1) \\
N &> \frac{\ln 0.1}{\ln(\frac{11}{12})} \quad \text{[Note: sign changes} \\
N &> 26.46 \quad \text{because } \ln(\tfrac{11}{12}) \\
& \quad \text{is negative]}
\end{aligned}$$

$\therefore N = 27$ since N must be an integer.

(d) Let Y be the number of people with birthdays in May, June, July or August.

$$Y \sim Binomial(\underbrace{100}_{n}, \underbrace{\frac{4}{12} = \frac{1}{3}}_{p})$$

We need to find $P(Y \geq 40)$. Since $np = \frac{100}{3} \geq 5$, and $nq = 100 \times \frac{2}{3} \geq 5$,

We can use the normal distribution to approximate Y.

$$\begin{aligned}
E(Y) &= np = \frac{100}{3} \\
Var(Y) &= npq = \frac{100}{1} \times \frac{1}{3} \times \frac{2}{3} = \frac{200}{9}
\end{aligned}$$

$$Y \overset{approx.}{\sim} N(\frac{100}{3}, \frac{200}{9})$$

$$\begin{aligned}
P(Y \geq 40) =& P(Y \geq 39.5) \qquad \text{[apply continuity correction]} \\
=& 1 - P(Y \leq 39.5) \\
=& 1 - P(Z \leq \frac{39.5 - \frac{100}{3}}{\sqrt{\frac{200}{9}}}) \\
=& 1 - P(Z \leq 1.31) \\
=& 1 - 0.9049 \\
=& 0.0951
\end{aligned}$$

(We've completed Ch 6.1, 6.2 and 7.2)