Normal Approximation to the Binomial

MATH 130, *Elements of Statistics I*

J Robert Buchanan

Department of Mathematics

Fall 2023

Background

Recall: a binomial experiment must satisfy the following criteria:

1. the experiment is performed a fixed number of times. Each repetition of the experiment is called a **trial**.
2. the trials are all independent. The outcome of one trial does not affect the outcome of any other trial.
3. for each trial, there are two mutually exclusive outcomes generally thought of as "success" or "failure".
4. the probability of success is the same for each trial.

Binomial Probabilities

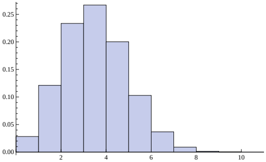
If the number of trials is n, the probability of success on a single trial is p, and random variable X is the number of successes, then

$$
\begin{aligned}
P(X = x) &= ({}_nC_x)\, p^x (1-p)^{n-x} \\
P(X \geq x) &= \sum_{k=x}^{n} P(X = k) = \sum_{k=x}^{n} ({}_nC_k)\, p^x (1-p)^{n-k} \\
P(X \leq x) &= \sum_{k=0}^{x} P(X = k) = \sum_{k=0}^{x} ({}_nC_k)\, p^x (1-p)^{n-k}
\end{aligned}
$$

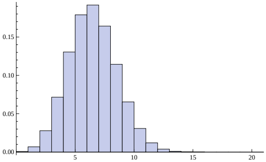
Sometimes Tables III and IV can help.

Histograms

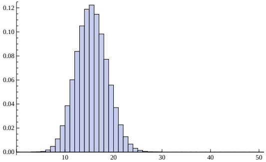
Consider the histograms of the binomial probability distribution for $p = 0.30$ and three different values of n.



$n = 10$



$n = 20$



$n = 50$

Normal Approximation

Theorem

If $np(1-p) \geq 10$ *the binomial random variable* X *is approximately normally distributed with mean* $\mu_X = np$ *and standard deviation* $\sigma_X = \sqrt{np(1-p)}$.

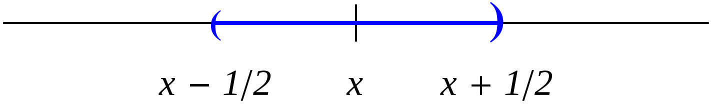
Continuity Correction

Continuity Correction: we are using a **continuous** random variable to approximate a **discrete** random variable. This affects how we think of $P(X = x)$.

Binomial (discrete)	**Normal (continuous)**
$P(X = x)$	$P(x - 0.5 < X < x + 0.5)$
$P(X \leq x)$	$P(X < x + 0.5)$
$P(X \geq x)$	$P(X > x - 0.5)$
$P(X < x)$	$P(X < x - 0.5)$
$P(X > x)$	$P(X > x + 0.5)$

Illustration (1 of 5)

Binomial: $P(X = x)$



Normal: $P(x - 0.5 < X < x + 0.5)$

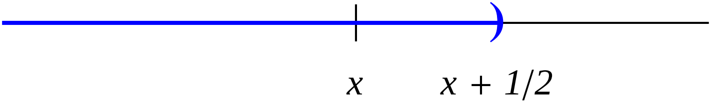
Example

If $n = 50$ and $p = 0.4$ then $np(1 - p) = 12 \geq 10$ and

$$P(X = 18) \approx P(17.5 < X < 18.5).$$

Illustration (2 of 5)

Binomial: $P(X \leq x)$



Normal: $P(X < x + 0.5)$

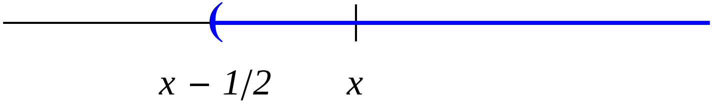
Example

If $n = 60$ and $p = 0.45$ then $np(1 - p) = 14.85 \geq 10$ and

$$P(X \leq 20) \approx P(X < 20.5).$$

Illustration (3 of 5)

Binomial: $P(X \geq x)$



Normal: $P(X > x - 0.5)$

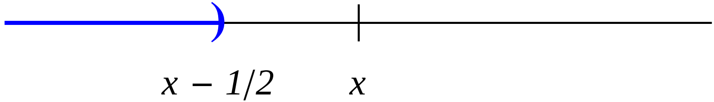
Example

If $n = 75$ and $p = 0.33$ then $np(1 - p) = 16.58 \geq 10$ and

$$P(X \geq 23) \approx P(X > 22.5).$$

Illustration (4 of 5)

Binomial: $P(X < x)$



Normal: $P(X < x - 0.5)$

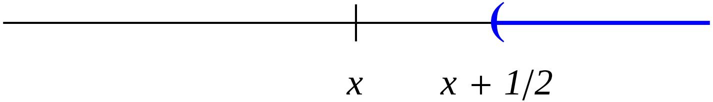
Example

If $n = 50$ and $p = 0.60$ then $np(1 - p) = 12 \geq 10$ and

$$P(X < 10) \approx P(X < 9.5).$$

Illustration (5 of 5)

Binomial: $P(X > x)$



Normal: $P(X > x + 0.5)$

Example

If $n = 80$ and $p = 0.42$ then $np(1 - p) = 19.49 \geq 10$ and

$$P(X > 32) \approx P(X > 32.5).$$

Example

Sixty-two percent of Americans own a cell phone. Suppose 100 Americans are randomly selected. Use the normal approximation to the binomial to determine the following.

1. The probability that at least 70 of them own a cell phone.
2. The probability that exactly 70 of them own a cell phone.
3. The probability that less than 70 of them own a cell phone.
4. The probability that between 40 and 70 of them own a cell phone.

Solution (1 of 4)

Let $n = 100$ and $p = 0.62$, then since $np(1-p) = 100(0.62)(1-0.62) = 23.56 > 10$ we may use the normal approximation to the binomial. Note that $\mu_X = n\,p = 62.0$ and $\sigma_X = \sqrt{np(1-p)} = 4.9$.

$$\begin{aligned}
P(X \geq 70) &= P(X \geq 69.5) \\
&= P\left(\frac{X-\mu_X}{\sigma_X} \leq \frac{69.5-62.0}{4.9}\right) \\
&= P(Z \geq 1.53) \\
&= 1 - P(Z < 1.53) = 1 - 0.9370 \\
P(X \geq 70) &= 0.0630
\end{aligned}$$

Solution (2 of 4)

Note that $\mu_X = 62.0$ and $\sigma_X = 4.9$.

$$\begin{aligned}
P(X = 70) &= P(69.5 < X < 70.5) \\
&= P\left(\frac{69.5 - 62.0}{4.9} < \frac{X - \mu_X}{\sigma_x} < \frac{70.5 - 62.0}{4.9}\right) \\
&= P(1.53 < Z < 1.73) \\
&= 0.9582 - 0.9370 = 0.0212
\end{aligned}$$

Solution (3 of 4)

Note that $\mu_X = 62.0$ and $\sigma_X = 4.9$.

$$\begin{aligned} P(X < 70) &= P(X < 69.5) \\ &= P\left(\frac{X - \mu_X}{\sigma_x} < \frac{69.5 - 62.0}{4.9}\right) \\ &= P(Z < 1.53) \\ &= 0.9370 \end{aligned}$$

Solution (4 of 4)

Note that $\mu_X = 62.0$ and $\sigma_X = 4.9$.

$$\begin{aligned} P(40 \leq X \leq 70) &= P(39.5 < X < 70.5) \\ &= P\left(\frac{39.5 - 62.0}{4.9} < \frac{X - \mu_X}{\sigma_X} < \frac{70.5 - 62.0}{4.9}\right) \\ &= P(-4.59 < Z < 1.73) \\ &= 0.9582 - 0.0000 \\ &= 0.9582 \end{aligned}$$

Example

Thirty-six percent of students at Millersville University were diagnosed with influenza last winter. Suppose 60 students were randomly selected. Use the normal approximation to the binomial to determine the following.

1. The probability that between 10 and 30 were diagnosed with influenza.
2. The probability that at least 15 were diagnosed with influenza.
3. The probability that no more than 20 were diagnosed with influenza.
4. The probability that exactly 22 were diagnosed with influenza.

Solution (1 of 4)

Let $n = 60$ and $p = 0.36$, then since $np(1-p) = 60(0.36)(1-0.36) = 13.824 > 10$ we may use the normal approximation to the binomial. Note that $\mu_X = np = 21.6$ and $\sigma_X = \sqrt{np(1-p)} = 3.7$.

$$\begin{aligned}
P(10 \leq X \leq 30) &= P(9.5 < X < 30.5) \\
&= P\left(\frac{9.5 - 21.6}{3.7} < \frac{X - \mu_X}{\sigma_X} < \frac{30.5 - 21.6}{3.7}\right) \\
&= P(-3.27 < Z < 2.41) \\
&= 0.9920 - 0.0005 \\
&= 0.9915
\end{aligned}$$

Solution (2 of 4)

Note that $\mu_X = 21.6$ and $\sigma_X = 3.7$.

$$\begin{aligned} P(X \geq 15) &= P(X > 14.5) \\ &= P\left(\frac{X - \mu_X}{\sigma_x} < \frac{14.5 - 21.6}{3.7}\right) \\ &= P(Z > -1.92) \\ &= 1 - 0.0274 \\ &= 0.9726 \end{aligned}$$

Solution (3 of 4)

Note that $\mu_X = 21.6$ and $\sigma_X = 3.7$.

$$\begin{aligned} P(X \leq 20) &= P(X < 20.5) \\ &= P\left(\frac{X - \mu_X}{\sigma_x} < \frac{20.5 - 21.6}{3.7}\right) \\ &= P(Z < -0.30) \\ &= 0.3821 \end{aligned}$$

Solution (4 of 4)

Note that $\mu_X = 21.6$ and $\sigma_X = 3.7$.

$$\begin{aligned} P(X = 22) &= P(21.5 < X < 22.5) \\ &= P\left(\frac{21.5 - 21.6}{3.7} < \frac{X - \mu_X}{\sigma_x} < \frac{22.5 - 21.6}{3.7}\right) \\ &= P(-0.03 < Z < 0.24) \\ &= 0.5948 - 0.4880 \\ &= 0.1068 \end{aligned}$$