Probability and Statistics

Topic 17 - The Normal Approximation to the Binomial Probability Distribution

Aamir Alaud Din, PhD

November 15, 2024

TABLE OF CONTENTS

TABLE OF CONTENTS

RECAP

- The normality of a data set is assessed using normal probability plot.
- For normal probability plot, the data must be arranged in ascending order.
- Normal probability plot is the plot of expected z-scores against the corresponding data points.
- If a strong correlation is established between the z-scores against the corresponding data points, we say that the data is normally distributed.

TABLE OF CONTENTS

OBJECTIVES

After learning this topic and studying, you should be able to:

1. Approximate binomial probabilities using the normal distribution

TABLE OF CONTENTS

THE WHY SECTION

- We already discussed the binomial probability distribution.
- Now, we will review the criteria for a probability experiment to be a binomial experiment.
- The only point to know why we want to use this approximation is a single point.
- The point is to use the normal probability plot or data.
- Our focus is to use the normal probability data table for binomial probabilities.

TABLE OF CONTENTS

APPROXIMATING BINOMIAL PROBABILITIES USING THE NORMAL DISTRIBUTION

Criteria for a Binomial Probability Experiment

A probability experiment is a binomial experiment if all the following are true:

1. The experiment is performed n independent times. Each repetition of the experiment is called a trial. Independence means that the outcome of one trial will not affect the outcome of the other trials.
2. For each trial, there are two mutually exclusive outcomes—success or failure.
3. The probability of success, p, is the same for each trial of the experiment.

- The binomial probability formula can be used to compute probabilities of events in a binomial experiment.

APPROXIMATING BINOMIAL PROBABILITIES USING THE NORMAL DISTRIBUTION

- A large number of trials of a binomial experiment, however, makes this formula difficult to use.
- For example, given 500 trials of a binomial experiment, to compute the probability of 400 or more successes requires that we compute the following probabilities:

$$P(X \geq 400) = P(400) + P(401) + \cdots + P(500)$$

- This would be time consuming to compute by hand! Fortunately, we have an alternative means for approximating binomial probabilities, provided that certain conditions are met.
- Recall, the following fact:

APPROXIMATING BINOMIAL PROBABILITIES USING THE NORMAL DISTRIBUTION

"For a fixed p, as the number of trials n in a binomial experiment increases, the probability distribution of the random variable X becomes more nearly symmetric and bell shaped. As a rule of thumb, if $np(1-p) \geq 10$, the probability distribution will be approximately symmetric and bell-shaped."

- This result suggests that binomial probabilities can be approximated by the area under a normal curve, provided that $np(1-p) \geq 10$.

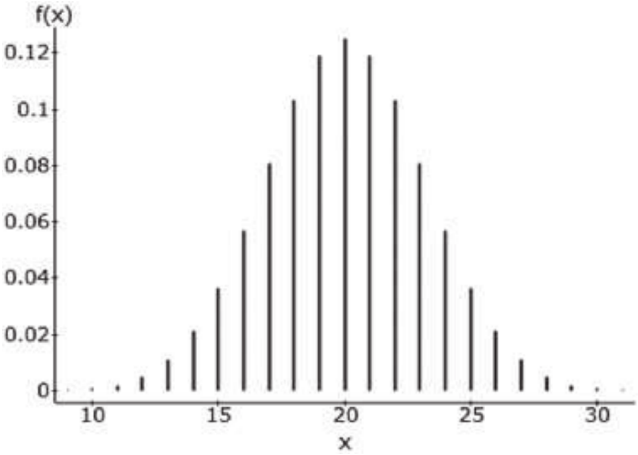
The Normal Approximation to the Binomial Probability Distribution

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

APPROXIMATING BINOMIAL PROBABILITIES USING THE NORMAL DISTRIBUTION

- Figure below shows a graph of the probability distribution for the binomial random variable X, with $\mu_x = np = 40(0.5) = 20$ and $p = 0.5$, drawn in StatCrunch.
- Because $np(1-p) = 40(0.5)(1-0.5) = 10$, we can use a normal model with $\mu_X = np = 40)0.5) = 20$ and standard deviation $\sigma_X = \sqrt{np(1-p)} = \sqrt{40(0.5)(1-0.5)} = \sqrt{10}$ to describe X.
- To approximate the probability of a specific value of the binomial random variable, such as $P(18)$, we find the area under the normal curve from $x = 17.5$ to $x = 18.5$.
- We add and subtract 0.5 from $x = 18$ as a correction for continuity because we are using a continuous density function to approximate a discrete probability.

APPROXIMATING BINOMIAL PROBABILITIES USING THE NORMAL DISTRIBUTION

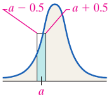
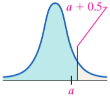
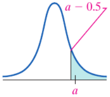
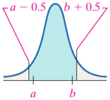


APPROXIMATING BINOMIAL PROBABILITIES USING THE NORMAL DISTRIBUTION

- Do you see why?
- To approximate $P(X \geq 18)$, we compute $P(X \geq 17.5)$.
- Do you see why?
- Table 6 summarizes how to use the correction for continuity.
- A question remains, however.
- What do we do if the probability is of the form $P(X > a)$, $P(X < a)$, or $P(a < X < b)$?
- The solution is to rewrite the inequality in a form with $\leq$ or $\geq$.

APPROXIMATING BINOMIAL PROBABILITIES USING THE NORMAL DISTRIBUTION

Table 6

Exact Probability Using Binomial	Approximate Probability Using Normal	Graphical Depiction
$P(a)$	$P(a - 0.5 \leq X \leq a + 0.5)$	
$P(X \leq a)$	$P(X \leq a + 0.5)$	
$P(X \geq a)$	$P(X \geq a - 0.5)$	
$P(a \leq X \leq b)$	$P(a - 0.5 \leq X \leq b + 0.5)$	

APPROXIMATING BINOMIAL PROBABILITIES USING THE NORMAL DISTRIBUTION

- For example, $P(X > 4) = P(X \geq 5)$ and $P(X < 4) = P(X \leq 3)$ for binomial random variables, because the values of the random variables must be whole numbers.

EXAMPLE 1

According to the American Red Cross, 7% of people in the United States have blood type O-negative. What is the probability that, in a simple random sample of 500 people in the United States, fewer than 30 have blood type O-negative?

APPROXIMATING BINOMIAL PROBABILITIES USING THE NORMAL DISTRIBUTION

EXAMPLE 2

According to the Gallup Organization, 65% of adult Americans are in favor of the death penalty for individuals convicted of murder. Erica selects a random sample of 1000 adult Americans in Will County, Illinois, and finds that 630 of them are in favor of the death penalty for individuals convicted of murder.

(a) Assuming that 65% of adult Americans in Will County are in favor of the death penalty, what is the probability of obtaining a random sample of no more than 630 adult Americans in favor of the death penalty from a sample of size 1000?

(b) Does the result from part (a) contradict the Gallup Organization's findings? Explain.

TABLE OF CONTENTS

SUMMARY

- The binomial probabilities can be approxiated by a normal distribution.
- The mean and standard deviation of the binomial probability distribution are computed the formulas already studied.
- The random variable x is approximated using a continuous probability distribution by consulting Table V as discussed in the previous topics.
- Majority of the software contain the modules/packages/options for continuous probability distributions which help in the approximation of the binomial probability distributions.


Thank You!