Chap. 6 Notes: Probability Distribution: The Essentials

J. Harner A. Billings

Department of Statistics
West Virginia University

Stat 211 Fall 2007

Outline

The Binomial Distribution

- Suppose an experiment possesses the properties;

1. There are a fixed number of trials, n.
2. Each trial results in one of only two possible outcomes:

Success and Failure

3. The probability of a success is the same for each trial.
4. The trials are independent of one another.
5. X = number of Successes.

Such an experiment is called a **Binomial eaperiment**.

Example:

determine if the following is a binomial experiment:

A fair coin is tossed 25 times and the number of Heads is observed.

1. n = 25 trials.
2. Success = Heads.
   Failure = Tails.
3. p = P(Success) = P(H) = $\frac{1}{2}$ is constant.
4. Tosses are independent.
5. X = number of heads.

General Binomial Distribution:

- Suppose X counts the number of successes in a binomial experiment consisting of n trials.
  Then X follows a binomial distribution.

Notation:

$$X \sim B(n, p)$$

- B stands for binomial distribution.
- p = P(Success on a single trial.)

- Unless n is small (n $\leq$ 5), it is inefficient to use probability trees.

  Instead, we can use the binomial distribution formula to calculate binomial probabilities.

$$P(X) = \binom{n}{x} . p^X . (1-p)^{n-X}$$

for X = 0, 1, 2, . . ., n.

- P(X) stands for "the theoretical probability of X Successes in n trials."

- $$\binom{n}{X} = \frac{n!}{X!.(n-X)!}$$ tells the number of different ways we can obtain X Successes in n trials.

- p = P(Success on a single trial)
  X= 0, 1, 2, . . ., n. means
  we can use this formula for each of these values of X.

Calculating Binomial Probabilities:

Example:

A couple plans to have 2 children.
Arbitrarily label "Male" as a success. Then

$$p = P(Success) = 0.5$$

(Recall this is a binomial experiment)

a. Find the probability that the couple has no male children.
(Here, X = 0).

-

$$\begin{aligned} P(0) &= \binom{2}{0}.(0.5)^0.(1-5)^{2-0} \\ &= \frac{2!}{0!.(2-0)!}.(1).(0.5)^2 \\ &= \frac{2.1}{1.(2.1)}.(0.25) \end{aligned}$$

(contd.)

$$\begin{aligned} P(0) &= 1.(0.25) \\ &= 0.25 \end{aligned}$$

b. Find the probability that the couple has 1 Male child.
(X = 1).

-

$$\begin{aligned} P(1) &= \binom{2}{1}.(0.5)^1.(1-5)^{2-1} \\ &= \frac{2!}{1!.(2-1)!}.(0.5).(0.5)^1 \\ &= \frac{2.1}{1.1}.(0.25) \\ &= 2.(0.25) \\ &= 0.50 \end{aligned}$$

c. Find the probability that the couple has 2 Male children.
(X = 2)

-

$$
\begin{aligned}
P(1) &= \binom{2}{2}.(0.5)^2.(1-5)^{2-2} \\
&= \frac{2!}{2!.0!}.(0.25).(0.5)^0 \\
&= \frac{2.1}{(2.1).1}.(0.25).(1) \\
&= 2.(0.25) \\
&= 0.25
\end{aligned}
$$

Binomial Probabilities Using Tables:

- **Table G** in the Text Book[1], gives cumulative binomial probabilities for certain values of n and p:

$$P(X \leq r)$$

Example:

P(X ≤ 8) for n = 20, p = 0.3

$$P(X \leq 8) = 0.8867$$

Example:

n = 10 p = 0.4

$$P(X \leq 5) = 0.8338$$

[1] refer to the Page: 702 of the Text Book

Cumulative Binomial Probabilities:

- What does P(X $\leq$ 7) mean?

$$\begin{aligned} P(X \leq 7) &= P(X = 0 \; or X = 1 \; or \ldots or \; X = 7) \\ &= P(X = 0) + P(X = 1) + \ldots + P(X = 7) \end{aligned}$$

(by Special Addition Rule)

- We can use the **cumulative binomial Table G** to find the simple binomial probabilities.

Example:

$$P(X = 6) = P(x \leq 6) - P(x \leq 5)$$

(P(x$\leq$ 6) and P(x$\leq$ 5) can be obtained from the **Table G** of the Text Book[2]).

[2]refer to the Page: 702 of the Text Book

- Why does this work?

$$P(X \leq 6) - P(X \leq 5)$$

- [P(X=0)+ P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5) + P(X=6)] - [P(X=0)+ P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)]

- Removing the square brackets and combining similar terms yields

$$P(X = 6)$$

Example:

$$n = 10 \qquad p = 0.5$$

$$\begin{aligned} P(X = 6) &= P(X \leq 6) - P(X \leq 5) \\ &= 0.8281 - 0.6230 \\ &= 0.251 \end{aligned}$$

Example:

$$n = 20 \qquad p = 0.4$$

$$\begin{aligned} P(X > 10) &= 1 - P(X \leq 10) \\ &= 1 - 0.8725 \\ &= 0.1275 \end{aligned}$$

- The previous example uses the idea of complementary events;

$$P(E) = 1P(\text{not E})$$

where,

$$\begin{aligned} E &= \{X > 10\} \\ &= \{X = 11, 12, 13, \ldots, \text{or } 20\} \end{aligned}$$

and

$$\begin{aligned} \text{not E} &= \{X \leq 10\} \\ &= \{X = 0, 1, 2, \ldots, \text{or } 10\} \end{aligned}$$

- Similarly, for the next example,

$$n = 10 \qquad p = 0.3$$

$$\begin{aligned} P(X \geq 2) &= P(X > 10) \\ &= 1 - P(X \leq 1) \\ &= 1 - 0.1493 \\ &= 0.8507 \end{aligned}$$

- Other Examples are in Sec. 6.1 of the Text Book.

- What if p > 0.5?
  - Change the roles of Success and failure.

Example:

70% of male students eventually marry. A random sample of 25 male students is selected.
Find the probability that at least 20 of these students will eventually marry.

- Success = Marry

p = P(Success) = P(Marry) = 0.7

$$\begin{aligned} X &= \text{number of Successes} \\ &= \text{Marry} \end{aligned}$$

Find $P(X \geq 20)$

Can't use Table. G since p = 0.7

(Contd.)

- Reverse roles of Success and Failure:
  Success = Not marry

$$p = \text{P(Success)} = \text{P(not marry)} = 0.3$$

$$Y = \text{number of Successes} = \text{number of not Marry}$$

(Contd.)

Then,

$$\begin{aligned} P(X \geq 20) &= P(Y \leq 5) \\ &= 0.1935 \end{aligned}$$

Value of P(Y≤5) is obtained, using **Table G** with n = 25 , p = 0.3 and r = 5.

Example:

Example 6.7 in the Text Book[3].

[3]refer to Page: 324 of the Text Book

- **Recall:** sample mean $\bar{X}$ estimates the population mean μ. The sample standard deviation s estimates the population standard deviation σ.

- **IF** we know the distribution of random variable X, we can calculate μ and σ directly - no need to estimate μ and σ.

Mean and Standard Deviation of a Binomial Distribution:

$$
\begin{aligned}
E(X) = \text{mean} &= \mu = n \times p \\
\text{standard deviation} &= \sigma = \sqrt{n \times p \times (1-p)}
\end{aligned}
$$

Example:

Couple with 2 children (Contd.)

$$n = 2 \quad p = 0.5$$

$$\mu = n.p = 2.(0.5) = 1$$

i.e, the mean number of Males in a family of two children is 1 male.

$$
\begin{aligned}
\sigma &= \sqrt{n.p.(1-p)} = \sqrt{2.(0.5).(1-0.5)} \\
&= \sqrt{2.(0.5).(0.5)} = \sqrt{0.5} \\
&\simeq 0.7
\end{aligned}
$$

Geometric Distribution:

- Geometric Distribution is concerned with the number of trials that must be conducted until a specific outcome is observed.

Example

Toss thumbtack (repeatedly).
Thumbtack can land with the spindle pointing up (u) or with spindle pointing down (D).
Suppose P(D) = p.
So, P(U) = 1- p.
We will toss the thumbtack repeatedly until the first tie the tack lands with the spindle down(D).

(Contd.)

a. what is the Sample Space?

$$\mathrm{SS} = \{D, UD, UUD, UUUD, \ldots\}$$

b. Let random variable X count the number of tosses.
Construct the probability distribution for X.

$$P(X = 1) = P(D) = p$$

$$\begin{aligned} P(X = 2) = P(UD) &= P(U).P(D) \\ &= (1 - p).p \end{aligned}$$

(Contd.)

$$\begin{aligned} P(X=3) &= P(UUD) \\ &= P(U).P(U).P(D) \\ &= (1-p).(1-p).p \end{aligned}$$

$$\begin{aligned} P(X=4) &= P(UUUD) \\ &= P(U).P(U).P(U).P(D) \\ &= (1-p).(1-p).(1-p).p \end{aligned}$$

- Generalizing, we can get the probability density function (p.d.f) of X:

$$P(X=x) = (1-p)^{x-1}.p$$

for x = 1,2,3,4, . . .

- **Geometric Distribution:**

- Suppose an experiment

a. consists of a sequence of independent trials.

b. each trial results in one of only two possible outcomes;

Success or failure

3. p = P(Success) is the the same on each trial.

4. random variable X counts the number of trials it takes to observe the first Success.

- Then x has a Geometric distribution with,

$$P(X = x) = (1 - p)^{x-1}.p$$

for x = 1,2,3,4, . . .

(Contd.)

- Also,

$$\begin{aligned} \mu_X &= \frac{1}{p} \\ \sigma_X &= \sqrt{\frac{1-p}{p^2}} \end{aligned}$$

Example:

Toss fair coin until H appears. X counts the number of tosses.

a. Find the probability that H appears on first toss.

$$\begin{aligned} P(X=1) &= (1-0.5)^{1-1}.(0.5) \\ &= (0.5)^0.(0.5) \\ &= 1.(0.5) \\ &= 0.5 \end{aligned}$$

b. Find the probability that first H appears on second toss.

$$\begin{aligned} P(X=2) &= (1-0.5)^{2-1}.(0.5) \\ &= (0.5)^1.(0.5) \\ &= 0.25 \end{aligned}$$

c. Find the probability that first H appears on 5^{th} toss.

$$\begin{aligned} P(X=5) &= (1-0.5)^{5-1}.(0.5) \\ &= (0.5)^4.(0.5) \\ &= (0.0625).(0.5) \\ &= 0.03125 \end{aligned}$$

d. What is the expected number of tosses for the first H to appear?

$$\mu_X = \frac{1}{p} = \frac{1}{0.5} = 2 = \mathrm{E(X)}$$

- Using similar reasoning, we can find the probability that it takes more than χ trials for the first Success to appear:

$$
\begin{aligned}
P(X > \chi) &= \text{P(first } \chi \text{ trials result in failure)} \\
&= \text{P(first trial Failure)} \times \\
&\quad \text{P(second trial Failure)} \times \\
&\quad \vdots \\
&\quad \text{P(}\chi^{th}\text{ trial Failure)} \\
&= (1 - p)^{\chi}
\end{aligned}
$$

Example: **Coin example** (Contd.)

c. Find the probability that more than 4 tosses will be required to get the first H.

$$\begin{aligned} P(X > 4) &= (1 - 0.5)^4 \\ &= (0.5)^4 \\ &= 0.0625 \end{aligned}$$

f. Find the probability that first H appears before the 6^{th} toss.

$$\begin{aligned} P(X < 6) &= P(X \leq 5) \\ &= 1 - P(X > 5) \\ &= 1 - (1 - 0.5)^5 \\ &= 1 - (0.5)^5 \\ &= 1 - 0.03125 \\ &= 0.96875 \end{aligned}$$

g. Find the probability that no more than 5 tosses are required to observe the first H.

$$\begin{aligned} P(X \leq 5) &= 1 - P(X > 5) \\ &= 1 - (0.5)^5 \\ &= 0.96875 \end{aligned}$$

h. Find the probability that at least 3 tosses will be required to get first H.

"At least 3" means **"3 or more"**

$$\begin{aligned} P(X \geq 3) &= P(X > 2) \\ &= (1 - 0.5)^2 \\ &= (0.5)^2 \\ &= 0.25 \end{aligned}$$

Poisson Distribution:

- Poisson distribution is concerned with **"the number of random events per unit time, or space, or both."**

- **Note:** Here **"random"** implies that there is a constant probability that the event will occur in one unit of time or space.

Poisson Random Variables:

Examples:

1. The number of people arriving at a check-out lane in Giant Eagle in a 5 minute interval.
2. The number of earthworms in one cubic yard of dirt.
3. The number of grass sprouts appearing in one week of a seeded one square yard plot of land.

General Poisson Distribution:

- The Poisson Distribution is an appropriate model for experiments in which:

1. random variable X counts the number of occurrences of some event of interest in a unit of time or space (or both).
2. the events occur randomly.
3. the mean number of events per unit of time/space is constant.
4. random variable X has no fixed upper limit.

- The the random variable X possesses a **Poisson Distribution** and we can use the General Poisson Distribution formula to compute probabilities.

- suppose λ is the mean number of events per unit time or space, and X denotes the number of possible events per unit time or space. Then

$$P(X) = \frac{e^{-\lambda}.\lambda^X}{X!}$$

$$\text{for } X = 0, 1, 2, \ldots$$

- **Note:**

$$e \simeq 2.7182818\ldots$$

Calculating Poisson Probabilities:

Example:

Suppose arrivals at a check-out line "average" 2 persons each 5 minutes.

a. Find the probability that no people arrive at the check-out line in the next 5 minutes

$$\begin{aligned} \text{Unit of time} &= 5 \text{ minutes} \\ \lambda &= 2 \\ X &= 0 \end{aligned}$$

So,

$$P(X) =$$

$$\text{for } X = 0, 1, 2, \ldots$$

(Contd.)

So,

$$\begin{aligned} P(0) &= \frac{e^{-2}.2^0}{0!} \\ &= \frac{(0.1353).(1)}{1} \\ &= 0.1353 \end{aligned}$$

b. Find the probability that 3 people arrive at the check-out line in the next 5 minutes.

$$\begin{aligned} P(X=3) &= \frac{e^{-2}.2^3}{3!} \\ &= \frac{(0.1353).(8)}{6} \\ &= 0.1804 \end{aligned}$$

c. Find the probability that 4 people arrive at check-out in the next 15 minutes

$$\begin{aligned} \text{time unit} &= 15 \text{ minutes} \\ \frac{2 \text{ persons}}{5 \text{ minutes}} &= \frac{6 \text{ people}}{15 \text{ minutes}} \end{aligned}$$

So,

$$\begin{aligned} \lambda &= 6 \\ \textit{and}, \quad X &= 4 \end{aligned}$$

$$\begin{aligned} P(4) &= \frac{e^{-6}.6^4}{4!} \\ &= \frac{(0.00247875).(1296)}{24} \\ &= 0.13385 \end{aligned}$$

Mean, Standard Deviation of Poisson Distribution

$$\begin{aligned} \text{mean} = \mu &= \lambda \\ \text{standard deviation} = \sigma &= \sqrt{\lambda} \end{aligned}$$

Example:

$$\begin{aligned} \lambda = 5 &= \mu \\ \sigma = \sqrt{5} &= 2.236 \end{aligned}$$

Poisson Probability Tables:

- **Table J** in the Textbook[4] gives the Cumulative Poisson probabilities of λ:

$$P(X \leq r)$$

Example:

$$\begin{aligned} \lambda &= 4 \\ P(X \leq 3) &= 0.4335 \\ P(X \leq 6) &= 0.8896 \end{aligned}$$

[4] refer to Page:714 of the Text Book

Example:

$\lambda = 3$

$$\begin{aligned} P(X=2) &= P(X \le 2) - P(X \le 1) \\ &= 0.423 - 0.199 \\ &= 0.224 \end{aligned}$$

$$\begin{aligned} P(X \ge 2) &= 1 - P(X < 8) \\ &= 1 - P(X \le 7) \\ &= 1 - 0.988 \\ &= 0.012 \end{aligned}$$

- Examples 6.11 - 6.17 of the Text book[5].

[5] refer to Page:335 to Page:339 of the Text Book

Poisson Approximation of the Binomial Distribution:

- When n is large and p is small, we can use the **Poisson distribution** to approximate the **Binomial probabilities**.
- Just let

$$\lambda = n \times p$$

  and use Poisson distribution.

- The conditions which must be satisfied

  1. p$\leq$ 0.1
  2. n$\geq$ 25
  3. np$\leq$ 10

Example:

Approximately 10% of people are left-handed (Success). Find the probability that 5 or fewer left-handed people will be found in a sample (random sample) of 100 people.

First note that this is a Binomial Experiment with $n = 100$ and $p = 0.1$.
(Check the 4 conditions.)

X = number of Successes = number of left-hander.

But Binomial table G does not contain probabilities for $n = 100$.

(Contd.)

Let approximate the requested probability using the Poisson distribution:

$$\lambda = np = 100.(0.10 = 10$$

Then,

$$\begin{array}{ccc} P(X \leq 5) & \simeq & P(Y \leq 5) \\ \textit{binomial} & & \textit{poisson} \\ & = & 0.67 \end{array}$$

Example:

Example 6.13 of Text Book[6].

[6]refer to Page:335 of the Text Book

Normal Distribution:

- Important in Statistics:
  1. Some variables are Normally distributed.
  2. Normal distributions have "nice" properties useful in statistical inference.

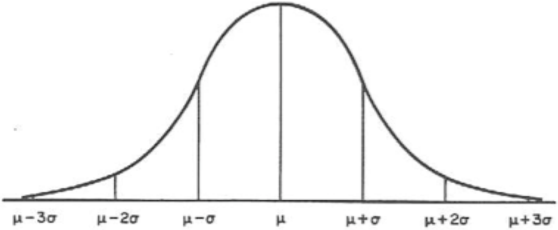


Fig. 7.4 The normal distribution: μ is the mean of the distribution, and σ is the standard deviation of the distribution.

Properties:

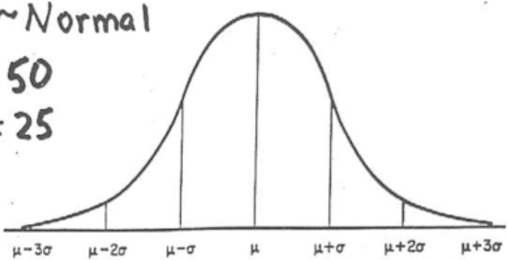
1. Mound-shaped (Bell-shaped).
2. Symmetric about μ, the population mean.
3. Continuous.
4. Total area beneath Normal curve is 1.
5. Infinite number of Normal distributions, each with its own μ and σ.

EX: X ~ Normal

with $\mu = 50$

and $\sigma^2 = 25$

So $\sigma =$

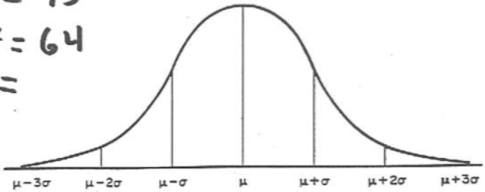


EX: X ~ Normal

with $\mu = 75$

and $\sigma^2 = 64$

So $\sigma =$



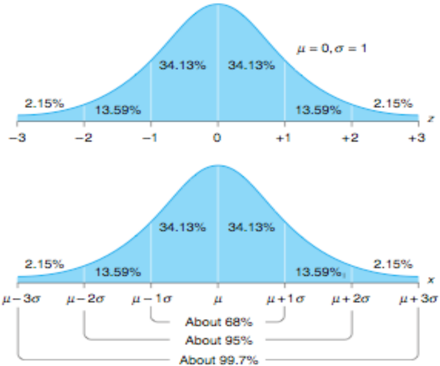


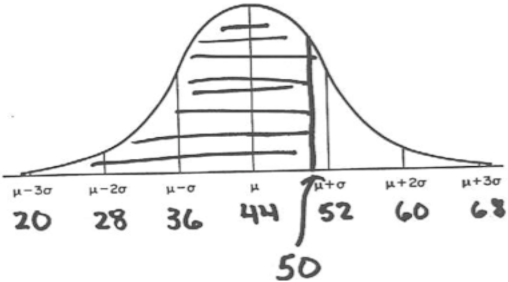
Figure 6.12 Areas under the standard normal curve derived from the normal table and the corresponding areas under the normal curve with mean μ and standard deviation σ.

$\simeq$ 68% of data values will be between $\mu - \sigma$ and $\mu + \sigma$

$\simeq$ 95% of data values will be between $\mu - \sigma$ and $\mu + \sigma$

$\simeq$ 99.7% of data values will be between $\mu - \sigma$ and $\mu + \sigma$

Example:



Suppose X, the weight of a full grown dog is normally distributed with μ= 44 lbs and σ= 8 lbs.
Find the probability that a randomly selected dog weighs less than 50 lbs.

$$P(X < 50) = \text{Area beneath the curve to the left of 50.}$$

- It is impossible to construct tables for every Normal Distribution.
- We construct a table for a very special Normal Distribution, called the**Standard Normal Distribution**($\mu = 0, \sigma = 1$).
- Then translate the original probability question into an equivalent question involving the Standard Normal Distribution.

$$Z = \frac{X - U}{\sigma}$$

- **Notation:** N(0,1) is the Standard Normal distribution.

- We usually let Z represents a variable which has a Standard Normal Distribution.

- **Note:** Table E of the Text Book[7] gives cumulative standard normal probabilities.

[7] refer to Page:699 of the Text Book

Example:

Example of Dog's weight (Contd.)

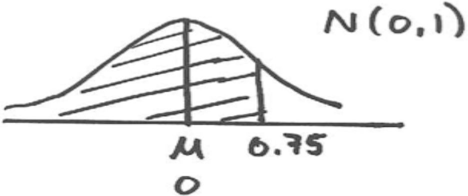
$$\mu = 44 \text{ lbs} \qquad \sigma = 8 \text{ lbs}$$

Then,

$$\begin{aligned} Z &= \frac{X - U}{\sigma} \\ &= \frac{50 - 44}{8} \\ &= 8 \end{aligned}$$

Therefore,

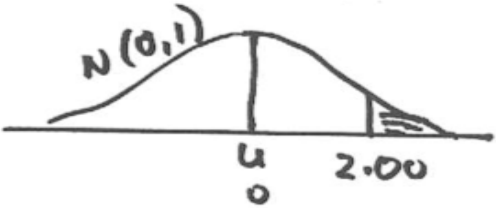
$$\begin{aligned} P(X < 50) &= P(Z < 0.75) \\ &= 0.7734 \end{aligned}$$



(Contd.)

Find P(X>60)

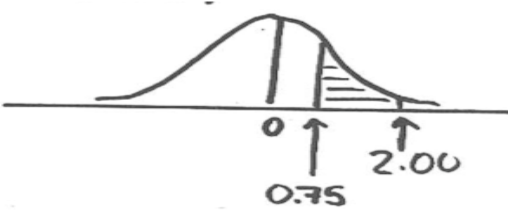
$$\begin{aligned} Z &= \frac{X - U}{\sigma} \\ &= \frac{60 - 44}{8} \\ &= 2 \end{aligned}$$



$$\begin{aligned} P(X > 60) &= P(Z > 2.00) \\ &= 1 - P(Z \leq 2.00) \\ &= 1 - 0.9772 \\ &= 0.0228 \end{aligned}$$

(Contd.)

Find P(50<X<60)



$$
\begin{aligned}
P(50 < X < 60) &= P(0.75 < Z < 2.00) \\
&= P(Z < 2.00) - P(Z < 0.75) \\
&= 0.9772 - 0.7734 \\
&= 0.2038
\end{aligned}
$$

(Contd.)

Find P(X<40)

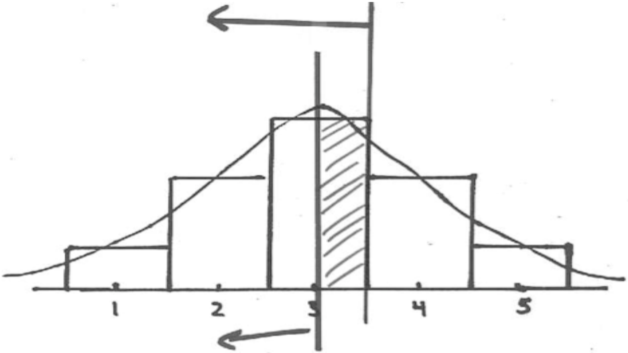
$$
\begin{aligned}
Z &= \frac{X - U}{\sigma} \\
&= \frac{40 - 44}{8} \\
&= -0.50
\end{aligned}
$$

$$
\begin{aligned}
P(X < 40) &= P(Z < -0.50) \\
&= 1 - P(Z \leq 2.00) \\
&= 0.3085
\end{aligned}
$$

Normal Approximation of the Binomial Distribution:

- We can use the **Normal Distribution** to approximate **Binomial probabilities**
  If:
  1. np>5
  2. n(1-p)>5

- **Note:** Since the Normal Distribution is continuous, but the Binomial Distribution is discrete, we must use the **"Continuity Correction"**.



Binomial		Normal without cont. correction
$P(X \leq 3)$	$\simeq$	$P(Y \leq 3)$
		(it is not correct)
		Normal with cont. correction
	$\simeq$	$P(Y \leq 3.5)$
		(it is correct)

Example:

Suppose n = 100 and p = 0.6 in a Binomial experiment.

Find P(X≤70)

- **Note:** Binomial Table does not have n = 100.
  - Can we use the Normal Approximation?

    Check

$$\begin{aligned} n.p &= 100.(0.6) \\ &= 60 > 5 \end{aligned}$$

and

$$\begin{aligned} n.(1-p) &= 100.(1-0.6) \\ &= 100.(0.4) \\ &= 40 > 5 \end{aligned}$$

  - **Yes!**

(Contd.)

First we need,

$$\mu = n.p = 100.(0.6) = 60$$

and

$$\begin{aligned}\sigma &= \sqrt{n.p.(1-p)} \\ &= \sqrt{100.(0.6).(1-0.6)} \\ &= \sqrt{24} \\ &= 4.899\end{aligned}$$

Therefore

$$\begin{aligned}Z &= \frac{X-U}{\sigma} \\ &= \frac{70.5-60}{4.899} \\ &= 2.14\end{aligned}$$

Now consider,

$$\begin{array}{ccl} P(X \leq 70) & \simeq & P(X \leq 70.5) \\ \text{Binomial} & & \text{Normal(with cont. corr.)} \\ \text{(discrete)} & & \text{(continuous)} \\ & = & P(Z \leq 2.14) \\ & & \text{Standard Normal} \\ & = & 0.9838 \end{array}$$

Example:

Example 6.23 of the Text Book[8].

[8] refer to Page: 353 of the Text Book.