Larson and Marx Section 3.2

Gene Quinn

The Binomial Probability Distribution

Binomial probabilities arise in the following situation:

- A series of identical and independent trials is conducted
- Each trial has exactly two possible outcomes, success or failure
- The probability p of success is the same for each trial

The Binomial Probability Distribution

Binomial probabilities arise in the following situation:

- A series of identical and independent trials is conducted
- Each trial has exactly two possible outcomes, success or failure
- The probability p of success is the same for each trial

We define a random variable X on the sample space consisting of outcomes of the series.

The value of the random variable X will be the number of successes obtained.

The Binomial Probability Distribution

Suppose our series consists of 4 trials. We will define a random variable X to be the number of successes in 4 trials. The sample space and the number of successes x_i is:

outcomes	x_i
$\{f,f,f,f\}$	0
$\{s,f,f,f\},\{f,s,f,f\},\{f,f,s,f\},\{f,f,f,s\}$	1
$\{s,s,f,f\},\{s,f,s,f\},\{s,f,f,s\},\{f,s,s,f\},\{f,s,f,s\},\{f,f,s,s\}$	2
$\{s,s,s,f\},\{s,s,f,s\},\{s,f,s,s\},\{f,s,s,s\}$	3
$\{f,f,f,f\}$	4

The Binomial Probability Distribution

We now assign a probability to each value x_i of the random variable X:

x_i	$P(X = x_i)$
0	$\binom{4}{0}p^0(1-p)^4$
1	$\binom{4}{1}p(1-p)^3$
2	$\binom{4}{2}p^2(1-p)^2$
3	$\binom{4}{3}p^3(1-p)$
4	$\binom{4}{4}p^4(1-p)^0$

The Binomial Probability Distribution

The probability of any sequence containing exactly k successes is:

$$p^k(1-p)^{4-k}$$

The Binomial Probability Distribution

The probability of any sequence containing exactly k successes is:

$$p^k(1-p)^{4-k}$$

The number of sequences (outcomes) with k successes is:

$$\binom{4}{k}$$

The Binomial Probability Distribution

The probability of any sequence containing exactly k successes is:

$$p^k(1-p)^{4-k}$$

The number of sequences (outcomes) with k successes is:

$$\binom{4}{k}$$

So the probability that the random variable X assumes the value k is:

$$\binom{4}{k}p^k(1-p)^{4-k}, \quad k = 0, 1, 2, 3, 4$$

The Binomial Probability Distribution

Theorem (3.2.1)

Suppose a series of n independent trials is conducted, each with exactly two possible outcomes, "success" or "failure".

The Binomial Probability Distribution

Theorem (3.2.1)

Suppose a series of n independent trials is conducted, each with exactly two possible outcomes, "success" or "failure".

Let

$$p = P(\text{success occurs on any given trial})$$

and assume that p remains constant from trial to trial.

The Binomial Probability Distribution

Theorem (3.2.1)

Suppose a series of n independent trials is conducted, each with exactly two possible outcomes, "success" or "failure".

Let

$$p = P(\text{success occurs on any given trial})$$

and assume that p remains constant from trial to trial.

Then

$$P(k\text{successes}) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

The Binomial Probability Distribution

Theorem (3.2.1)

Suppose a series of n independent trials is conducted, each with exactly two possible outcomes, "success" or "failure".

Let

$$p = P(\text{success occurs on any given trial})$$

and assume that p remains constant from trial to trial.

Then

$$P(k\text{successes}) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

This probability assignment is called the **binomial distribution**

The Binomial Probability Distribution

Example

A traveller has to take three flights and a train to reach his remote destination.

The schedules allow 20 minutes for each connection.

If 4 percent of flights are more than 20 minutes late, what is the probability that the traveller arrives in time to board the train?

The Binomial Probability Distribution

This is a binomial model, and if the random variable X represents the number of arrivals within 20 minutes of the scheduled time. If any of the three planes is 20 minutes or more late, the traveller will miss the train, so the probability we want is

$$P(X = 3)$$

The Binomial Probability Distribution

This is a binomial model, and if the random variable X represents the number of arrivals within 20 minutes of the scheduled time. If any of the three planes is 20 minutes or more late, the traveller will miss the train, so the probability we want is

$$P(X = 3)$$

$$P(X = 3) = \binom{3}{0}(.96)^3(.04)^0 = .96^3 = 0.885$$

So, the traveller has about an eleven percent chance of missing the train.

The Binomial Probability Distribution

Suppose on a given day the Yankees have a 60 percent chance of beating the Red Sox. If the two teams meet in a best of seven series, what is the probability that the Red Sox win the series?

The Binomial Probability Distribution

Suppose on a given day the Yankees have a 60 percent chance of beating the Red Sox. If the two teams meet in a best of seven series, what is the probability that the Red Sox win the series?

We will use a binomial model with "success" defined as a Red Sox win and $p = .4$.

For the Red Sox to win in a sweep, they have to win 4 out of 4 games:

$$P(X = 4) = \binom{4}{4}(.4)^4(.6)^0 = .4^4 = .0256$$

So, the chances of a sweep are small. We need to consider longer series as well.

The Binomial Probability Distribution

Now consider the probability that the Red Sox win a 5 game series.

We consider this as two events:

First, the Red Sox win exactly 3 of the first 4 games.

Using a binomial probability distribution, the probability of winning 3 of the first 4 games is:

$$P(X = 3) = \binom{4}{3}(.4)^3(.6) = 4 \cdot (.4)^3 \cdot (.6) = .154$$

The Binomial Probability Distribution

Now consider the probability that the Red Sox win a 5 game series.

We consider this as two events:

First, the Red Sox win exactly 3 of the first 4 games.

Using a binomial probability distribution, the probability of winning 3 of the first 4 games is:

$$P(X = 3) = \binom{4}{3}(.4)^3(.6) = 4 \cdot (.4)^3 \cdot (.6) = .154$$

Now, they have to win game 5, which occurs with probability $.4$, so assuming independence, the probability of the Red Sox winning a 5 game series is

$$(.154) \cdot (.4) = .0614$$

The Binomial Probability Distribution

Now consider the probability that the Red Sox win a 6 game series.

We consider this as two events:

First, the Red Sox win exactly 3 of the first 5 games.

Using a binomial probability distribution, the probability of winning 3 of the first 5 games is:

$$P(X = 3) = \binom{5}{3}(.4)^3(.6)^2 = 10 \cdot (.4)^3 \cdot (.6)^2 = .230$$

The Binomial Probability Distribution

Now consider the probability that the Red Sox win a 6 game series.

We consider this as two events:

First, the Red Sox win exactly 3 of the first 5 games.

Using a binomial probability distribution, the probability of winning 3 of the first 5 games is:

$$P(X = 3) = \binom{5}{3}(.4)^3(.6)^2 = 10 \cdot (.4)^3 \cdot (.6)^2 = .230$$

Now, they have to win game 6, which occurs with probability $.4$, so assuming independence, the probability of the Red Sox winning a 6 game series is

$$(.230) \cdot (.4) = .0922$$

The Binomial Probability Distribution

Finally, consider the probability that the Red Sox win a 7 game series.

We consider this as two events:

First, the Red Sox win exactly 3 of the first 6 games.

Using a binomial probability distribution, the probability of winning 3 of the first 6 games is:

$$P(X = 3) = \binom{6}{3}(.4)^3(.6)^3 = 20 \cdot (.4)^3 \cdot (.6)^3 = .276$$

The Binomial Probability Distribution

Finally, consider the probability that the Red Sox win a 7 game series.

We consider this as two events:

First, the Red Sox win exactly 3 of the first 6 games.

Using a binomial probability distribution, the probability of winning 3 of the first 6 games is:

$$P(X=3) = \binom{6}{3}(.4)^3(.6)^3 = 20 \cdot (.4)^3 \cdot (.6)^3 = .276$$

Now, they have to win game 7, which occurs with probability $.4$, so assuming independence, the probability of the Red Sox winning a 7 game series is

$$(.276) \cdot (.4) = .111$$

The Binomial Probability Distribution

Since Red Sox series wins in 4, 5, 6, and 7 games are mutually exclusive, we can add the probabilities of these events to get the probability that the Red Sox win the series regardless of how many games are played:

P(Red Sox win series)=.0256+.0614+.0922+.111=.290

The Hypergeometric Probability Distribution

The binomial distribution arises if we conduct repeated trials with a constant probability p of success.

We can model this as an urn problem, where chips are drawn from an urn with proportion p of red (success) chips, and $1-p$ of white (failure) chips.

The Hypergeometric Probability Distribution

The binomial distribution arises if we conduct repeated trials with a constant probability p of success.

We can model this as an urn problem, where chips are drawn from an urn with proportion p of red (success) chips, and $1 - p$ of white (failure) chips.

In order to keep p constant, we have to return each chip to the urn after it is drawn.

This is called "sampling with replacement"

The Hypergeometric Probability Distribution

It is also possible to sample without replacement.

In sampling without replacement, the chip drawn is **not** returned to the urn.

The Hypergeometric Probability Distribution

It is also possible to sample without replacement.

In sampling without replacement, the chip drawn is **not** returned to the urn.

If we draw n chips without replacement, and define a random variable X as the number of red chips drawn, the resulting probabilities follow a **hypergeometric distribution**.

The Hypergeometric Probability Distribution

Theorem: (3.2.2)

Suppose an urn contains r red chips, and w white chips, with $r + w = N$.

If n chips are drawn randomly without replacement, and if k is the number of red chips drawn, then

$$P(k \text{ red chips are drawn}) = \frac{\binom{r}{k}\binom{w}{n-k}}{\binom{N}{n}}$$

The Hypergeometric Probability Distribution

Theorem: (3.2.2)

Suppose an urn contains r red chips, and w white chips, with $r + w = N$.

If n chips are drawn randomly without replacement, and if k is the number of red chips drawn, then

$$P(k \text{ red chips are drawn}) = \frac{\binom{r}{k}\binom{w}{n-k}}{\binom{N}{n}}$$

In this formula, k varies over all integers for which

$$\binom{r}{k} \quad \text{and} \quad \binom{w}{n-k} \quad \text{are defined}$$

These probabilities are known as the **hypergeometric distribution**.

The Hypergeometric Probability Distribution

The Las Vegas game Keno uses a card with 80 numbers.

The player selects a subset of these, containing between 1 and 15 numbers.

The caller then draws 20 numbers without replacement from the 80 listed on the card.

The Hypergeometric Probability Distribution

The Las Vegas game Keno uses a card with 80 numbers.

The player selects a subset of these, containing between 1 and 15 numbers.

The caller then draws 20 numbers without replacement from the 80 listed on the card.

What (if anything) the player wins is determined by how many of the numbers in the subset the player selected were chosen, and how many were in the subset to begin with.

The Hypergeometric Probability Distribution

The game can be thought of as drawing some number k (with k between 1 and 15) of chips from an urn that contains 20 red chips and 60 white chips.

The Hypergeometric Probability Distribution

The game can be thought of as drawing some number k (with k between 1 and 15) of chips from an urn that contains 20 red chips and 60 white chips.

In this case, the number of "red" chips drawn corresponds to the count of numbers from the k selected in advance by the player that were drawn.

The associated probabilities will have a hypergeometric distribution.

The Hypergeometric Probability Distribution

Example: Find the probability that 3 numbers of 5 bet on are selected in a game of Keno.

The Hypergeometric Probability Distribution

Example: Find the probability that 3 numbers of 5 bet on are selected in a game of Keno.

This is equivalent to getting 3 "red" chips out of 5 drawn from an urn containing 20 red chips and 60 white chips.

$$P(3 \text{ red chips chosen}) = \frac{\binom{20}{3}\binom{60}{2}}{\binom{80}{5}}$$