STAT 234 Lecture 5
Binomial Distributions
Section 3.5

Yibi Huang
Department of Statistics
University of Chicago

Coverage

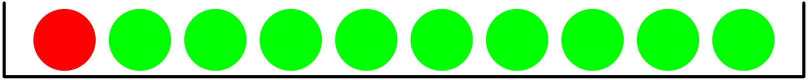
Lecture 5 covers Section 3.5 of MMSA

Please skip

- Moment General Function for Binomial on p.135 in Section 3.5
- Section 3.4 Moment Generating Functions
- Section 3.6 Hypergeometric and Negative Binomial Distributions
- Section 3.7 Poisson Distributions

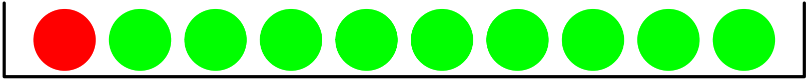
The Binomial Distribution

Five draws are made at random with replacement from a box containing one red ball and 9 green balls.



What is the probability that the first two draws are Red and the next 3 are Green?

Five draws are made at random with replacement from a box containing one red ball and 9 green balls.



What is the probability that the first two draws are Red and the next 3 are Green?

$$\begin{aligned} P(\mathsf{R\,R\,G\,G\,G}) &= P(\mathsf{R}) \cdot P(\mathsf{R}) \cdot P(\mathsf{G}) \cdot P(\mathsf{G}) \cdot P(\mathsf{G}) \\ &= 0.1 \times 0.1 \times 0.9 \times 0.9 \times 0.9. \end{aligned}$$

As the draws are made with replacement,
the outcomes of the 5 draws are **independent**.
The multiplication rule for independent events can be applied.

What is the probability of getting **exactly two Reds in 5 draws**?

Is it also equal to

$$0.1 \times 0.1 \times 0.9 \times 0.9 \times 0.9?$$

There are 10 possible orderings of the 2 Reds and the 3 Greens.

Possible Orders	Probability
RRGGG	$0.1 \times 0.1 \times 0.9 \times 0.9 \times 0.9 = (0.1)^2(0.9)^3$
RGRGG	$0.1 \times 0.9 \times 0.1 \times 0.9 \times 0.9 = (0.1)^2(0.9)^3$
RGGRG	$0.1 \times 0.9 \times 0.9 \times 0.1 \times 0.9 = (0.1)^2(0.9)^3$
RGGGR	$0.1 \times 0.9 \times 0.9 \times 0.9 \times 0.1 = (0.1)^2(0.9)^3$
GRRGG	$0.9 \times 0.1 \times 0.1 \times 0.9 \times 0.9 = (0.1)^2(0.9)^3$
GRGRG	$0.9 \times 0.1 \times 0.9 \times 0.1 \times 0.9 = (0.1)^2(0.9)^3$
GRGGR	$0.9 \times 0.1 \times 0.9 \times 0.9 \times 0.1 = (0.1)^2(0.9)^3$
GGRRG	$0.9 \times 0.9 \times 0.1 \times 0.1 \times 0.9 = (0.1)^2(0.9)^3$
GGRGR	$0.9 \times 0.9 \times 0.1 \times 0.9 \times 0.1 = (0.1)^2(0.9)^3$
GGGRR	$0.9 \times 0.9 \times 0.9 \times 0.1 \times 0.1 = (0.1)^2(0.9)^3$

P(exactly 2 Reds in 5 draws) is the sum of the probabilities of the 10 cases above because the 10 cases are *disjoint*. So

$$\mathrm{P}(\text{exactly 2 Reds in 5 draws}) = 10 \times (0.1)^2(0.9)^3.$$

What is P(getting **exactly k Reds in n draws**)?

What is P(getting **exactly k Reds in n draws**)?

Consider all possible ways to order the k Reds and the $n - k$ Greens.

Possible Orders	Probability
$\underbrace{\text{RRR}\ldots\text{R}}_{k}\,\underbrace{\text{G}\ldots\text{G}}_{n-k}$	$\underbrace{0.1 \times \ldots \times 0.1}_{k} \times \underbrace{0.9 \times \ldots \times 0.9}_{n-k} = (0.1)^k(0.9)^{n-k}$
$\text{RGR}\ldots\text{R}\,\text{G}\ldots\text{G}$	$0.1 \cdot 0.9 \cdot 0.1 \ldots 0.1 \times 0.9 \ldots 0.9 = (0.1)^k(0.9)^{n-k}$
$\vdots$	$\vdots$

Note

- the events for different orderings are *disjoint*, and
- each occurs with identical probability $(0.1)^k(0.9)^{n-k}$.

By the Addition Rule, P(exactly k Reds in n draws) equals

$$(\text{\# of ways to order } k \text{ Reds and } n-k \text{ Greens}) \times (0.1)^k(0.9)^{n-k}$$

Factorial

The notation $n!$, read n *factorial*, is defined as

$$n! = 1 \times 2 \times 3 \times \ldots \times (n-1) \times n$$

e.g.,

$$1\,! = 1, \qquad 3\,! = 1 \times 2 \times 3 = 6,$$
$$2\,! = 1 \times 2 = 2, \qquad 4\,! = 1 \times 2 \times 3 \times 4 = 24.$$

By convention,

$$0\,! = 1.$$

Binomial Coefficients

The number of ways to order k Reds and $n - k$ Greens equals

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- $\binom{n}{k}$ is read as "*n choose k*", also denoted as ${}_nC_k$, or C_k^n.

e.g.,

$$\binom{5}{2} = \frac{5\,!}{2\,! \times (5-2)\,!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{(2 \times 1)(3 \times 2 \times 1)} = \frac{5 \times 4}{2 \times 1} = 10,$$

$$\binom{n}{n} = \frac{n\,!}{n\,! \times 0\,!} = \frac{n\,!}{n\,! \times 1} = 1$$

You can also use R for these calculations:

```
choose(5,2)
[1] 10
```

Binomial Coefficients

The number of ways to order k Reds and $n-k$ Greens equals

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- $\binom{n}{k}$ is read as "*n choose k*", also denoted as ${}_nC_k$, or C_k^n.

e.g.,

$$\binom{5}{2} = \frac{5\,!}{2\,! \times (5-2)\,!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{(2 \times 1)(3 \times 2 \times 1)} = \frac{5 \times 4}{2 \times 1} = 10,$$

$$\binom{n}{n} = \frac{n\,!}{n\,! \times 0\,!} = \frac{n\,!}{n\,! \times 1} = 1$$

You can also use R for these calculations:

```
choose(5,2)
[1] 10
```

- $\binom{n}{0} =$

- $\binom{n}{n} =$

- $\binom{n}{1} =$

- $\binom{n}{n-1} =$

- $\binom{n}{0} = \frac{n\,!}{0\,! \times n\,!} = 1$

  $\Rightarrow$ there is only 1 way to order 0 Reds and n Greens

- $\binom{n}{n} =$

- $\binom{n}{1} =$

- $\binom{n}{n-1} =$

- $\binom{n}{0} = \dfrac{n\,!}{0\,! \times n\,!} = 1$
  $\Rightarrow$ there is only 1 way to order 0 Reds and n Greens
- $\binom{n}{n} = \dfrac{n\,!}{n\,! \times 0\,!} = 1$
  $\Rightarrow$ there is only 1 way to order n Reds and 0 Green
- $\binom{n}{1} =$
- $\binom{n}{n-1} =$

- $\binom{n}{0} = \dfrac{n\,!}{0\,! \times n\,!} = 1$
  $\Rightarrow$ there is only 1 way to order 0 Reds and n Greens

- $\binom{n}{n} = \dfrac{n\,!}{n\,! \times 0\,!} = 1$
  $\Rightarrow$ there is only 1 way to order n Reds and 0 Green

- $\binom{n}{1} = \dfrac{n\,!}{1\,! \times (n-1)\,!} = n$
  $\Rightarrow$ there are n ways to order 1 Red and $n-1$ Greens

- $\binom{n}{n-1} =$

- $\binom{n}{0} = \dfrac{n\,!}{0\,! \times n\,!} = 1$
  $\Rightarrow$ there is only 1 way to order 0 Reds and n Greens

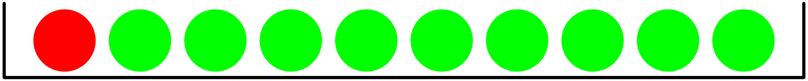
- $\binom{n}{n} = \dfrac{n\,!}{n\,! \times 0\,!} = 1$
  $\Rightarrow$ there is only 1 way to order n Reds and 0 Green

- $\binom{n}{1} = \dfrac{n\,!}{1\,! \times (n-1)\,!} = n$
  $\Rightarrow$ there are n ways to order 1 Red and $n-1$ Greens

- $\binom{n}{n-1} = \dfrac{n\,!}{(n-1)\,! \times 1\,!} = n$
  $\Rightarrow$ there are n ways to order $n-1$ Reds and 1 Green

Summary of All the Calculations So Far

When n draws are made at random with replacement from a box contains one red ball and 9 green ones,



the probability to get exactly k Reds (and $n - k$ Greens) equals

$$\binom{n}{k}(0.1)^k(0.9)^{n-k}.$$

Such calculations can be generalized to other similar problems and the general formula is called the *Binomial Formula*.

Bernoulli Trials

A random trial having only 2 possible outcomes (Success, Failure) is called a *Bernoulli trial*, e.g.,

- whether a coin lands heads or tails when tossing a coin
- whether one gets a six or not a six when rolling a die
- whether a drug works on a patient or not
- whether a electronic device is defected
- whether a subject answers Yes or No to a survey question

Binomial Formula

Suppose n **independent** Bernoulli trials are to be performed, each of which results in

- a *success* with probability p and
- a *failure* with probability $1 - p$.

The probability of getting k successes and $n - k$ failures in n Bernoulli trials is given by

$$(\text{\# of ways to order the } k \text{ successes and } n-k \text{ failures}) \times p^k(1-p)^{n-k}$$

$$= \binom{n}{k} p^k (1-p)^{n-k}$$

Binomial Distribution

Suppose n **independent** Bernoulli trials are to be performed, each of which results in

- a *success* with probability p and
- a *failure* with probability $1 - p$.

If we define

$$X = \text{the number of successes that occur in the } n \text{ trials},$$

then X is said to have a *binomial distribution* with parameters (n, p), denoted as

$$X \sim Bin(n, p).$$

with the probability mass function (pmf)

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

Does the Binomial Probabilities Add Up to 1?

Recall the *Binomial expansion* in math:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

which is valid for all real numbers a and b.

Applying the Binomial expansion with $a = p$ and $b = 1 - p$, we get

$$\sum_{k=0}^{n} \mathrm{P}(X=k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p+1-p)^n = 1^n = 1$$

which means that the pmf

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

is a valid pmf.

Conditions Required to be Binomial

Conditions required to apply the binomial formula:

1. each trial outcome must be classified as a *success* or a *failure*
2. the probability of success, p, must be the same for each trial
3. the number of trials, n, must be fixed
4. the trials must be independent

Binomial or Not? — 10 Rolls of a Die

Rolling a die 10 times, what is the probability of getting exactly 3 aces?

Binomial or Not? — 10 Rolls of a Die

Rolling a die 10 times, what is the probability of getting exactly 3 aces?

- a trial: whether one gets an ace when rolling a die once
- prob. of success $p = 1/6$
- number of trials $n = 10$

Binomial or Not? — 10 Rolls of a Die

Rolling a die 10 times, what is the probability of getting exactly 3 aces?

- a trial: whether one gets an ace when rolling a die once
- prob. of success $p = 1/6$
- number of trials $n = 10$
- the trials (rolls) are independent

So, it's okay to use the Binomial formula.

$$
\begin{aligned}
\mathrm{P}(\text{3 aces in 10 rolls}) &= \frac{10!}{3!\,7!}\left(\frac{1}{6}\right)^3\left(1-\frac{1}{6}\right)^7 \\
&= \frac{10\times 9\times 8\times(7!)}{(3\times 2\times 1)(7!)}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^7 \\
&= 120\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^7 \approx 0.155
\end{aligned}
$$

Binomial or Not? — Rolling a Die Until 3rd Ace

Rolling a die continuously, is the probability of getting the 3rd aces in the 10th roll equals to

$$\frac{10!}{3!\,7!}\left(\frac{1}{6}\right)^3\left(1-\frac{1}{6}\right)^7?$$

Binomial or Not? — Rolling a Die Until 3rd Ace

Rolling a die continuously, is the probability of getting the 3rd aces in the 10th roll equals to

$$\frac{10!}{3!\,7!}\left(\frac{1}{6}\right)^3\left(1-\frac{1}{6}\right)^7?$$

No. The number of trials (sample size) is not determined in advance.

Binomial or Not? — Restaurants

Suppose an inspector randomly selects **5** restaurantsfrom the 20 restaurants in a town, of which 10 currently have health code violation(s) and the other 10 have no violations. Let X be the number of selected restaurants with violations. Is X binomial?

- a trial: whether a randomly selected restaurant has violation(s)

$$S = \text{violation}, \quad F = \text{no violation}$$

- number of trials: $n = 5$

Binomial or Not? — Restaurants

Suppose an inspector randomly selects **5** restaurantsfrom the 20 restaurants in a town, of which 10 currently have health code violation(s) and the other 10 have no violations. Let X be the number of selected restaurants with violations. Is X binomial?

- a trial: whether a randomly selected restaurant has violation(s)

$$S = \text{violation}, \quad F = \text{no violation}$$

- number of trials: $n = 5$
- prob. of success?

Binomial or Not? — Restaurants (2)

- P(S on first trial)= $\dfrac{10}{20} = 0.5$

Binomial or Not? — Restaurants (2)

- P(S on first trial)= $\frac{10}{20} = 0.5$
- P(S on 2nd trial | S on 1st trial)= $\frac{9}{19}$

Binomial or Not? — Restaurants (2)

- P(S on first trial)= $\frac{10}{20} = 0.5$
- P(S on 2nd trial | S on 1st trial)= $\frac{9}{19}$
- P(S on 2nd trial | F on 1st trial)= $\frac{10}{19}$

Binomial or Not? — Restaurants (2)

- P(S on first trial)= $\frac{10}{20} = 0.5$
- P(S on 2nd trial | S on 1st trial)= $\frac{9}{19}$
- P(S on 2nd trial | F on 1st trial)= $\frac{10}{19}$
- P(S on 5th trial | FFFF)= $\frac{10}{16} = 0.625$

Binomial or Not? — Restaurants (2)

- P(S on first trial)= $\frac{10}{20} = 0.5$
- P(S on 2nd trial | S on 1st trial)= $\frac{9}{19}$
- P(S on 2nd trial | F on 1st trial)= $\frac{10}{19}$
- P(S on 5th trial | FFFF)= $\frac{10}{16} = 0.625$
- P(S on 5th trial | SSSS)= $\frac{6}{16} = 0.375$

Binomial or Not? — Restaurants (2)

- P(S on first trial)= $\frac{10}{20} = 0.5$
- P(S on 2nd trial | S on 1st trial)= $\frac{9}{19}$
- P(S on 2nd trial | F on 1st trial)= $\frac{10}{19}$
- P(S on 5th trial | FFFF)= $\frac{10}{16} = 0.625$
- P(S on 5th trial | SSSS)= $\frac{6}{16} = 0.375$
- Trials are NOT independent since the selection are made **without replacement**.

Binomial or Not — UC Undergrads

50 UC undergrads are randomly selected and each of them is asked whether he/she has a driver's licence. Suppose 4000 of the 5000 UC undergrads have driver's licence. Let X be the number who reply yes. Is X binomial?

Binomial or Not — UC Undergrads

50 UC undergrads are randomly selected and each of them is asked whether he/she has a driver's licence. Suppose 4000 of the 5000 UC undergrads have driver's licence. Let X be the number who reply yes. Is X binomial?

- a trial: a randomly selected student reply yes = S or no = F
- number of trials $n = 50$

Binomial or Not — UC Undergrads

50 UC undergrads are randomly selected and each of them is asked whether he/she has a driver's licence. Suppose 4000 of the 5000 UC undergrads have driver's licence. Let X be the number who reply yes. Is X binomial?

- a trial: a randomly selected student reply yes = S or no = F
- number of trials $n = 50$
- Strictly speaking, NOT binomial, since selection are made without replacement — trials are dependent. However,. . .

- P(S on first trial)= $\frac{4000}{5000} = 0.8$

- P(S on first trial)= $\dfrac{4000}{5000} = 0.8$
- P(S on 2nd trial | S on 1st trial)= $\dfrac{3999}{4999} \approx 0.79996$

- P(S on first trial)= $\frac{4000}{5000} = 0.8$
- P(S on 2nd trial | S on 1st trial)= $\frac{3999}{4999} \approx 0.79996$
- P(S on 2nd trial | F on 1st trial)= $\frac{4000}{4999} \approx 0.80016$

- P(S on first trial)= $\frac{4000}{5000} = 0.8$
- P(S on 2nd trial | S on 1st trial)= $\frac{3999}{4999} \approx 0.79996$
- P(S on 2nd trial | F on 1st trial)= $\frac{4000}{4999} \approx 0.80016$
- P(S on 50th trial | all S up to 49th)= $\frac{3951}{4951} \approx 0.7980$

- P(S on first trial)= $\frac{4000}{5000} = 0.8$
- P(S on 2nd trial | S on 1st trial)= $\frac{3999}{4999} \approx 0.79996$
- P(S on 2nd trial | F on 1st trial)= $\frac{4000}{4999} \approx 0.80016$
- P(S on 50th trial | all S up to 49th)= $\frac{3951}{4951} \approx 0.7980$
- P(S on 50th trial | all F up to 49th)= $\frac{4000}{4951} \approx 0.8079$

- P(S on first trial)= $\dfrac{4000}{5000} = 0.8$
- P(S on 2nd trial | S on 1st trial)= $\dfrac{3999}{4999} \approx 0.79996$
- P(S on 2nd trial | F on 1st trial)= $\dfrac{4000}{4999} \approx 0.80016$
- P(S on 50th trial | all S up to 49th)= $\dfrac{3951}{4951} \approx 0.7980$
- P(S on 50th trial | all F up to 49th)= $\dfrac{4000}{4951} \approx 0.8079$
- P(S on 50th trial | n S's up to 49th)= $\dfrac{4000 - n}{4951}$

- P(S on first trial)= $\frac{4000}{5000} = 0.8$
- P(S on 2nd trial | S on 1st trial)= $\frac{3999}{4999} \approx 0.79996$
- P(S on 2nd trial | F on 1st trial)= $\frac{4000}{4999} \approx 0.80016$
- P(S on 50th trial | all S up to 49th)= $\frac{3951}{4951} \approx 0.7980$
- P(S on 50th trial | all F up to 49th)= $\frac{4000}{4951} \approx 0.8079$
- P(S on 50th trial | n S's up to 49th)= $\frac{4000 - n}{4951}$
- Since the sample size 50 is only 1% of the population size (5000), the 50 draws has little effect on the makeup of the population. $P(S)$ stays close to 0.8 regardless of the outcome of prior draws. Trials are nearly independent

- P(S on first trial)= $\frac{4000}{5000} = 0.8$
- P(S on 2nd trial | S on 1st trial)= $\frac{3999}{4999} \approx 0.79996$
- P(S on 2nd trial | F on 1st trial)= $\frac{4000}{4999} \approx 0.80016$
- P(S on 50th trial | all S up to 49th)= $\frac{3951}{4951} \approx 0.7980$
- P(S on 50th trial | all F up to 49th)= $\frac{4000}{4951} \approx 0.8079$
- P(S on 50th trial | n S's up to 49th)= $\frac{4000 - n}{4951}$
- Since the sample size 50 is only 1% of the population size (5000), the 50 draws has little effect on the makeup of the population. $P(S)$ stays close to 0.8 regardless of the outcome of prior draws. Trials are nearly independent

- P(S on first trial)= $\frac{4000}{5000} = 0.8$
- P(S on 2nd trial | S on 1st trial)= $\frac{3999}{4999} \approx 0.79996$
- P(S on 2nd trial | F on 1st trial)= $\frac{4000}{4999} \approx 0.80016$
- P(S on 50th trial | all S up to 49th)= $\frac{3951}{4951} \approx 0.7980$
- P(S on 50th trial | all F up to 49th)= $\frac{4000}{4951} \approx 0.8079$
- P(S on 50th trial | n S's up to 49th)= $\frac{4000-n}{4951}$
- Since the sample size 50 is only 1% of the population size (5000), the 50 draws has little effect on the makeup of the population. $P(S)$ stays close to 0.8 regardless of the outcome of prior draws. Trials are nearly independent
- So X is approx. binomial, $Bin(n = 50, p = 0.8)$.

Rule: Approx. Binomial or Not When Sampling w/o Replacement

Consider sampling without replacement from a dichotomous population of size N. If the sample size (number of trials) n is at most 5% of the population size, the experiment can be analyzed as though it were exactly a binomial experiment.