STAT 23400 Lecture 5: Binomial Distribution

Review

- Variance to measure the variability of a random variable: $Var(X) = E((X - \mu)^2)$:
- Standard deviation: $SD = \sqrt{Var(X)}$
- Linear transformation: $E(aX + b) = aE(X) + b$, $Var(aX + b) = a^2 Var(X)$
- Moments: $m_k = E(X^k)$
- Moment generating function: $M_X(t) = E(e^{tX})$
- Using mgf to kth moments: $m_k = M^{(k)}(0) = E(X^k)$

Outline for Today

- Topics: Binomial distribution, intro to continuous random variable.
- Reading: Section 4.3 in OIS and Section 3.5, 4.1 in MMSA.
- Lab03: Probability. Link posted on canvas.

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 an electronic device is defected;
- whether a subject answers Yes or No to a survey question.

Bernoulli distribution

Bernoulli distribution

A r.v. X follows the Bernoulli distribution, $X \sim B(p)$, if $R(X) = \{0, 1\}$, and the probability function is

$$p(x) = \begin{cases} p & \text{if } x = 1, \\ 1 - p & \text{if } x = 0. \end{cases}$$

E.x. flip a fair coin.

Bernoulli Distribution: Expected Value, Variance & SD

Let $X \sim B(p)$. What is its probability distribution? How about $E(X), Var(X)$ and $SD(X)$.

Binomial Distribution

Suppose a basic experiment is repeated,

1. Each trial has success (S) or failure (F);
2. On each trial, $P(S) = p$ and $P(F) = 1 - p = q$;
3. Trials are independent.
4. It is repeated for n times.

Define $X =$ the number of successes in the n trials.

X is a binomial r.v. We write $X \sim \text{Binom}(n, p)$.

Note: Binomial distribution has two parameters n and p.

Examples

- E.x. Toss a fair coin 25 times and observe the number of heads.

- E.x. Five items are chosen at random with replacement from a lot with 3% defectives. Observe the number of defectives selected.

- E.x. Same as previous example, except that we sample without replacement.

Example: Drawing Balls from a Box

A box contains 1 red ball and 9 green ones. Five draws are made at random from the box with replacement.

- What is the probability that the first two draws are Red and the next 3 are Green?
- What is the probability of getting **exactly 2 Reds in 5 draws**? Is it also equal to $0.1 \times 0.1 \times 0.9 \times 0.9 \times 0.9$?

In general, what is P(getting exactly k Reds in n draws)? Need to consider all possible orderings of the k Reds and the $n - k$ Greens.

Binomial Coeffcients

The number of ways to choose k items, regardless of order, from a total of n distinct items is

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

- $\binom{n}{k}$ is read as "n choose k", also denoted as C_k^n.
- The # of ways to order k Reds and $n-k$ Greens is simply the # of ways to choose k draws to be R out of a total of n draws, and hence equals $\binom{n}{k}$
- For example,

$$\binom{5}{2} = \frac{5!}{2! \times (5-2)!}$$

Factorial

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

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

e.g.

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

- $0! = 1$
- Back to $\binom{5}{2} =?$.

Back to the Example of Drawing Balls from a Box

- 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 calculation can be generalized to other similar problems and the general formula is called the Binomial Formula (next slide).
- See supplementary material at the end of this lecture for more details.

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 $1 - p$.

Let X be the number of successes that occur in the n trials, then X is said to have a binomial distribution with parameters (n, p), denoted as

$$X \sim \text{Binom}(n, p),$$

with the probability distribution (pmf)

$$p(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, \ldots, n$$

Example 1

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

Example 2

Playing the card game 10 times we described in Lecture 4, what is the prob. to get \$10 in total? Can one calculate using the Binomial formula?

Event	X	$P(X)$
Heart (not ace)	1	12/52
Ace	5	4/52
King of spades	10	1/52
All else	0	35/52
Total		1

Binomial Distribution: Expected Value, Variance & SD

In general, the mean, variance and the SD of binomial random variable can be obtained

$$\mu = E(X) = np, \quad Var(X) = np(1-p), \quad \sigma = SD(X) = \sqrt{np(1-p)}.$$

By definition, the expected value is

$$E(X) = \sum_{\text{all } x} x p(x) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$$

- It's not trivial to show the expression above equals np.
- Later We'll use an easier way to prove this.
- Note the SD increases proportionally to $\sqrt{n}$ not n.

Example 3: 2019 COPSS award

- COPSS (Committee of Presidents of Statistical Societies) award is considered the Nobel Prize in statistics. In 2019, it was awarded to Dr. Hadley Wickham (RStudio).

- Suppose we randomly sample 30 people from an ASA (American Statistical Association) event with 75 participants in total, and ask each one selected "Do you support the decision for the 2019 COPSS award?" Suppose the truth is that 45 of the 75 participants support the decision

- Let X be the count in the sample who say yes." Is X Binomial?

Binomial Distribution in R

Binomial Distribution in R

R provides the following 4 functions for Binomial distribution. These functions are

- Pmf: dbinom(x, size, prob, log = FALSE);
- Cdf: pbinom(q, size, prob, lower.tail = TRUE, log.p = FALSE);
- Quantile: qbinom(p, size, prob, lower.tail = TRUE, log.p = FALSE);
- Random Number Generation: rbinom(n, size, prob);
- Type help(rbinom) for R help.

Refer to the supplemented R markdown document for more details (posted on canvas).

Supplementary Materials: Binomial Coeffcients

Examples (Binomial Coeffcients)

- A die is rolled 10 times. What is the total number of ways to get exactly 3 aces?

$$\binom{10}{3} = \frac{10!}{3!7!} = 120$$

- In poker, the order of the cards you're dealt is irrelevant. How many possible poker hands are there?

$$\binom{52}{5} = \frac{52!}{5!47!} = 2598960.$$

- From a box of n distinct items, # of ways to select k items and take them out of the box = # of ways to select $n - k$ items and take the remaining k items out of the box

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

Examples (Binomial Coeffcients)

- $\binom{n}{0} = \frac{n!}{0! \times n!} = 1$
  There is only 1 way to order 0 Reds and n Greens

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

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

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

Technical Notes on Binomial Coeffcients I

How many ways are there to arrange n people in a row?

- There are n possibilities for the first position.
- The 2nd person can be anybody except the first one. There are $n - 1$ possibilities.
- ...
- As a whole, there are

$$n \times (n-1) \times \cdots \times 2 \times 1 = n! \text{ ways.}$$

How many ways are there to select k people out of $n (k \leq n)$, when order matters?

- There are n people to choose to the first one
- The second one can be anyone but the first one selected. There are $n-1$ possibilities.
- ... The kth one can be anyone but the first $k-1$ people that have been selected. There are $n-k+1$ possibilities.

In total, there are

$$n \times (n-1) \times \cdots \times (n-k+1) = \frac{n!}{(n-k)!} \text{ ways.}$$

How many ways are there to select k people out of n, **regardless of order**?

Let N be the number of ways. We can find the value of N by considering the problem of finding "**the number of ways to select k out of n people, when order matters**". We can generate all such ways by

- First selecting k people out of n, regardless of order. There are N ways to do this.
- There are k! ways to order the k selected people
- Therefore $N = \frac{n!}{(n-k)!k!} = \binom{n}{k}$.

Continuous Random variable

Continuous Random Variables

- Discrete random variables can not describe all stochastic experiments. For example, consider

$$X : \text{the height (in cm) of students in a class}$$
$$X \in (140, 210) \Rightarrow \text{X is not discrete}$$

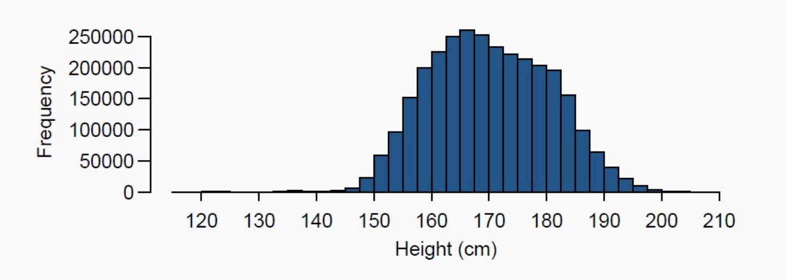
- In order to describe the distribution of X, we need to introduce continuous r.v.

Frequency Scale of Histograms

For histograms in a frequency scale,

$$\text{bar height} = \text{ count of observations in that bin.}$$

Below is a histogram of the distribution of heights of US adults.

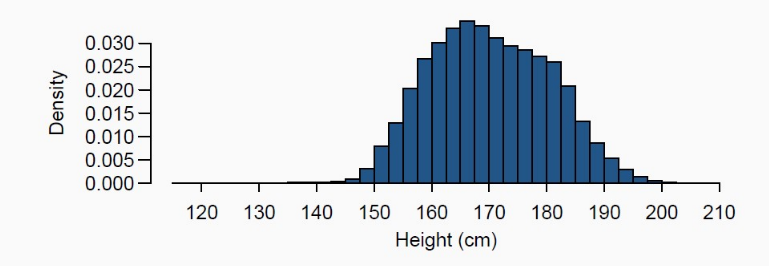


Density Scale of Histograms

For a histogram in a density scale,

$$\textbf{bar area} = \text{ proportion of observations in that bin.}$$

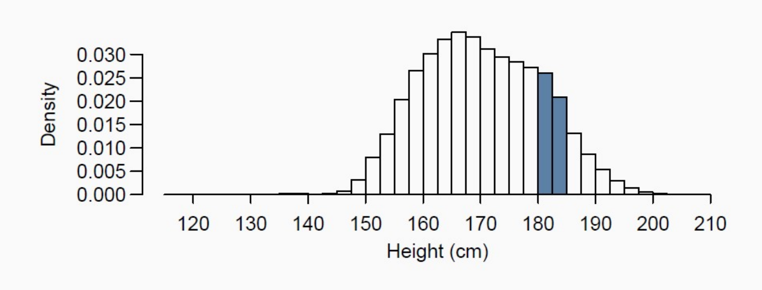
So, bar height $= \dfrac{\#\text{ of observations in the bin}}{(\text{total }\#\text{ of observations})(\text{bin width})}$



Whichever scale is used, the shape of a histogram is not affected.

Density Scale of Histograms

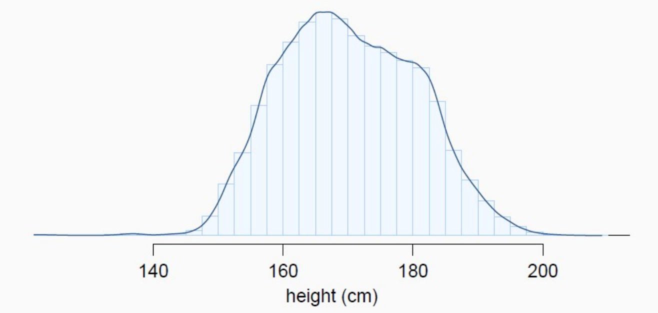
In a density scale, proportion of US adults that are 180-185 cm tall = area under the histogram between 180-185 (shaded region)



In a density scale, the total area under a histogram is 1 (why?).

From Histograms to Density Curves

We might attempt to approximate a histogram by a smooth curve, called a (probability) density functions.



- A density curve is nonnegative, i.e., always on or above the zero line.
- The total area under the density curve is always 1, or 100%.

Therefore, the proportion US adult between 180 cm and 185 cm tall can be estimated as the shaded area under the curve. (The exact proportion is the area under the histogram).