04 - Binomial distribution

HCI/PSYCH 522
Iowa State University

February 10, 2022

Overview

- Random variables
  - Bernoulli distribution
    - Model for success/failure
  - Binomial distribution
    - Model for success/failure counts

Random variables

Suppose you will run a study (any data collection) and you will have some outcome. A random variable is any numerical summary of the outcome of that study.

We may know the following quantities for random variables:

- Distribution:
  - Image, i.e. the possible values for X
  - For discrete random variables, probability mass function (pmf) $P(X = x)$.
  - Cumulative distribution function (cdf), $P(X \leq x)$.
- Expectation (average value), $E[X]$
- Variance (variability), $Var[X]$
- Standard deviation (variability), $\sqrt{Var[X]}$

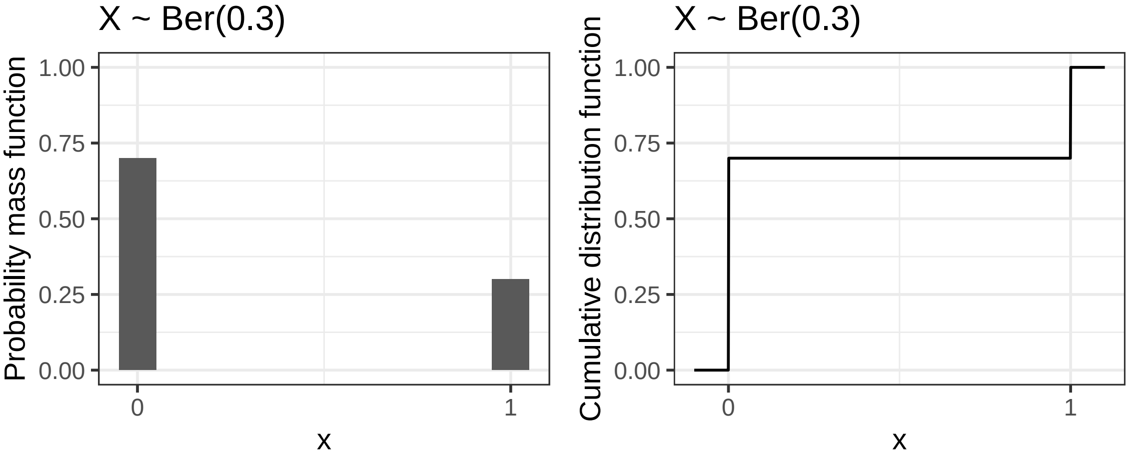
Bernoulli

Suppose we are interested in recording the success or failure. By convention, we code 1 as a success and 0 as a failure and call this value X.

If $X \sim Ber(p)$, then X is a Bernoulli random variable with probability of success p and

- $P(X = 1) = p$,
- $P(X = 0) = (1 - p)$,
- $E[X] = p$,
- $Var[X] = p(1 - p)$, and
- $SD[X] = \sqrt{p(1 - p)}$.

Bernoulli


X ~ Ber(0.3)
Probability mass function
1.00
0.75
0.50
0.25
0.00
0
1
x
X ~ Ber(0.3)
Cumulative distribution function
1.00
0.75
0.50
0.25
0.00
0
1
x

6-sided die example

Let X be an indicator that a 1 was rolled on a 6-sided die. More formally

$$X = \begin{cases} 1 & \text{if a 1 is rolled} \\ 0 & \text{if anything else is rolled.} \end{cases}$$

Then we write $X \sim Ber(1/6)$ and know

- $P(X = 1) = 1/6$,
- $P(X = 0) = 1 - 1/6 = 5/6$,
- $E[X] = 1/6$,
- $Var[X] = 1/6 \times (1 - 1/6) = 1/6 \times 5/6 = 5/36$, and
- $SD[X] = \sqrt{5/36} = \sqrt{5}/6$.

Binomial

Suppose we count the number of successes in n attempts with a common probability of success p where each attempt is independent and call this count Y.

If $Y \sim Bin(n, p)$, then Y is a binomial random variable with n attempts and probability of success p and

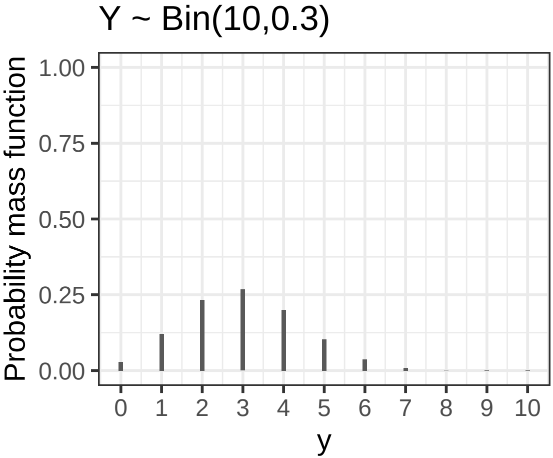
- $P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}$ for $y = 0, 1, \ldots, n$
- $E[Y] = np$
- $Var[Y] = np(1-p)$

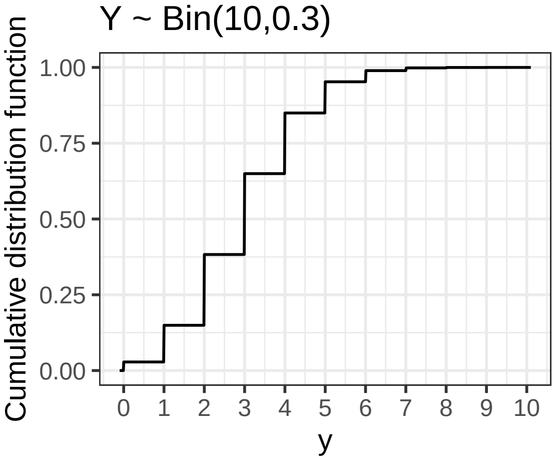
We can use R to calculate the probability mass function values, e.g. if $Y \sim Bin(10, 1/6)$ and we want to calculate $P(Y = 2)$ we use

```
n <- 10; p <- 1/6; y <- 2
dbinom(y, size = n, prob = p)

## [1] 0.29071
```

Binomial


Y ~ Bin(10,0.3)
Probability mass function
1.00
0.75
0.50
0.25
0.00
0 1 2 3 4 5 6 7 8 9 10
y


Y ~ Bin(10,0.3)
Cumulative distribution function
1.00
0.75
0.50
0.25
0.00
0 1 2 3 4 5 6 7 8 9 10
y

6-sided die example

Suppose we roll a 6-sided die 10 times and record the number of times we observed a 1. Assume independence between our roles, we have $Y \sim Bin(10, 1/6)$ and we know

- $E[Y] = 10 \times 1/6 = 10/6$,
- $Var[Y] = 10 \times 1/6 \times (1 - 1/6) = 10/6 \times 5/6 = 50/36$, and
- $SD[Y] = \sqrt{10 * 5/36} = \sqrt{50}/6$.