- Quiz II on Oct. 17th (a week from Thursday)

- HW 4 sol. uploaded

- HW 5 due Thursday, Oct. 10th.

- Filled slides were uploaded!

STAT 305: Chapter 5

Part II

Amin Shirazi

ashirazist.github.io/stat305.github.io

Discrete Random Variables

Meaning, Use, and Common Distributions

General Info About Discrete RVs

Reminder: What is a Random Variable?

Random Variables, we have already defined, take real-numbered ($\mathbb{R}$) values based on outcomes of a random experiment.

- If we know the outcome, we know the value of the random variable (so that isn't the random part).
- However, before we perform the experiment we do not know the outcome - we can only make statements about what the outcome is likely to be (i.e., we make "probabilistic" statements).
- In the same way, we do not know the value of the random variable before the experiment, but we can make probability statements about what value the RV might take.

General Info

Reminder: RVs

Discrete?

Terms & Notation

Common Terms and Notation for Discrete RVs

Of course, we can't introduce a *sort of* new concept without introducing a whole lot of new terminology.

We use capital letters to refer to discrete random variables: X, Y, Z, ...

We use lower case letters to refer to values the discrete RVs can take: x, x_1, y, z, ...

While we can use $P(X = x)$ to refer to the probability that the discrete random variable takes the value x, we usually use what we call the **probability function**:

- For a discrete random variable X, the probability function $f(x)$ takes the value $P(X = x)$
- In otherwords, we just write $f(x)$ instead of $P(X = x)$.

General Info

Reminder: RVs

Discrete?

Terms & Notation

Common Terms and Notation for Discrete RVs

We also have another function related to the probabilities, called the **cumulative probability function**.

For a discrete random variable X taking values $x_1, x_2, \ldots$ the CDF or **cumulative probability function** of X, $F(x)$, is defined as

$$F(x) = \sum_{z<x} f(z)$$

Which in other words means that for any value x,

$$f(x) = P(X = x)$$

and

$$\boxed{F(x) = P(X \leq x)}$$

General Info

Reminder: RVs

Discrete?

Terms & Notation

Common Terms and Notation for Discrete RVs (cont)

The values that X can take and the probabilities attached to those values are called the **probability distribution** of X (since we are talking about how the total probability 1 gets spread out on (or distributed to) the values that X can take).

Example

Suppose that the we roll a die and let T be the number of dots facing up. Define the probability distribution of T. Find $f(3)$ and $F(6)$.

General Info

Reminder: RVs

Discrete?

Terms & Notation

Example: [Torque]

Let $Z =$ the torque, rounded to the nearest integer, required to loosen the next bolt on an apparatus.

Z	11	12	13	14	15	16	17	18	19	20
f(z)	0.03	0.03	0.03	0.06	0.26	0.09	0.12	0.20	0.15	0.03

Calculate the following probabilities:

- $P(Z \leq 14)$
- $P(Z > 16)$
- $P(Z \text{ is even})$
- $P(Z \in \{15, 16, 18\})$

General Info

Reminder: RVs

Discrete?

Terms & Notation

Example: [Torque]

Z	11	12	13	14	15	16	17	18	19	20
f(z)	0.03	0.03	0.03	0.06	0.26	0.09	0.12	0.20	0.15	0.03

- $P(Z \leq 14)$

- $P(Z > 16)$

General Info

Reminder: RVs

Discrete?

Terms & Notation

Example: [Torque]

Z	11	12	13	14	15	16	17	18	19	20
f(z)	0.03	0.03	0.03	0.06	0.26	0.09	0.12	0.20	0.15	0.03

- $P(Z \text{ is even})$

- $P(Z \in \{15, 16, 18\})$

General Info

Reminder: RVs

Discrete?

Terms & Notation

More on CDF

> The *cumulative probability distribution (cdf)* for a random variable X is a function $F(x)$ that for each number x gives the probability that X takes that value or a smaller one, $F(x) = P[X \leq x]$.

Since (for discrete distributions) probabilities are calculated by summing values of $f(x)$,

$$F(x) = P[X \leq x] = \sum_{y \leq x} f(y)$$

More on CDF

Properties of a mathematically valid cumulative distribution function:

- $F(x) \geq 0$ for all real numbers x
- $F(x)$ is monotonically **increasing**
- $F(x)$ is right continuous
- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$
  - This means that $0 \leq F(x) \leq 1$ for **any CDF**

> In the discrete cases, the graph of $F(x)$ will be n stair-step graph with jumps at possible values of our random variable and height equal to the probabilities associated with those values

General Info

Reminder: RVs

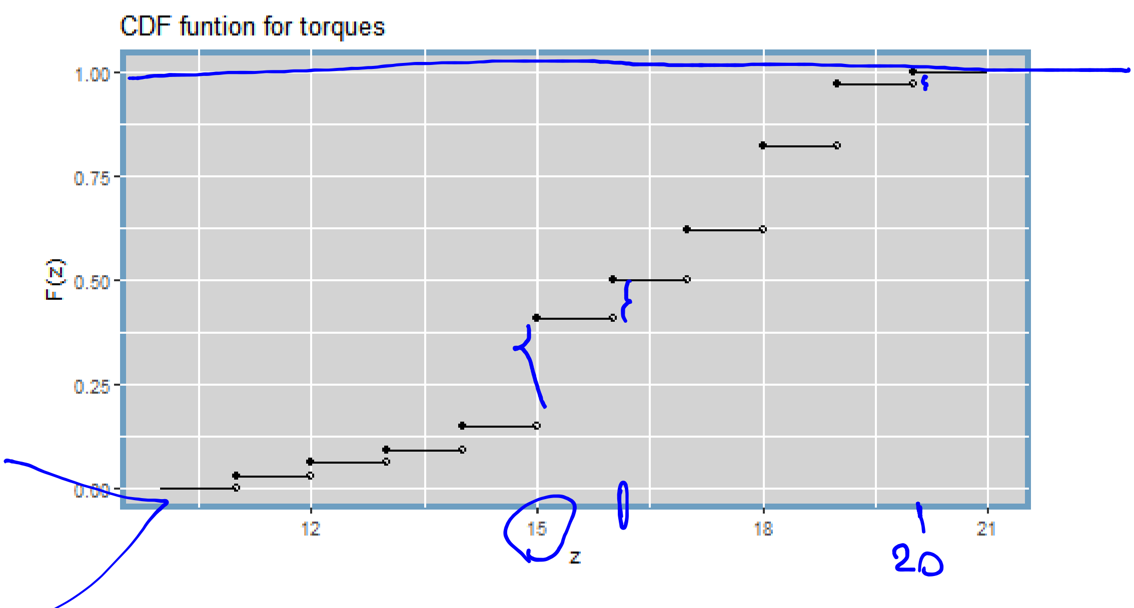
Discrete?

Terms & Notation

More on CDF

Example: [Torque] Let $Z =$ the torque, rounded to the nearest integer, required to loosen the next bolt on an apparatus.

Z	11	12	13	14	15	16	17	18	19	20
f(z)	0.03	0.03	0.03	0.06	0.26	0.09	0.12	0.20	0.15	0.03



General Info

Reminder: RVs

Discrete?

Terms & Notation

More on CDF

Calculate the following probabilities using the **cdf only**:

- $F(10.7)$
- $P(Z \leq 15.5)$
- $P(12.1 < Z \leq 14)$
- $P(15 \leq Z < 18)$

General Info

Reminder: RVs

Discrete?

Terms & Notation

More on CDF

One more example

Say we have a random variable Q with pmf:

q	f(q)
1	0.34
2	0.10
3	0.22
7	0.34

Draw the CDF

General Info

Reminder: RVs

Discrete?

Terms & Notation

Summaries

Almost all of the devices for describing relative frequency (empirical) distributions in Ch. 3 have versions that can describe (theoretical) probability distributions.

1. Measures of location == Mean
   population : μ
   sample : $\bar{x}$
2. Measures of spread == variance
   pop : σ^2
   sample : s^2
3. Histogram == probability histograms based on theoretical probabilities

Mean

and

Variance

of Discrete Random Variables

Mean of a Discrete Random Variable

For a discrete random varable, X, which can take values $x_1, x_2, \ldots$ we define **the mean of X** (also known as **the expected value of X**) as:

$$E(X) = \sum_{i=1}^{n} \overset{\text{value}}{x_i} \cdot \underbrace{f(x_i)}_{\text{probability}}$$

We often use the symbol μ instead of $E(X)$.

Also, just to be confusing, you will often see $\boxed{EX}$ instead of $\boxed{E(X)}$. Use context clues.

Example:

Suppose that the we roll a die and let T be the number of dots facing up. Find the expected value of T.

x	1	2	3	4	5	6
$P(X=x)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

$$E(X) = \sum_{x=1}^{6} x\, P(X=x)$$

$$= 1\,P(X=1) + 2\,P(X=2) + 3\,P(X=3) + 4\,P(X=4) + 5\,P(X=5)$$

$$+ 6\,P(X=6)$$

$$= 1 \cdot \tfrac{1}{6} + 2 \cdot \tfrac{1}{6} + 3 \cdot \tfrac{1}{6} + 4 \cdot \tfrac{1}{6} + 5 \cdot \tfrac{1}{6} + 6 \cdot \tfrac{1}{6}$$

$$= \tfrac{1}{6}(1+2+3+4+5+6) = \frac{21}{6} = 3.5$$

Interpretation:

If rolling a fair die once, on average, we see 3.5, Expected value of "anything" with a probability function $f_X(x)$ < continuous / discrete

$$E(\text{anything}) = \textstyle\sum \text{ or } \int (\text{anything}) \times (\text{Prob.})$$

$X \sim f_X(x) = P(X=x)$: discrete

$$E(X) = \sum_{x \in S_X} x \cdot P(X=x)$$

support of X

$$E(X^2) = \sum_{x \in S_X} x^2 P(X=x)$$

X	0	1	2	3
$P(\boxed{X=x})$	0.2	0.1	0.5	0.2

$$E(X) = 0 \cdot P(X=0) + 1 \cdot \underline{P(X=1)} + 2P(X=2) + 3 \cdot P(X=3)$$

$$= 0 + 1 \cdot 0.1 + 2(0.5) + 3(0.2) = 1.7$$

$$E(\sqrt{X}) = \sqrt{0}\, P(X=0) + \sqrt{1}\, P(X=1) + \sqrt{2}\, P(X=2) + \sqrt{3}\, P(X=3)$$

$$=$$

Recall: population variance:
$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$

General Info

Reminder: RVs

Discrete?

Terms & Notation

Variance of a Discrete Random Variable

For a discrete random varable, X, which can take values $x_1, x_2, \ldots$ and has mean μ we define **the variance of X** as:

E(X)

$$Var(X) = \sum_{i=1}^{n}(x_i - \mu)^2 \cdot f(x_i)$$

There are other usefule ways to write this, most importantly:

$$Var(X) = \sum_{i=1}^{n} x_i^2 \cdot f(x_i) - \mu^2$$

which is the same as

$$\mathrm{Var}X = \sum_x (x - \mathrm{E}X)^2 f(x) = \boxed{\mathrm{E}(X^2) - (\mathrm{E}X)^2}.$$

General Info

Reminder: RVs

Discrete?

Terms & Notation

Variance of a Discrete Random Variable

Example:

Suppose that the we roll a die and let T be the number of dots facing up. What is the variance of T?

t	1	2	3	4	5	6
$P(T=t)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

$E(T) = 21/6 = 3.5$

$Var(T) = E(x^2) - [E(x)]^2$

$$E(x^2) = \sum_{x=1}^{6} x^2 P(X=x)$$

$$= 1^2 \overbrace{P(X=1)}^{1/6} + 2^2 \overbrace{P(X=2)}^{1/6} + 3^2 \overbrace{P(X=3)}^{1/6} + 4^2 \overbrace{P(X=4)}^{1/6}$$

$$+ 5^2 \overbrace{P(X=5)}^{1/6} + 6^2 \overbrace{P(X=6)}^{1/6}$$

$$= \frac{1}{6}\left(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2\right)$$

$$= 15.17$$

$$var(\underline{X}) = E(\underline{X}^2) - [E(\underline{X})]^2 = 15.17 - (3.5)^2$$

$$\boxed{= 2.92}$$

The standard deviation of X is

$$S = \sqrt{var(X)} = \sqrt{2.92} = 1.7088$$

Variance of a Discrete Random Variable

Example

Say we have a random variable Q with pmf:

q	f(q)	$q f(q)$	$q^2 f(q)$
1	0.34	0.34	0.34
2	0.10	0.2	$2^2(0.1)$
3	0.22	0.66	$3^2(0.22)$
7	0.34	2.28	$7^2(0.34)$

Find the variance and standard deviation

$E(Q) = ? \longrightarrow \sum_{q \in \{1,2,3,7\}} q f_Q(q) = 0.34 + 0.2 + .66 + 2.28$

$= 3.58$

$E(Q^2) = ? \longrightarrow \sum_q q^2 f(q) = 19.38$

$$Var(Q) = E(Q^2) - [E(Q)]^2$$

$$= 19.38 - (3.58)^2$$

$$= 6.56.$$

Variance is ALWAYS ≥ 0

General Info

Reminder: RVs

Discrete?

Terms & Notation

Summary

Discrete Random Variables

- Discrete RVs are RVs that will take one of a countable set of values.
- When working with a discrete random variable, it is common to need or use the RV's
  - probability distribution: the values the RV can take and their probabilities
  - probability function: a function where $f(x) = P(X = x)$
  - cumulative probability function: a function where $F(x) = P(X \leq x)$.
  - mean: a value for X defined by $EX = \sum_x x \cdot f(x)$
  - variance: a value for X defined by $VarX = \sum_x (x - \mu)^2 \cdot f(x)$

Your Turn:

Chapter 5 Handout 1

Common Distributions

Working with Off The Shelf Random Variables

Common Distributions

Why Are Some Distributions Worth Naming?

Even though you may create a random variable in a unique scenario, the way that it's probability distribution behaves (mathematically) may have a lot in common with other random variables in other scenarios. For instance,

> I roll a die until I see a 6 appear and then stop. I call X the number of times I have to roll the die in total.
>
> I flip a coin until I see heads appear and then stop. I call Y the number of times I have to flip the coin in total.
>
> I apply for home loans until I get accepted and then I stop. I call Z the number of times I have to apply for a loan in total.

Why Are Some Distributions Worth Naming? (cont)

In each ot the above cases, we count the number of times we have to do some action until we see some specific result. The only thing that really changes from the random variables perspective is the likelyhood that we see the specific result each time we try.

Mathematically, that's not a lot of difference. And if we can really understand the probability behavior of one of these scenarios then we can move our understanding to the different scenario pretty easily.

By recognizing the commonality between these scenarios, we have been able to identify many random variables that behave very similarly. We describe the similarity in the way the random variables behave by saying that they have a common/shared distribution.

We study the most useful ones by themselves.

Important: On HW 4 grading.

- Double check problem 5, there has been different values of $\sum x_i$, $\sum y_i$, $\sum x_i^2$, $\sum y_i^2$, $\sum x_i y_i$ in the question & in the solution

→ If you missed some points on that question & think you did everything well, scan your answer of "question 5" and email me. I'll update your grade.

- Quiz II on October 17^{th}. (one week from today)
  - sample quiz + solution + note sheet were posted in "Exam materials" on the page.

Quiz II coverage:
- chapter 4: line/surface fitting
- chapter 5: Discrete random variables. (until the end of Poisson Distribution)

- Hw 6 will be posted after class.
- Questions related to ch 5 → Discrete r.v. (They are good practice for quiz II).
- Due Thursday, OCT 17th in-class. (The same day as quiz II)

The Bernoulli Distribution

General Info

Common Distributions

Background

Bernoulli

The Bernoulli Distribution

Origin: A random experiment is performed that results in one of two possible outcomes: success or failure. The probability of a successful outcome is p.

Definition: X takes the value 1 if the outcome is a success. X takes the value 0 if the outcome is a failure.

probability function:

$$f(x) = \begin{cases} p & x = 1, \\ 1-p & x = 0, \\ 0 & o.w. \end{cases}$$

success

failure

otherwise.

which can also be written as

$$f(x) = \begin{cases} p^x(1-p)^{1-x} & x = 0, 1 \\ 0 & o.w. \end{cases}$$

Example:

- In fliping a coin:

Define success as observing a "Head" in fliping a coin.

let X be a r.v associated with observing a "Head".

Then X is a binary trial (r.v) with probability of

success $\boxed{p = 1/2}$

and probability of failure $\boxed{1-p = 1 - 1/2 = \frac{1}{2}}$.

- what if we flip the coin twice! what is the prob. that we see Heads in both trials? (each trial is independent)

success

P(Heads in both trials) = P(X=1 in both trials)

= P(X=1 on the first trial $\boxed{\text{and}}$ X=1 on the second trial)

$= P(X=1$ on the first trial$) \cdot P(X=1$ on the second trial$)$

$= P(X=1) \cdot P(X=1)$

Bernoulli trial $= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

Example: A Cyclone Basketball player takes 4 independent free throws with a probability of 0.7 of getting a basket on each shot. The number of baskets he makes is recorded. Each throw is a bernoulli trial then and they are independent. what is the probability that he:

① He makes 4 baskets?

$X \sim \text{Bernoulli}(p=0.7)$

$P(\text{"He makes 4 baskets"}) = P(X=1 \text{ and } X=1 \text{ and } X=1 \text{ and } X=1)$

success in the first shot

in the second

indep. trials $= P(X=1) \cdot P(X=1) \cdot \underbrace{P(X=1)}_{\text{success 3rd}} \cdot \underbrace{P(X=1)}_{\text{success in the 4th}}$

$= (0.7)^4$

(2) He makes zero baskets?

P(He doesn't make a basket in the 1st and 2nd and 3rd and 4th)

$$= P(\overbrace{X=0}^{1st} \text{ and } \overbrace{X=0}^{2nd} \text{ and } \overbrace{X=0}^{3r} \text{ and } \overbrace{X=0}^{4th})$$

$$= [P(X=0)]^4 = [1-0.7]^4$$

(3) He makes at least 3 baskets?

$$P(X \geq 3) = P(X=3) + P(X=4)$$

(There are 4 trials)

= P(He makes 3 baskets and one out) + P(4 baskets)

= P(3 successes and one failure) + P(X=4)

$$= P(X=1 \text{ and } X=1 \text{ and } X=1 \text{ and } X=0) + P(X=4)$$

$$= [P(X=1)]^3 [P(X=0)] + [P(X=1)]^4$$

$$= (0.7)^3 (1-0.7)^1 + (0.7)^4$$

Bernoulli Distribution

Expected Value and Variance

General Info

Common Distributions

Background

Bernoulli

The Bernoulli Distribution

Expected value: $\boxed{E(X) = p}$

Proof optional.

$(1-p)$

$$EX = \sum_{x=0}^{1} x\, p_X(x) = 0 * p_X(0) + 1 * p_X(1)$$

$$= 0 + 1 * p$$

$$\boxed{= p}$$

General Info

Common Distributions

Background

Bernoulli

The Bernoulli Distribution

Variance: $\boxed{Var(X) = (1-p) \cdot p}$

proof optional

$$EX^2 = \sum_{x=0}^{1} x^2 f_X(x) = 0^2 \cdot f_X(0) + 1^2 \cdot f_X(1)$$

$$= 0 + 1 \cdot f_X(1)$$

$$= p$$

$$\rightarrow Var(X) = \boxed{EX^2 - (EX)^2}$$

$$= p - (p)^2 = p - p^2 = \boxed{p(1-p).}$$

General Info

Common Distributions

Background

Bernoulli

The Bernoulli Distribution

A few useful notes:

- In order to say that " X has a bernoulli distribution with success probability p " we write $X \sim Bernoulli(p)$
- Trials which results in which the only possible outcomes are "success" or "failure" are called **Bernoulli Trials** .
- The value p is the Bernoulli distribution's **parameter**. We don't treat parameters like random values - they are fixed, related to the real process we are studying.
- "Success" does not mean something we would perceive as "good" has happened. It just means the outcome we were watching for was the outcome we got.
- Please note: we have two outcomes, but the probability for each outcome is **not** the same (duh!).

The Binomial Distribution

Common Distributions

Background

Bernoulli

Binomial

The Binomial Distribution

Origin: A series of n independent random experiments, or trials, are performed. Each trial results in one of two possible outcomes: success or failure. The probability of a successful outcome, p, is the same across all trials.

Definition: For n trials, X is the number of trials with a successful outcome. X can take values $0, 1, \dots, n$.

probability function:

With $\boxed{0 < p < 1,}$

$$\boxed{f(x)} = \begin{cases} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} & x = 0, 1, \dots, n \\ 0 & o.w. \end{cases}$$

fixed prior to the experiment

X: # of "successes" in "n" Bernoulli trials

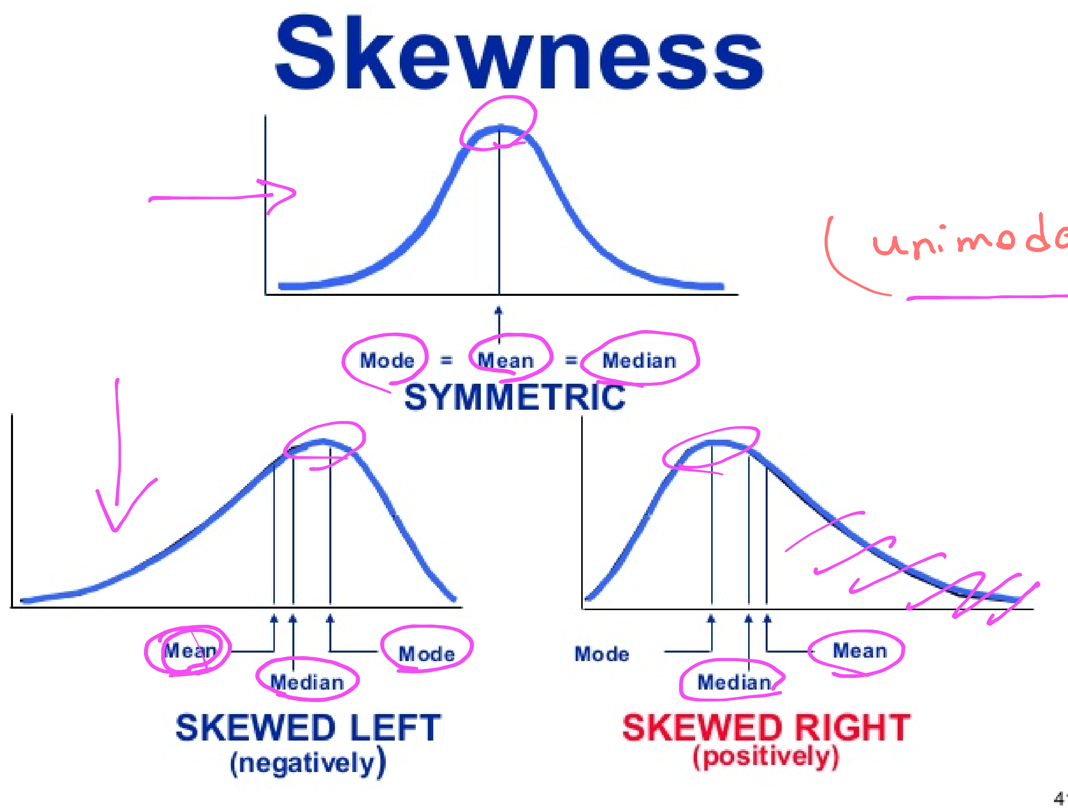
"n choose x" $\binom{n}{x}$

where $n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 1$ and $\boxed{0! = 1.}$

Examples of Binomial Distribution

- Number of hexamine pallets in a batch of $n = 50$ total pallets made from a palletizing machine that conform to some standard.
- Number of runs of the same chemical process with percent yield above 80 given that you run the process 1000 times.
- Number of winning lottery tickets when you buy 10 tickets of the same kind.

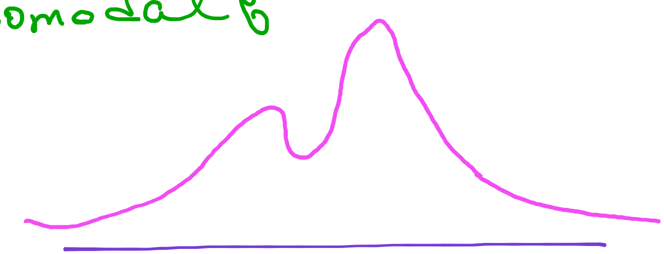
Digression: shape of distributions:



(unimodal)

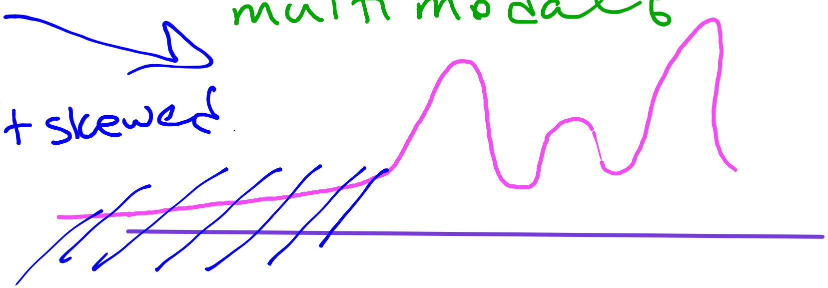
symmetric

biomodal:



left skewed

multimodal:

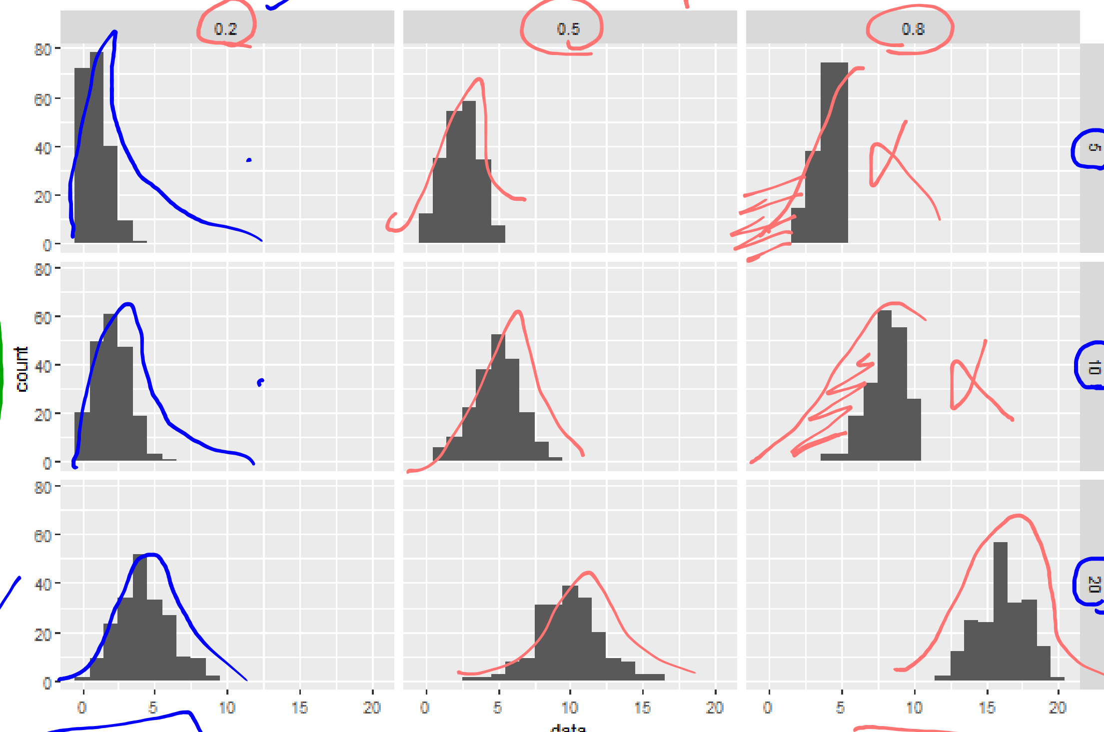


for different values of parameters (n,p) in binomial dist., the shape is different.

The Binomial Distribution

Plots of Binomial distribution based on different success probabilities and sample sizes.

p



n

skeweness decreases as "n" increases

$p<0.5$ => right skewed

$p=0.5$ symmetric

$p>0.5$ left skewed.

The Binomial Distribution

Example [10 component machine]

Suppose you have a machine with 10 independent components $\boxed{\text{in series}}$. The machine only works if all the components work. Each component succeeds with probability $\boxed{p = 0.95}$ and fails with probability $1 - p = 0.05$.

Let Y be the number of components that succeed in a given run of the machine. Then

$\boxed{\text{fully specified distribution}}$

$$\rightarrow Y \sim \text{Binomial}(n = 10, p = 0.95)$$

Question: what is the probability of the machine working properly?

$$P(\text{machine working}) = P(\text{all components work})$$

$$= P(Y=10)$$

$$\left[= \frac{n!}{y!\,(n-y)!}\, p^{y} (1-p)^{n-y} \right]$$

$$\begin{cases} y = 10 \\ n = 10 \end{cases}$$

$$= \frac{\cancel{10!}}{\cancel{10!}\,\underbrace{(10-10)!}_{0!=1}} (0.95)^{10} \underbrace{(1-0.95)^{\cancel{10-10}\;0}}_{(0.05)^{0}=1}$$

$$= (0.95)^{10}$$

$$= 0.5987$$ (not very reliable machine)

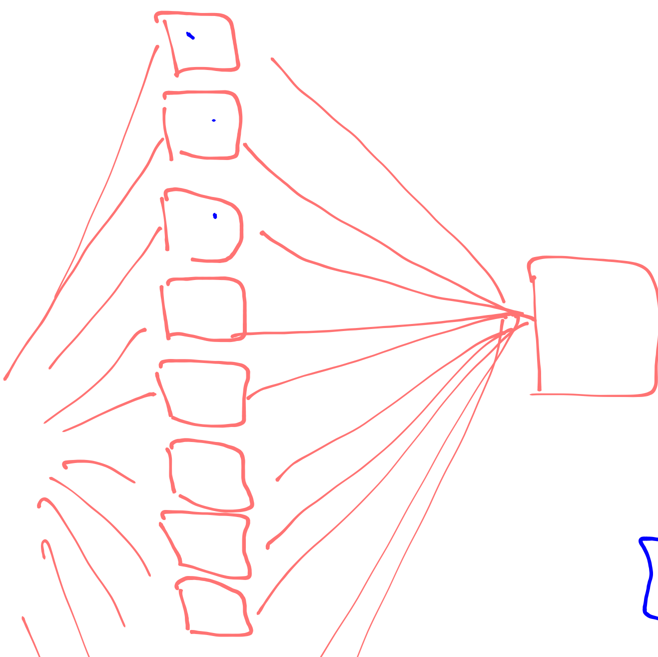
The Binomial Distribution

Example [10 component machine]

$$Y \sim \text{Binomial}(n = 10, p = 0.95)$$

What if I arrange these 10 components in parallel? This machine succeeds if at least 1 of the components succeeds.

What is the probability that the new machine succeeds?



a parallel system works if at least one component works.

P(the machine works)

= P(at least one component works)

$[P(Y \geq 1) = 1 - P(Y < 1)]$

$$= 1 - P(\text{all fail})$$

$$\boxed{= 1 - P(Y=0)}$$

$$= 1 - \underbrace{\frac{10!}{0!(10-0)!}}_{1} \underbrace{(0.95)^0}_{1} (1-0.95)^{10-0}$$

$$= 1 - (0.05)^{10}$$

$$\approx 1$$