STAT 206

INTRODUCTION

- Probability: quantify uncertainty
- Notation:
  - Sample space (S, Ω): all possible outcomes
  - Event: subset of sample space that you're interested in
  - Random experiment: setting where outcomes are uncertain
- Ex: Tossing a coin 3 times. P(getting 3 heads)

$$\Omega = \{\underline{HHH}, HHT, HTH, HTT, THT, TTT, TTH, \ldots\}$$

- Definition:

1. Classical definition:

$$P(A) = \frac{\#\text{ of favorable outcomes}}{\#\text{ of outcomes in } \Omega}$$ provided <u>equal likiliness</u>.

Problems:

i) Ω must be finite

ii) Circular defined

iii) $\boxed{\text{Equally likely}}$ ⟹ Make sure this assumption is true.

~~P(airplane crash) = $\frac{1}{2}$~~ = 50%

↳ 2 events in Ω are not equally likely

2. Relative frequency: long term proportion of success
   ↳ Not practical
3. Subjective prob. distr: probability is based on info

COUNTING RULES

- Addition + multiplication rules:
  - Addition: event A <u>or</u> event B $\Rightarrow$ # of total = $\#_A + \#_B$
  - Multiplication: event A <u>and</u> event B $\Rightarrow$ # of total = $\#_A \times \#_B$

Ex:// 3 jeans, 4 shirts, 5 pairs of socks. # of total outfits?

Need to have all 3 $\Rightarrow$ multiplication.

$\therefore$ # of outfits = # of jeans $\times$ # of shirts $\times$ # of socks

$= 3 \times 4 \times 5$

$= 60$

Ex:// License plates: 3 letters and 3 numbers

i) How many license plates?

$$\underline{26} \cdot \underline{26} \cdot \underline{26} \cdot \underline{10} \cdot \underline{10} \cdot \underline{10} = 26^3 \cdot 10^3$$

↳ Selection graphics: space for each decision, put # of outcomes in space.

ii) Starts w/ B, ends w/ 9

$$\underline{1} \cdot \underline{26} \cdot \underline{26} \cdot \underline{10} \cdot \underline{10} \cdot \underline{1} = 26^2 \cdot 10^2$$

Ex:// Randomly selecting 2 digit # from $\{1, \ldots, 9\}$.

Find prob. that exactly 1 digit is odd.

Break down into simpler $\Rightarrow$ P(exactly 1 digit) = P(first is odd) + P(second is odd)

$|\Omega| = 81$

# of outcomes where 1^{st} is odd = 20

" 2^{nd} is odd = 20

$P(\text{exactly 1 digit odd}) = \frac{20}{81} + \frac{20}{81} = \frac{40}{81}$

- Total # of ways to select k items from n items.
  - n^k rule: order matters, repetition is allowed.

    Ex:// Pin has 4 digits. Find # of outcomes.

    # of out comes: 10^4

    $\underline{10} \cdot \underline{10} \cdot \underline{10} \cdot \underline{10}$

  - Permutation: Order matters, repetition is not allowed.

    Ex:// Pin has 4 digits, but each digit is unique. Find # of outcomes:

    $\underline{10} \cdot \underline{9} \cdot \underline{8} \cdot \underline{7}$

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

    Ex:// 6 letter word. Constructed from A - F. Find prob. that ABC is together.

    $|\Omega| = 6!$

    P(ABC is together) = Orderings w/in ABC × # of ways arranging ABC in word.

    # of outcomes w/ ABC orderings: 6

    $\underline{3}\ \underline{2}\ \underline{1}$

    # of ways of arranging ABC in word:

$$\underline{4} \cdot \underline{3} \cdot \underline{2} \cdot \underline{1} = 24$$

$$\therefore \quad P(ABC \text{ together}) = \frac{24 \times 6}{6!}$$

- Combinations: order does not matter, repetition is allowed.
  - Divide permutations by the # of ways you can order w/in permutation $\Rightarrow k!$

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

  - Ex:// Random 5 card hand. Find $P(\text{full house})$

    Full house: 3 of 1 kind, 2 of another kind.

    ↳ kind: digit/face (A → J, 10 → 2)

    # of total outcomes $= {}^{52}C_5$

$$13 \cdot \underbrace{\overbrace{\_\_ \ \_\_ \ \_\_}^{^{4}C_3}}_{1^{st}\text{ kind}} \ \underbrace{\overbrace{\_\_ \ \_\_}^{^{4}C_2}}_{2^{nd}\text{ kind}} \cdot 12$$

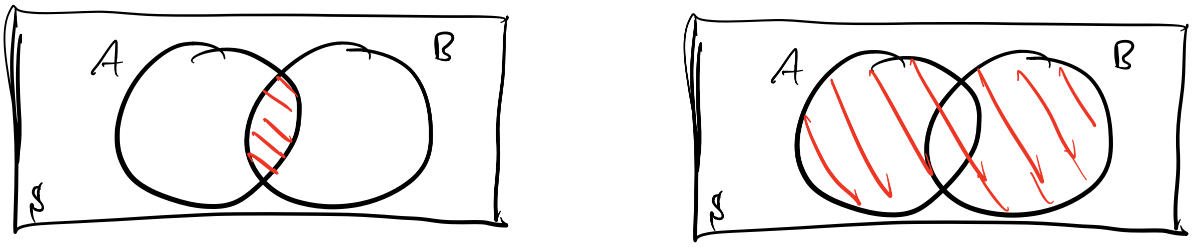
$$= 13 \cdot {}^{4}C_3 \cdot {}^{4}C_2 \cdot 12$$

$$\therefore \quad P(A) = \frac{13 \cdot {}^{4}C_3 \cdot {}^{4}C_2 \cdot 12}{^{52}C_5}$$

$$\underset{\substack{\downarrow \\ ^{4}C_3}}{\underline{13}} \cdot \underset{\substack{\downarrow \\ ^{4}C_2}}{\underline{12}}$$

AXIOMATIC APPROACH TO PROBABILITY

- Probability is a function: $\Omega \rightarrow \mathbb{R}$ (every subset of sample space will have a corresponding number)
  - Must fulfill 2 criteria:
    1. $P(\Phi) = 0$. $P(S/\Omega) = 1$
    2. If A and B are mutually exclusive, then
       $P(A \cup B) = P(A) + P(B)$
- Quick reminder on unions and intersection



Intersections   Unions

  - Use venn diagrams w/ complex problems
- Ex:// $S = \{A, B, C\}$.

  Possible subsets: Φ, A, B, C, AB, AC, BC, ABC (8)

  Define $\mathbb{R}$ for each subset:

  $P(\Phi) = 0$, $P(S) = 1$

  $P(A) = 0.3$
  $P(B) = 0.2$ $\Rightarrow$ $P(C) = 0.5$

- Ex:// $S = \{A, B, C\}$.

  Possible subsets: Φ, A, B, C, AB, AC, BC, ABC (8)

  Define $\mathbb{R}$ for each subset:

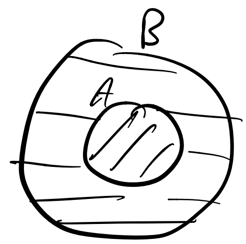
  $P(\Phi) = 0$, $P(S) = 1$

  $P(A) = 1/3$
  $P(B) = 1/3$ $\Rightarrow$ $P(C) = 1/3$

- Classical definition $\subseteq$ axiomatic defn specifically when the probability of each of the events is equal!

LAWS OF PROBABILITY

- If $A \subseteq B \Rightarrow P(A) \leq P(B)$.



Proof: $B = A \cup (B - A)$

$$P(B) = P(A) + \underline{P(B-A)}$$

by axiomatic definition

Extend: $P(B) \geq P(A)$

- Complement: $P(A^c) = 1 - P(A)$ $\Rightarrow$ "at least"

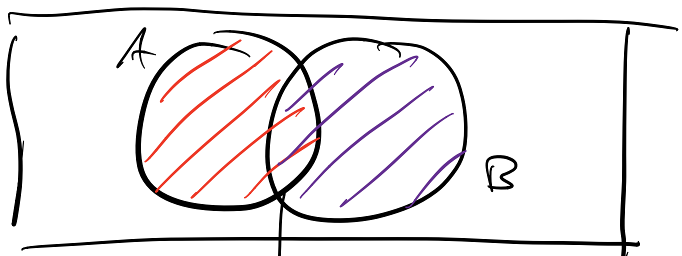
Proof: $S = A \cup A^c$

$$P(S) = P(A) + P(A^c) \quad \text{by ii) in ax. defn.}$$

$$1 = P(A) + P(A^c) \quad \text{by i) in ax. defn.}$$

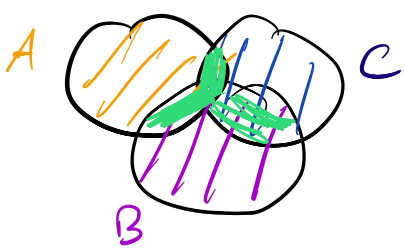
$$P(A^c) = 1 - P(A)$$

- De Morgan's Law: $P(A^c \cup B^c) = P(A \cap B)^c$
- Inclusion exclusion principle: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



Double counting intersection

- Extension for n events:



$P(A \cup B \cup C) =$

$= P(A) + P(B) + P(C) - (P(A \cap B) + P(A \cap C) + P(B \cap C)) + P(A \cap B \cap C)$

Add "odd # of intersections", subtract "even # of intersections"

- Ex:// 2 fair die. P(at least 1 6)

Approach 1: Inclusion - exclusion principle

A_1 : 6 on 1^{st}, A_2 = 6 on 2^{nd} die

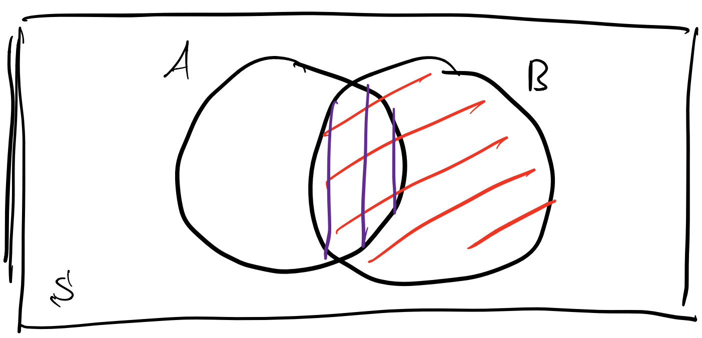
$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$
$$= 1/6 + 1/6 - 1/36$$
$$= 11/36$$

Approach 2: Complement

$$P(\text{at least } 1\ 6) = 1 - P(\text{no } 6)$$
$$= 1 - \left(\frac{5}{6}\right)\cdot\left(\frac{5}{6}\right)$$
$$= 1 - \frac{25}{36}$$
$$= 11/36$$

CONDITIONAL PROBABILITY

- $P(A \cap B) = P(A) \cdot P(B \mid A)$
- $P(X \mid Y)$ = probability of X given Y has occurred.
- Intuition:



$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

- Ex:// 2 dice are rolled. A = sum of 8. B = first roll is 3. $P(A \mid B)$

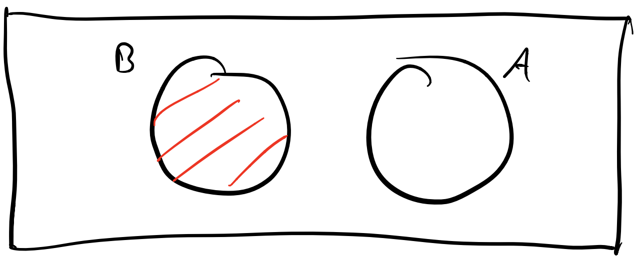
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1/36}{1/6} = 1/6$$

- Ex:// Fair coin tossed 3 times. A: exactly 1 head. B: at least 1 head. $P(A \mid B)$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{3/8}{1 - 1/8} = \frac{3/8}{7/8} = \frac{3}{7}$$

INDEPENDENCE

- A is independent of B if $P(A) = P(A \mid B)$
  - Another way: $P(A \cap B) = P(A)P(B)$ ⟹ use for proofs
- 3 or more events: check that all subset pairs are independent (multiplication rule checks out)
- Mutually exclusive ≠ independent



$$\underbrace{P(A \mid B)}_{0} \neq \underbrace{P(A)}_{\neq 0}$$

- Ex:// Roll 2 dice. A = sum of 8. B = first roll is 3. A indep. of B?

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = 1/6$$

$$P(A) = \frac{\#\text{ of outcomes}}{\text{total}} = 5/36$$

$$P(A) \neq P(A \mid B) \Rightarrow \text{not independent}$$

- Ex://

	$P(A)$
A	0.2
B	0.3
C	0.1

Suppose independent exams. P(exactly 2 A's)?

$$P(\text{exactly 2 A's}) = P(A \cap B \cap C^c) + P(A^c \cap B \cap C) + P(A \cap B^c \cap C)$$

indep.

$$= P(A)P(B)P(C^c) + P(A^c)P(B)P(C) + P(A)P(B^c)P(C)$$

- Ex:// A and B play a game until somebody wins.

$$\left.\begin{aligned} P(A \text{ win}) &= p \\ P(B \text{ win}) &= q \\ P(\text{draw}) &= r \end{aligned}\right\}$$ Assume independent. $P(A \text{ wins})$

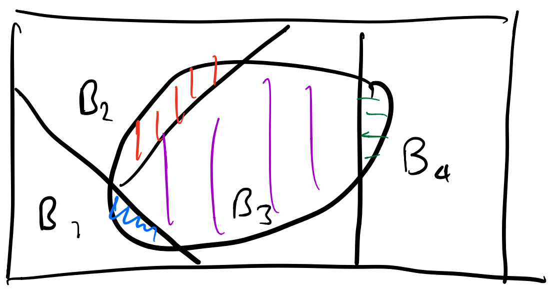
Ways A can win: imm., Draw → A, Draw → Draw → A ...

$$P(A \text{ winning}) = p + \overbrace{r \cdot p}^{\text{indep.}} + r^2 p + r^3 p + \dots$$
$$= p(1 + r + r^2 + \dots)$$
$$= p \cdot \frac{1}{1-r}$$
$$= \frac{p}{p+q}$$

Geometric series ⇓ useful trick.

LAW OF TOTAL PROBABILITY

- Disjoint sets B_i. $P(A) = P(A|B_1) \cdot P(B_1) + \dots + P(A|B_i) \cdot P(B_i)$



⇒ Add $P(A|B_i)$ weighted by $P(B_i)$

Event conditioned on another event? ↓ Law of Total Prob.!

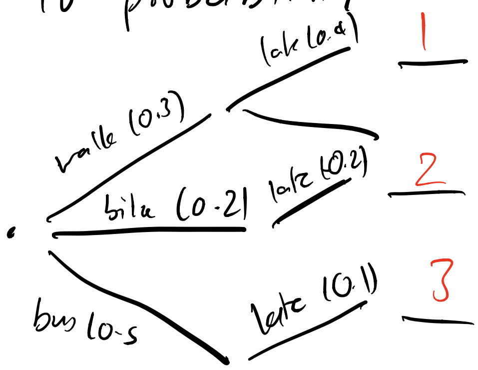
- Ex://

		Late
Walk	0.3	0.4
Bike	0.2	0.2
Bus	0.5	0.1

$P(\text{late})$

$$P(\text{late}) = P(\text{late}|\text{walk}) \cdot P(\text{walk}) + P(\text{late}|\text{bike}) \cdot P(\text{bike}) + P(\text{late}|\text{bus}) \cdot P(\text{bus})$$
$$= (0.4 \times 0.3) + (0.2 \times 0.2) + (0.1 \times 0.5)$$
$$= 0.21$$

- Connection to probability tree: exact same as Law of Total Prob.



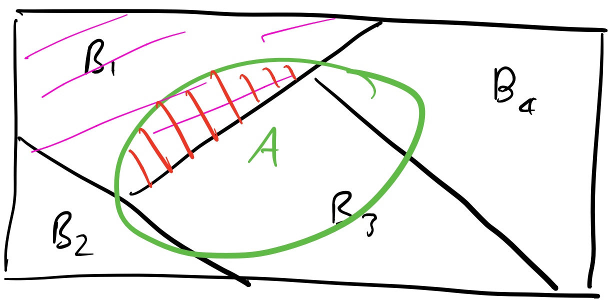
$1 + 2 + 3 = P(\text{late})$

BAYES RULE

- Consider disjoint sets B_i

$$\therefore \quad P(A \mid B_i) = \frac{P(B_i \mid A) \cdot P(A)}{P(B_i)}$$

- Intuition:
  - Visual



  - Proof:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

$$= \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

- Usage: information on A but not on B / changing information.
- Ex:// 1% of pop. has disease. Test: 5% false pos., 10% false neg. rate. Tested positive ⇒ P(disease)?

① Convert into events + notation.

Objective: $P(\text{disease} \mid + \text{test})$

a) $P(\text{disease}) = 0.01$

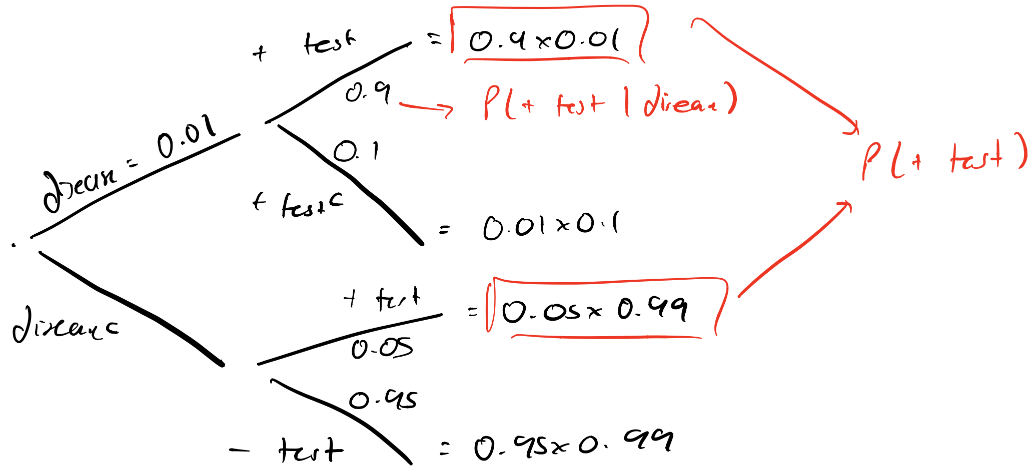
b) $P(+ \text{test} \mid \text{disease}^c) = 0.05$

c) $P(+ \text{test}^c \mid \text{disease}) = 0.1$

② Find out what is needed for Bayes Rule

$$P(\text{disease} \mid + \text{test}) = \frac{P(+\text{test} \mid \text{disease}) \times P(\text{disease})}{P(+\text{test})}$$

③ Probability tree or Law of Total probability:



④ Total prob:

$$P(\text{disease} \mid + \text{test}) = \frac{P(+\text{test} \mid \text{disease}) \times P(\text{disease})}{P(+\text{test})}$$

$$= \frac{0.9 \times 0.01}{0.9 \times 0.01 + 0.05 \times 0.99}$$

– Ex:// 60 Red cards + 40 blue cards. Red: born in J/F/M?. Blue: do you like JB. 0.27 said yes. P(student likes JB).

① Events:

Red / blue, yes/no

② Total probability law

Interest!! ✓

$$P(Y) = P(Y \mid \text{red}) \cdot P(\text{red}) + P(Y \mid \text{blue}) \cdot P(\text{blue})$$

$$0.27 = 0.25 \cdot 0.6 + p \cdot 0.4$$

$$p = \frac{0.27 - 0.25 \cdot 0.6}{0.4}$$

– Ex:// Bag: either has green / blue marble ⇒ prob. is 1/2. Green marble put in bag. Take a random marble out ⇒ green. P(other = green)?

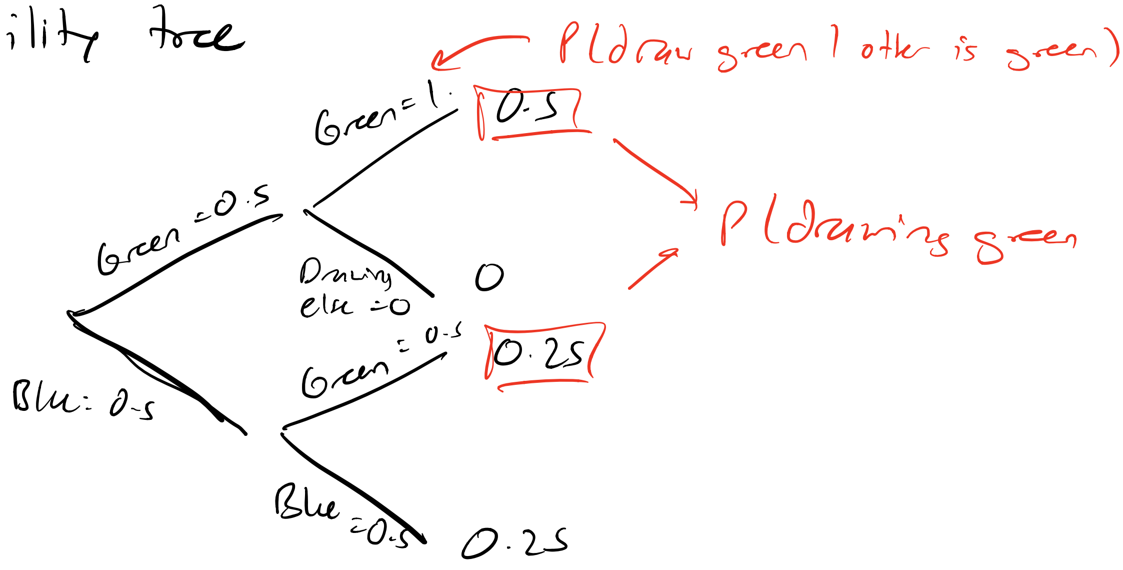
① Events:

Original = green / blue , draw = green / blue.

② Bayes Rule:

$$P(\text{other is green} \mid \text{draw green}) = \frac{P(\text{draw green} \mid \text{other is green}) \cdot P(\text{other is green})}{P(\text{drawing a green})}$$

③ Probability tree



④ Final ans:

$$P(\text{green}) = \frac{1 \cdot 0.5}{0.75} = 2/3$$

RANDOM VARIABLES

- Definition: numerical value for an outcome of an experiment
  - Ex:// Toss a coin 3 times. $X = |\#\text{ heads} - \#\text{ of tails}|$

    Roll a die 2 times. $X =$ sum of faces

Random variables

Discrete: $X \in \mathbb{Z}$

Continuous: $X \in \mathbb{R}$

- Distribution functions of random variables:

$$f(a) = P(X = a)$$

↳ Frequency of occurrences of X

- Ex:// 13 card hand drawn. $X = \#$ of aces. $f(x)$?

① Outcomes:

X
0
1
2
3
4

② Frequency:

X	$f(x)$
0	$f(0)$
1	$f(1)$
2	$f(2)$
3	$f(3)$
4	$f(4)$

$$f(0) = \frac{{}^{48}C_{13}}{{}^{52}C_{13}}$$

$$f(1) = \frac{{}^{4}C_{1} \cdot {}^{48}C_{12}}{{}^{52}C_{13}}$$

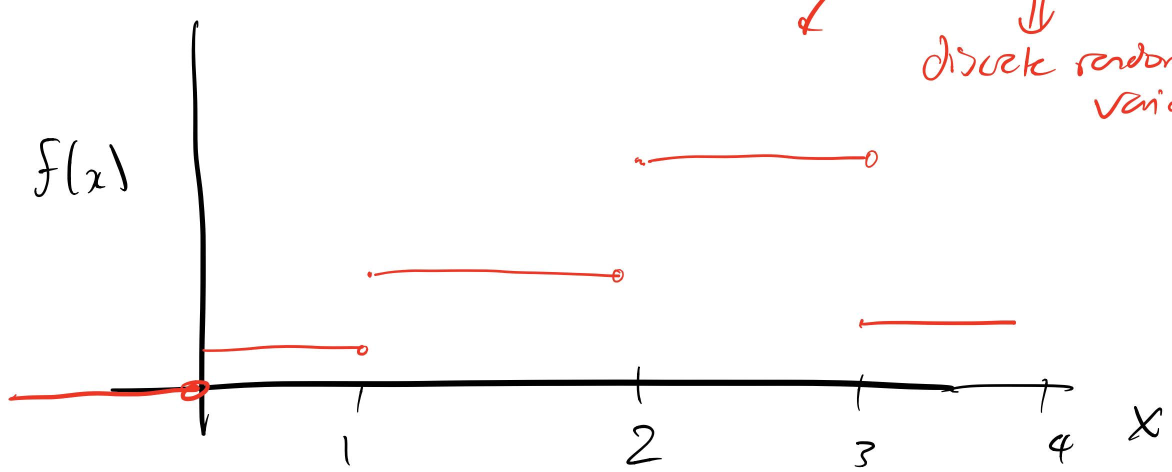
⋮

- Properties:

1. $\forall x,\ f(x) \geq 0 \wedge f(x) \leq 1$

2. $\sum_{x} f(x) = 1$

- Graphically:



Step function ⇓ discrete random variable

CUMULATIVE DISTRIBUTION FUNCTION

- Definition: $F(x) = P(X \leq x)$ ⇒ X can take values up to + including x
- Ex:// Tossing coin 2x. X = # of heads. Find CDF of X.

X	$F(x)$
0	①

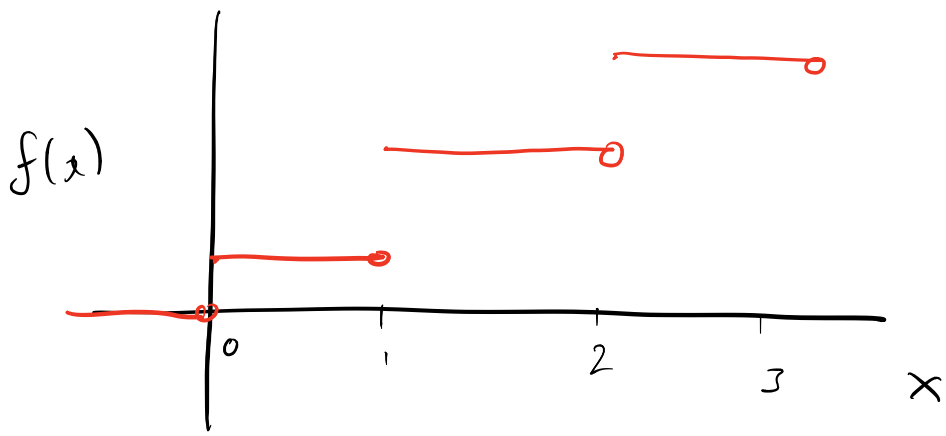
1	②
2	③

① : $P(X \le 0) = 1/4$

② $P(X \le 1) = P(X=1) + P(X=0)$
$= 1/2 + 1/4$
$= 3/4$

③ $P(X \le 2)$
$= 3/4 + 1/4$
$= 1$

- Graphically



⇒ Always going to 1

- Properties:
  1. $0 \le F(x) \le 1$
  2. Non decreasing
  3. Step function
  4. $\lim_{x \to \infty} F(x) = 1$, $\lim_{x \to -\infty} F(x) = 0$

- Converting CDF ⇔ nomal distribution function:

x	f
0	0.2
1	0.3
2	0.5

$\xrightarrow{\text{CDF}}$

x	F
0	0.2
1	0.5
2	1

$\xrightarrow{\text{PDF}}$

x	F
0	0.2
1	0.3
2	0.5

- PDF → CDF: adding prev. cum. sum to original value
- CDF → PDF: subtract differences

- CDF can be easier to calculate:
  - Ex:// 4 dice are rolled. Let $X = \max\{x_1, x_2, x_3, x_4\}$

X	PDF	CDF
1	$(1/6)^4$	$(1/6)^4$
2	challenging to calc.	$(2/6)^4$
3		$(3/6)^4$
4		⋮
5		
6		

$$P(X \le 2) = P(x_1 \le 2, x_2 \le 2, x_3 \le 2, x_4 \le 2)$$
$$= 2/6 \cdot 2/6 \cdot 2/6 \cdot 2/6$$

- Tip: if PDF is hard to calculate, find CDF instead.

EXPECTED VALUE & VARIANCE

- Expected value: average value of random variable over several rep.

$$E(x) = \sum x \cdot f(x) \Rightarrow$$ x is different values of X

- Ex:// Lotto 649. Following distribution of X (payout). Find your expected value:

X	f(x)
-5	0.99
10,000	0.0075
1M	0.0025

$$E(x) = -5(0.99) + 10{,}000(0.0075) + 1{,}000{,}000(0.0025)$$

- Properties:
  1. $E(\alpha) = \alpha$ (α is a constant)
  2. $E(a + bX) = a + bE(X)$
  3. $E(aX + bY)$
     $= aE(X) + bE(Y)$
  4. $E(g(x)) = \sum g(x) \cdot f(x)$

Distribution of $X+Y \neq$ sum of distributions

- Variance: average change of a random variable

$$V(X) = E(\underbrace{X - E(X)}_{\text{Diff. between variable + avg.}})^2$$

- Why square:
  1. Cancel out would occur if not squared
  2. Penalizes larger values
- Standard deviation: $\sqrt{V(x)}$ $\Rightarrow$ same unit as X (σ)
- Ex://

X	$f(x)$
0	1/4
1	1/2
2	1/4

1. Calculate expected value:

$E(X) = 0 \cdot 1/4 + 1 \cdot 1/2 + 2 \cdot 1/4$
$= 1$

2. Calculate squared deviations

X	$f(x)$	$(X - E(X))^2$
0	1/4	1
1	1/2	0
2	1/4	1

3. Calculate variance

$V(X) = 1 \cdot 1/4 + 0 \cdot 1/2 + 1/4 \cdot 1$
$= 1/2$

- Properties:
  1. $V(X) = E(X - \mu)^2 = E(x^2) - \mu^2$ $\Rightarrow$ Average of squares - square of average
  2. $V(\alpha) = 0$

3. $V(a+bX) = b^2 V(X)$

4. $\sigma(a+bX) = |b|\,\sigma(X)$

5. $Var(X+Y) = Var(X) + Var(Y)$ if X and Y are indep.

BINOMIAL DISTRIBUTION

- Following conditions met:
  1. Binary experiments
  2. Constant trial probability
  3. Independent trials
  4. Trials are fixed
  5. X = # of successes in n trials

- Definition:

$$\underbrace{P(X=x)}_{\text{Probability of getting } x \text{ successes in } n \text{ trials}} = \underbrace{{}^{n}C_{x}}_{\text{Choosing "positions" of success}}\ \underbrace{p^{x}}_{\text{Probability of success}}\ \underbrace{(1-p)^{n-x}}_{\text{Probability of failure.}}$$

- Problem solving tip for "at least" in binomial distribution.

$$P(X \geq \alpha) = 1 - P(X < \alpha)$$
$$= 1 - P(X=0) + P(X=1) + \ldots + P(X=\alpha-1)$$

- Bernoulli distribution: $n = 1$

- Ex.// Random walk problem: person takes 20 steps. P(right step) = 0.7, P(left step) = 0.3. P(back at starting point).

  Number of right + left steps must be equal.

  $\therefore$ 10 right steps!

  Let X be the random variable of the number of right steps in 20 steps.

$$P(X=10) = {}^{20}C_{10} \cdot 0.7^{10} \cdot 0.3^{10}$$

- Properties:
  1. If X is binomial $\Rightarrow X \sim Bin(n,p)$
  2. $E(X \sim Bin(n,p)) = np$

3. $V(X \sim Bin(n, p)) = np(1-p)$

GEOMETRIC DISTRIBUTIONS

- Goal: find probability that you get first success on the x^{th} trial.
- Given $X \sim Geom(p)$:

$$P(X = x) = \underbrace{(1-p)^{x-1}}_{\text{probability of failure } (x-1 \text{ times})} \cdot p$$

probability of success

- Exact same conditions as binomial distribution
- Ex:// Need O+ blood $(p = 0.1)$. Probability of testing at least 4 people to find 1 person w/ O+ blood.

Take the complement:

$P(\text{testing at least 4 people}) = 1 - P(\text{testing 3 people})$

$= 1 - P(1 \text{ person}) - \underline{P(2 \text{ people})} - P(3 \text{ people})$

① Define random variable

X = # of people you want to test

② Consider if geometric → proceed:

$P(\text{testing at least 4}) = 1 - 0.1 - (0.9 \cdot 0.1) - (0.9^2 \cdot 0.1)$

NEGATIVE BINOMIAL DISTRIBUTION

- Goal: geometric but finding # of failures before k successes
- Given $X \sim NB(p, k)$ ⇒ p: success prob., k: # of successes

$$P(X = r) = \underbrace{{}^{r+k-1}C_{k-1}\, p^{k-1}(1-p)^r}_{\text{Binomial: find prob. of } k-1 \text{ success before trial}}\; p$$

last trial w/ last success

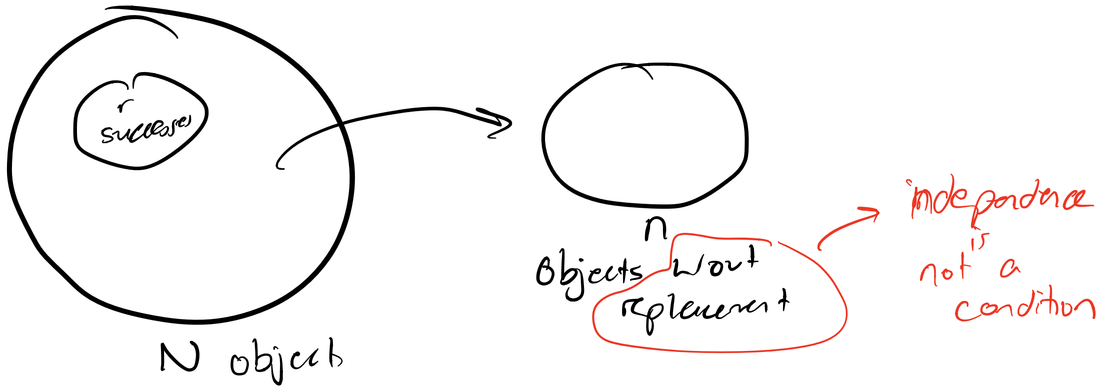
- Ex:// Banach has 20 matchsticks in 2 matchboxes each. Tosses a coin to determine from which matchbox he should draw from. $P(\text{left} = 5 \mid \text{right is empty})$?

Already consumed 35 matchsticks. Success is match stick from right.

$$P(X=5) = {}^{35}C_{20}\, p^{20} (1-p)^{15} \times p$$

HYPERGEOMETRIC DISTRIBUTION

- Goal: binomial ⇒ # of successes. Unique process:



X : # of successes in n objects

- Sampling + resampling are good examples
- Given $X \sim \text{Hyper}(N, r, n)$

$$P(X=k) = \frac{{}^{r}C_{k} \cdot {}^{N-r}C_{n-k}}{{}^{N}C_{n}}$$

Choosing successes ($^{r}C_{k}$)

Choosing failures ($^{N-r}C_{n-k}$)

# of ways to choose n objects ($^{N}C_{n}$)

- Ex:// $P(2 \text{ aces} \mid 13 \text{ card hand})$.

$$X \sim \text{Hyper}(52, 4, 13)$$

$$P(X=2) = \frac{{}^{4}C_{2} \cdot {}^{48}C_{11}}{{}^{52}C_{13}}$$

- $N \to \infty$, $X \sim Hyper(\ldots) \to X \sim Bin(\ldots)$

Poisson Distribution

- Poisson process:
  1. Indivduality: time interval is really small ∧ P(2 events) = 0
  2. Independence: events are independent of each other
  3. Proportionality: time interval ∝ P(event)
- Often, rates are involved w/ Poisson
- Given $X \sim Poi(\mu)$

k successes in an interval

$$P(X = k) = \frac{e^{-\mu}\mu^{k}}{k!}$$

  - μ: # of expected successes in a certain time interval.

$$\mu = \lambda t \Rightarrow$$ Scale rate λ into time interval that we are interested in

- Hints that we are dealing w/ Poisson:
  - No obvious event upper bound for successes
  - Failures don't make sense
  - N (# of intervals) is large but p is small
    ↳ $X \sim Bin(n,p)$. $n \to \infty$, $p \to 0$ but np is constant

$$\therefore Bin(n,p) = Poi(\mu = np)$$

- Ex:// Airline sold 122 seats but 120 seats capacity.
  P(passenger showing) = 0.97. P(more ppl. showing than seats).

  ① Binomial:
  Y = # of no shows

$$P(Y=0) + P(Y=1) = {}^{122}C_0 \times 0.03^0 \times 0.97^{122} + {}^{122}C_1 \times 0.03 \times 0.97^{121}$$

② Poisson:

$$Y \sim Poi(\mu = np) = Poi(\mu = 3.66)$$

$$P(Y=0) + P(Y=1) = \frac{e^{-3.66}\, 3.66^{-0}}{0!} + \frac{e^{-3.66}\, 3.66^{1}}{1!}$$

$$① = ②$$

- Ex:// Rate of earthquakes in 6/year
  - a) P(7 earthquakes in 2 years)
    - ① Identify if situation is Poisson
    - ② Find expected # of events in interested time interval

$$\mu = \lambda \times 2 = 12$$

    - ③ Probability:

$$P(X=7) = \frac{e^{-12}\, 12^{7}}{7!}$$

INDICATOR VARIABLES

- Defn: event A

$$I_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{if } A \text{ doesn't occur} \end{cases}$$

$$\therefore E(I_A) = P(A)$$

- Really useful tool if $E(X) \Leftrightarrow P(X)$
- Ex:// Choose 5 cards from 52 w/out replacement. X = # of aces. What is $E(X)$.

Three options

Pre-computed distribution
$\Downarrow$
No good $E(X)$ formula

Made own distribution
$\Downarrow$
Hard

Indicator variable

$$I_j = \begin{cases} 1 & j^{th} \text{ draw is an ace} \\ 0 & \neg(\quad " \quad) \end{cases}$$

$$E(X) = E\left(\sum_{j=1}^{5} I_j\right)$$

$$= \sum_{j=1}^{5} E(I_j)$$

$$= 5 \times 4/52$$

- Ex:// 20 chocolates → 20 students (equal chance of getting chocolate). Y = # of students w/ at least 1 chocolate. $E(Y)$.

$$I_j = \begin{cases} 1 & j^{th} \text{ student got} \geq 1 \text{ chocolates} \\ 0 & \neg(\quad " \quad) \end{cases}$$

$$E(I_j) = P(\text{chocolate}) = 1 - {}^{20}C_0 \, 0.95^{20}$$

$$\therefore \quad E(Y) = \sum_{i=1}^{20} E(I_j) = 20\left(1 - {}^{20}C_0 \cdot 0.95^{20}\right)$$

MULTIVARIATE DISTRIBUTION

- Distribution that involves 2 discrete variables

$$f(x, y) = P(X = x, Y = y) \Rightarrow \text{Joint distribution}$$

- Ex:// Tossing coin three times. X = # of heads. Y = |# of heads - # of tails|. Find joint distribution

① Ranges of X, Y

$$X = 0, 1, 2, 3$$

$$Y = 1, 3$$

② Construct matrix + define probabilities

X\Y	1	3
0	0	1/8
1	3/8	0
2	3/8	0
3	0	1/8

$\frac{1}{8} \rightarrow P(X=0 \cap Y=3)$

- Marginal distribution: distribution of X and Y given joint distribution

X\Y	Y			
X	a	b	c	4
	d	e	f	5
	g	h	i	6
	1	2	3	

4, 5, 6: Distribution of X

1, 2, 3: Distribution of Y

- Opposite does not work: marginal → joint iff independent
  - ↳ $P(X=x, Y=y) = P(X=x) \cdot P(Y=y)$ if independent

COVARIANCE & CORRELATION

- Covariance: $E\{\underbrace{(X-E(X))}_{\text{deviation of } X} \cdot \underbrace{(Y-E(Y))}_{\text{deviation of } Y}\}$
  - Sign of covariance to determine direction of relationship

$X\uparrow, Y\downarrow \Rightarrow (X-E(X))\cdot(Y-E(Y)) < \underline{0}$

$X\uparrow, Y\uparrow \Rightarrow \quad\quad\quad \text{"} \quad\quad\quad > 0$

- $Cov(X,Y) < 0 \Rightarrow$ inverse relationship. $Cov(X,Y) > 0 \Rightarrow$ direct

- Properties:
  - i) $Cov(X, k) = 0$
  - ii) $Cov(X, X) = Var(X)$
  - iii) $Cov(X, Y) = \boxed{E(XY) - E(X) \cdot E(Y)}$
  - iv) If X and Y are independent: $E(XY) = E(X) \cdot E(Y)$
    ↳ $Cov(X, Y) = 0$
  - v) Units matter

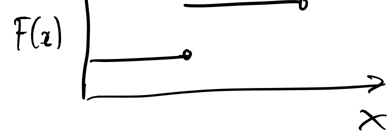
- Correlation coefficients:

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma(X) \cdot \sigma(Y)} \Rightarrow \text{Pearson's}$$

- Sign of ρ_{XY} = sign of $Cov(X, Y)$
- Units are normalized w/ $\sigma(X)$ + $\sigma(Y)$
- Properties:
  - i) $-1 \leq \rho_{XY} \leq 1$ (closer to $\pm 1 \Rightarrow$ stronger relation)
  - ii) Unit independent
  - iii) If X, Y are independent: $\rho_{XY} = 0$
  - iv) If $Y = a + bX$ then $\rho = \begin{cases} 1 & \text{if } b > 0 \\ -1 & \text{if } b < 0 \end{cases}$
- Only measure linear relationships

CONTINUOUS DISTRIBUTIONS

- Defn: X can take all $\mathbb{R}$ values in $[a, b]$
- Comparisons between continuous + discrete distributions
  - CDF:
    - Discrete: $F(x) = P(X \leq x)$



Properties remain same

- Continuous: $F(x) = P(X \le x)$

$F(x) = \int_{-\infty}^{x} f(y)\,dy$

- PDF:
  - Discrete: $f(x) = P(X = x)$
  - Continuous case:

    $P(X = x) = 0 \Rightarrow \lim_{n \to \infty} \frac{1}{n} = 0$

    ↳ Discrete: n is limited

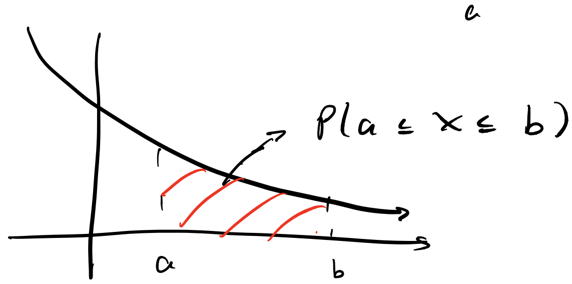
    Can only take probabilities of ranges

    $P(X \ge a)$, $P(X \le b)$, $P(a \le X \le b)$

    B/c $P(X = x) = 0 \Rightarrow P(X \ge x) = P(X > x)$

    Density function is closest match to discrete usage

    $$P(a \le X \le b) = \int_a^b f(x)\,dx$$



$$f(x) \ge 0, \quad \int_{-\infty}^{\infty} f(x)\,dx = 1, \quad f(x) = F'(x)$$

- Expected value:
  - Discrete: $\sum x f(x)$
  - Continuously: $\int_{-\infty}^{\infty} x f(x)\,dx$

- Ex:// $f(x) = kx^3$ for $0 \le x \le 1$. 0 otherwise

a) Find k

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \Rightarrow \int_0^1 kx^3\,dx = \frac{kx^4}{4}\Big|_0^1 = 1 \Rightarrow \frac{k}{4} = 1 \Rightarrow k = 4$$

b) Find CDF:

$$F(x) = \int_{-\infty}^{x} f(y)\,dy = \int_0^x 4y^3\,dy = y^4\Big|_0^x = x^4$$

c) Find $E(x)$:

$$E(x) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^1 x \cdot 4x^3\,dx = \int_0^1 4x^4\,dx = \frac{4}{5}x^5\Big|_0^1 = 4/5$$

d) Find $Var(x)$:

Recall: $Var(x) = E(x^2) - E(x)^2$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^1 x^2 \cdot 4x^3\,dx = 2/3\, x^6\Big|_0^1 = 2/3$$

$$\therefore\ Var(x) = 2/3 - (4/5)^2$$

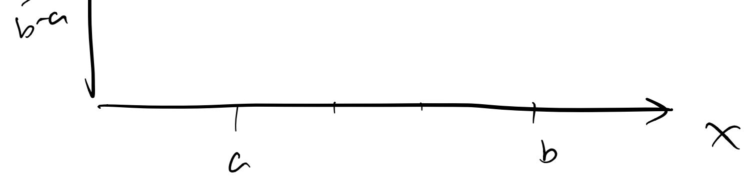
Continuous

Uniform | Exponential | Normal

UNIFORM DISTRIBUTION

- Defn: X is a Uniform distr. along $[a,b]$ if:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases} \Rightarrow \underline{X \sim U[a,b]}$$



- CDF:

$$F(x) = \int_a^x f(y)\,dy = \frac{1}{b-a}\int_a^x dy = \frac{x-a}{b-a}$$

- Special case: $U[0,1] \Rightarrow f(x) = 1, \; F(x) = x$

- $E(X)$: $\frac{b+a}{2}$

- $Var(X)$: $\frac{(b-a)^2}{12}$

- Ex:// Let $X = U[3,5]$. Find $P(X \leq 3.7)$, $P(4 \leq X \; 4.2)$

a) $P(X \leq 3.7)$

*Note: if $P(X \leq a) = F(a)$ *

$$P(X \leq 3.7) = F(3.7) = \frac{3.7-3}{5-3} = \frac{0.7}{2} = 0.35$$

b) $P(4 \leq X \leq 4.2)$

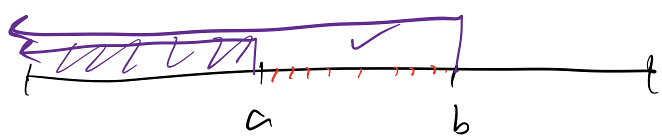
Method #1:

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

$$\therefore P(4 \leq X \leq 4.2) = \int_4^{4.2} \frac{1}{2}\,dx = \frac{1}{2}\cdot(4.2) - 4 = 0.1$$

Method #2:

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$$



$= F(b) - F(a)$

$\therefore P(4 \le X \le 4.2) = \frac{4.2-3}{5-3} - \frac{4-3}{5-3} = 0.1$

- Universality of Uniform:
   - Thm: To find n samples from any distributed r.v. X w/ CDF F and density f, then
     1. Generat $U_1, U_2, \dots, U_n$ from $[0,1]$
     2. $x_n = F^{-1}(U_n)$, $x \dots = F^{-1}(U_{\dots})$
   - Ex:// Let X be a r.v. s.t. $f(x) = e^{-x}$, $x \ge 0$. Find n observations.

① Check density function if it can be used for probability (non-neg + sums to 1)

#1: $e^{-x} \ge 0$ ✓

#2: $\int_0^\infty e^{-x} = 1$ ✓

② Find $F(x)$:

$$F(x) = \int_0^x f(y)\,dy = 1 - e^{-x}$$

③ Use inverse to find relation between X and U

$F^{-1}(U) = X \Rightarrow F(x) = U$

↳ Applied $F(\dots)$ to both sides

$$1 - e^{-x} = U$$

$$e^{-x} = 1 - U$$

$x = -\ln(1-U) \Rightarrow$ Generate n uniform vars to find n observations on X

EXPONENTIAL DISTRIBUTIONS

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \ge 0 \\ 0 & \text{o.w.} \end{cases} \Rightarrow X \sim \text{Exp}(\lambda)$$

- CDF: $F(x) = \int_0^x f(y)\,dy = 1 - e^{-\lambda x}$

- Expectation: $E(X) = \frac{1}{\lambda}$

- Variance: $Var(X) = \frac{1}{\lambda^2}$

- Ex:// $X \sim Exp(\lambda = ½)$.

a) $P(1.5 \le X \le 2.5)$

<u>Method #1</u>

$$P(1.5 \le X \le 2.5) = \int_{1.5}^{2.5} \frac{1}{2} e^{-\frac{1}{2}x}\,dx = \frac{1}{2}\int_{1.5}^{2.5} e^{-\frac{1}{2}x}\,dx$$

$$= \frac{1}{2}\left[2e^{-\frac{1}{2}}\right]_{1.5}^{2.5}$$

$$= \ldots$$

<u>Method #2</u>

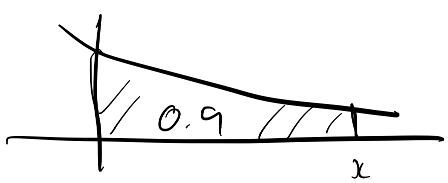
$$P(1.5 \le X \le 2.5) = F(2.5) - F(1.5)$$

$$= \left(1 - e^{-0.5 \cdot 2.5}\right) - \left(1 - e^{-0.5 \times 1.5}\right)$$

b) Find the 90^{th} percentile

* To find percentile:

1. Let x be the value of A^{th} percentile
2. $F(x) = A \Rightarrow A$ is percentile



$$F(x) = 0.9$$

$$1 - e^{-0.5x} = 0.9$$

$$e^{-0.5x} = 0.1$$

$$-0.5x = \ln(0.1)$$

$$x = \frac{\ln(0.1)}{-0.5}$$

- Memoryless: $X \sim Exp(\lambda) \Rightarrow P(X > s+t \mid X > s) = P(X > t)$
  - s has no bearing on $P(X > s+t)$!! $\Rightarrow$ Prior events = no impact
- Use the exponential distribution if modelling wait time between two Poisson events.

NORMAL DISTRIBUTION

- Defn:

$$X \in (-\infty, \infty) \wedge \boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}} \Rightarrow X \sim N(\mu, \sigma^2)$$

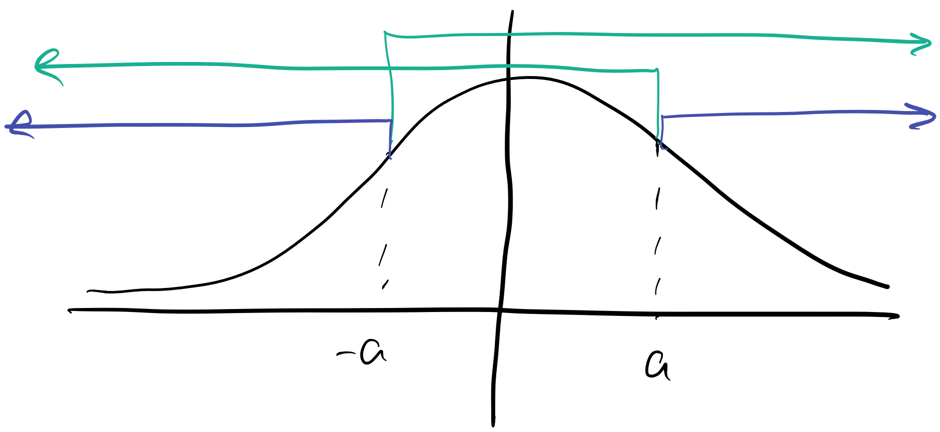
- Problems:
  1. Proofs are very difficult
  2. CDF has no formula:

$$F(y) = \int_{-\infty}^{y} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

- Expectation: $E(x) = \mu$
- Variance: $E(x) = \sigma^2$
- Standard normal: $X \sim N(\mu = 0, \sigma^2 = 1)$ (Z distr.)

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

- Use a Z table to find probabilities.
  - If $P(X \leq 1.23)$ $\Rightarrow$ go to row that corresponds to 1.2, column that corresponds to 0.03
  - Z-table only gives CDF of $\underline{\oplus}$ values.
  - If $P(X \leq -a)$:



* Symmetry *

  - Ex:// a) $P(X \geq -1)$

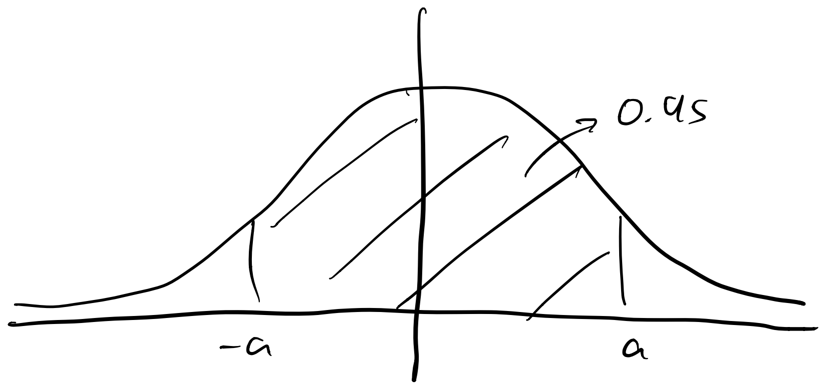
    $P(X \geq -1) = P(X \leq 1)$ $\Rightarrow$ Find 1.00 on Z-table

    b) $P(-1.2 \leq X \leq 0.2)$

    $$P(-1.2 \leq X \leq 0.2) = P(X \leq 0.2) - P(X \leq -1.2)$$
    $$= P(X \leq 0.2) - P(X \geq 1.2)$$
    common!!
    $$= P(X \leq 0.2) - (1 - P(X \leq 1.2))$$

    c) Find a s.t. $P(-a \leq X \leq a) = 0.95$

① Draw a diagram



② General formula

$$P(X \le a) - P(X \le -a) = 0.95$$

$$P(X \le a) - P(X \ge a) = 0.95$$

$$P(X \le a) - (1 - P(X \le a)) = 0.95$$

$$2P(X \le a) - 1 = 0.95$$

$$P(X \le a) = \frac{1.95}{2}$$

↳ Z table

- Problem solving strategy:
  1. Draw diagram
  2. General formula
  3. Convert to CDF:
     a. Range $(a \le X \le b)$: split into $F(b) - F(a)$
     b. Negative: use symmetry
     c. $X \ge a$: $1 - P(X \le a)$
  4. Z table lookup

- Finding probabilities on other normal curves.
  1. Normalize values

$$z^* = \frac{x - \mu}{\sigma}$$

2. Use same standard normal calculations

- Ex:// Grade distribution: $N(70, 100)$

a) $P(X \geq 80)$

$$z^* = \frac{80 - 70}{10} = 1$$

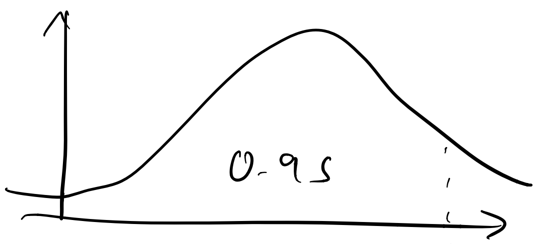
$$\therefore P(Z \geq 1) = 1 - P(Z \leq 1)$$

b) $P(X \geq 65)$

$$z^* = \frac{65 - 70}{10} = -0.5$$

$$P(Z \geq -0.05) = P(Z \leq 0.5)$$

c) Find 95^{th} percentile.



① Let a be valu s.t. area = 0.95

↳ Find this on standard normal via Z lookup

② Put this on grade distribution curve.

$$z^* = \frac{a - \mu}{\sigma}$$

$$z^* \sigma^2 + \mu = a$$

- If $x_1, \ldots, x_n$ are observations from $X \sim N(\mu, \sigma^2)$:

$$\sum x_i = N(n\mu, n\sigma^2)$$

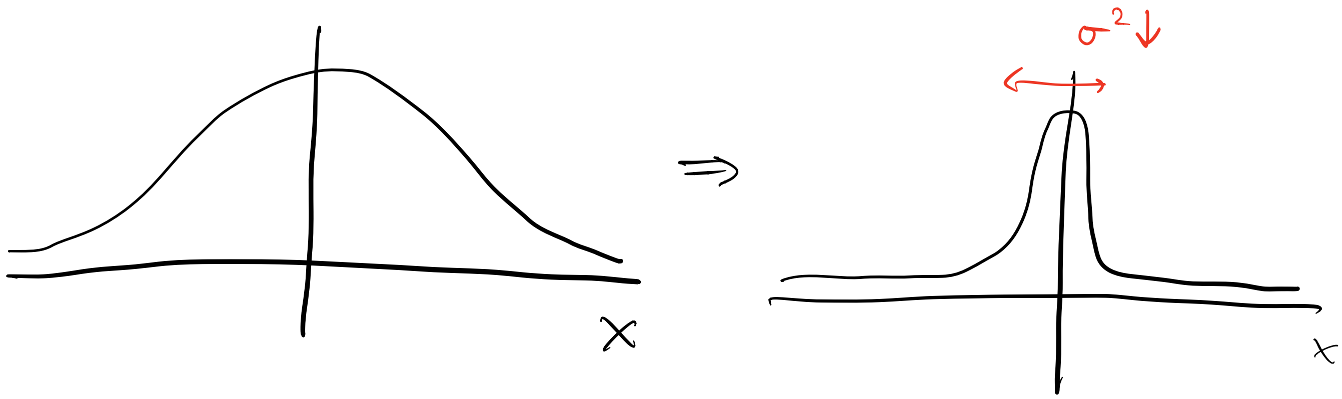
$$\overline{x_i} = N\left(\mu, \frac{\sigma^2}{n}\right)$$

- Laymen's terms: if we were to take n observations, sum/avg, note down + repeat several times ⇒ normal distribution!!

- This can be extended for ANY distribution:
  - $x_1, \ldots, x_n$, indep. + identically distributed (iid).

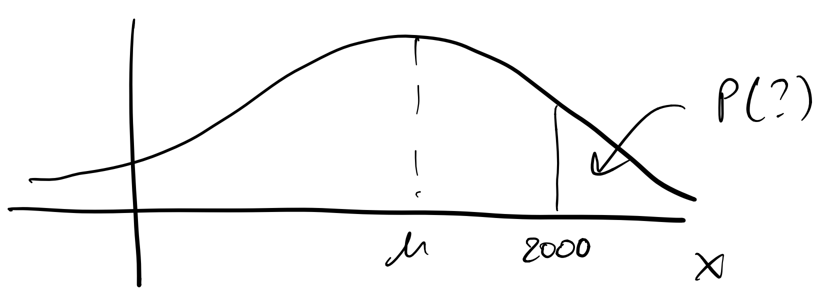
$$\therefore S_n \sim N(n\mu, n\sigma^2), \quad \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

- $\overline{X_n}$:
  - $E(\bar{x}_n) = E\left[\frac{x_1 + \ldots + x_n}{n}\right] = \frac{1}{n} \cdot E(x_1 + \ldots x_n) = \frac{1}{n} n\mu = \mu$
  - $Var(\bar{x}_n) = Var\left(\frac{x_1 + \ldots + x_n}{n}\right) = \frac{1}{n^2} Var(x_1 + \ldots x_n) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$



- Ex:// Max capacity of 2000 kg. 20 people ⇒ weight ~ N(95, 100)
  P(breakdown)?

  ① Diagram



② Distribution of sums of weight

$$X \sim N(1900, 2000)$$

③ Constraints + standardize:

$$z^* = \frac{2000 - 1900}{\sqrt{2000}}$$

④ Use Z table to find probability

$$P(X \geq 2000) = 1 - P\left(z < \frac{2000-1900}{\sqrt{2000}}\right)$$

Law of Large Numbers + Central Limit Theorem

- Law of Large Numbers:

  $X_1, \dots, X_n$ iid with μ, σ^2. As $n \to \infty$, $\bar{X}_n \to \mu$

  - Application: sample averages $\longrightarrow$ population average if n is large

- Central Limit Theorem:

  Regardless of sample distributions, if you take many samples, the sample mean/sum behave normal.

  - Binomial: $X \sim Bin(n, p) \Rightarrow \bar{X} \sim N(np, np(1-p))$
    - Binomial $= \sum$ Bernoulli $\Rightarrow$ CLT
    - Ex:// Coin tossed 1000 times. $p = 0.6$. $X =$ # of successes.

      $P(580 < x < 620)$?

      So many trials $\Rightarrow$ $X \sim N(1000 \cdot 0.6,\ 1000 \cdot 0.6 \cdot 0.4)$

      $\therefore$ roughly normal

  - Poisson: $X \sim Poi(\mu) \Rightarrow \bar{X} \sim N(\mu, \mu/n)$, $S_n \sim N(n\mu, n\mu)$

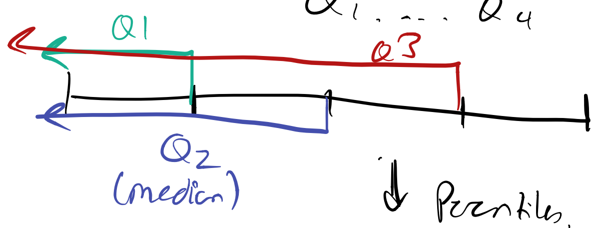
Measures of Location

Data
- Numerical
  - Discrete
  - Continuous
- Categorical
  - Ordinal
  - . . . .

- For numerical:

Measure of Center

- Sample mean: $\bar{y} = \frac{1}{n} \sum y_i$
- Median
  - $\Downarrow$ Resistant to outliers
- Mode
- Quartiles: $Q_1 \dots Q_4$



$\Downarrow$ Percentiles, deciles

Measures of Variability

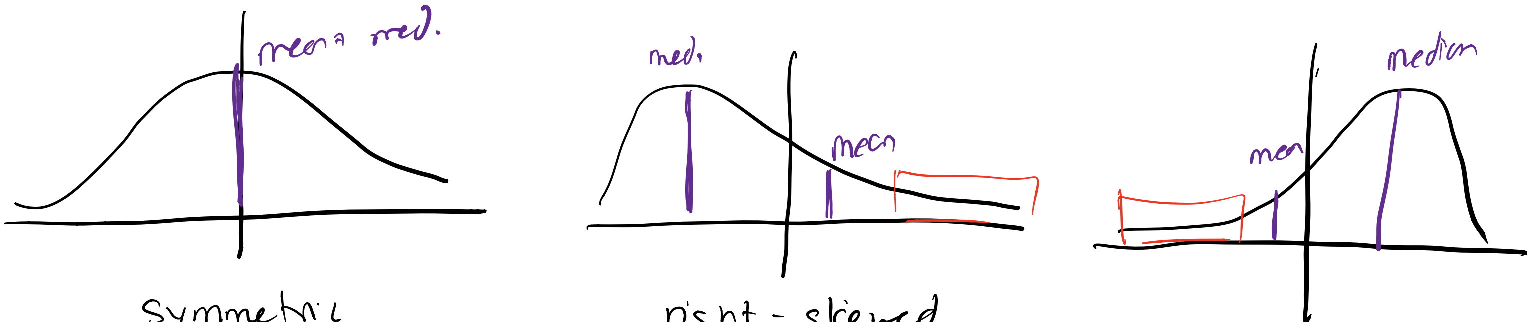
- Sample variance: $s^2 = \frac{\sum (y_i - \bar{y})^2}{n-1}$
  - ↳ n: pop, n-1: sample
- Range: max - min
- Sample s.d.: $s = \sqrt{s^2}$
- IQR: $Q_3 - Q_1$
  - $\Downarrow$ Resistant to outliers.

- Properties for sample mean & variance:

1. $\sum (y_i - \bar{Y}) = 0$ $\Rightarrow$ Deviations sum to 0
2. $s^2 = \frac{\sum (y_i - \bar{y})^2}{n-1} = \frac{1}{n-1} \left(\sum (y_i^2 - n\bar{y}^2) \right)$ Quite useful if you have missing data

Measures of Symmetry, Kurtosis + Association

- Symmetry: mirror image around mean.



symmetric   right-skewed   left-skewed

- Ways to figure out if dist. is symmetric?
    1. Graph it out
    2. $\bar{y}$ - median
        - > 0: right-skewed data
        - < 0: left-skewed data
        - $= 0$: symmetric

- Kurtosis: $E(x-\mu)^4$
    - Used to check if model distribution is correct

- Association:

Associ.

- Numerical
    - Sample corr. ⇓ same as pop. corr. but use sample mean + dev.
    - Sample cov.
- Categorical ("Relative Risk")

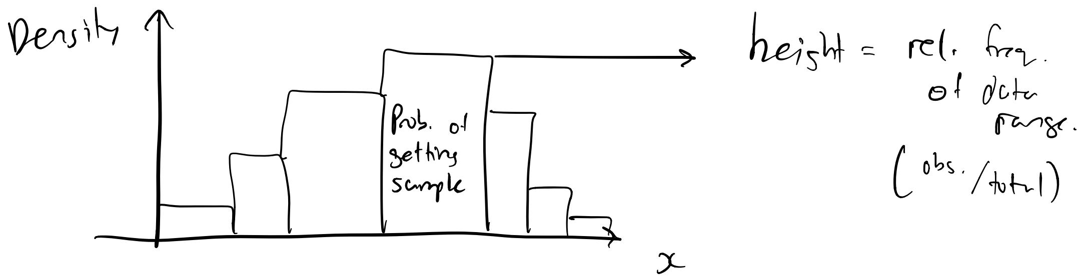
X \ Y	Y1	Y2
X1	n_{11}	n_{12}
X2	n_{21}	n_{22}

$$\Rightarrow R.R = \frac{\frac{n_{11}}{n_{11}+n_{12}}}{\frac{n_{21}}{n_{21}+n_{22}}}$$

Indep. if R.R = 1 ⇒ choice of X has no impact on distribution of Y

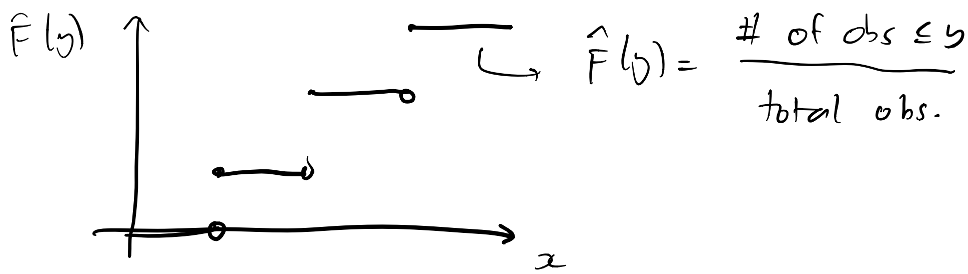
GRAPHICAL SUMMARIES

- Density histogram:

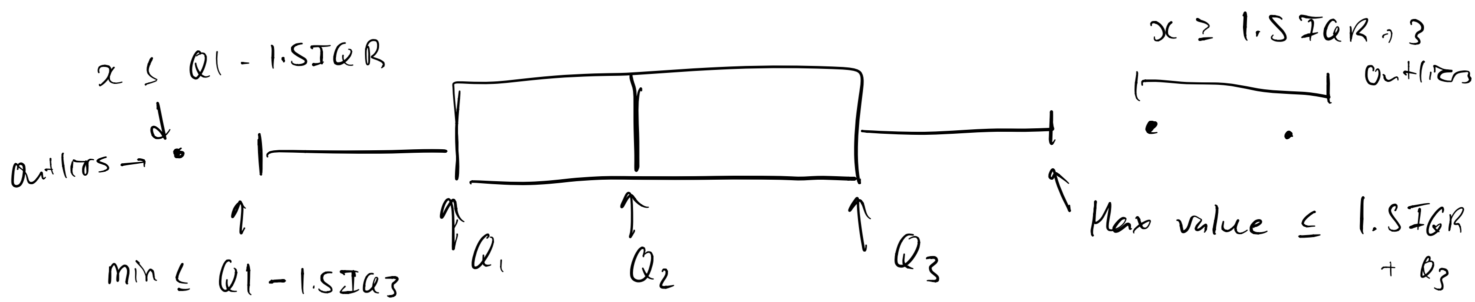


  - Area = 1. Density function of our sample

- Empirical CDF:

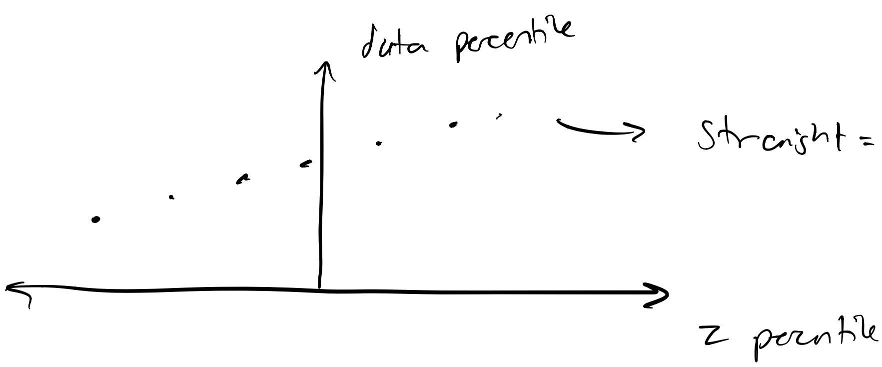


- Box plot:
  - 5 # summary: min, Q_1, Q_2, Q_3, max

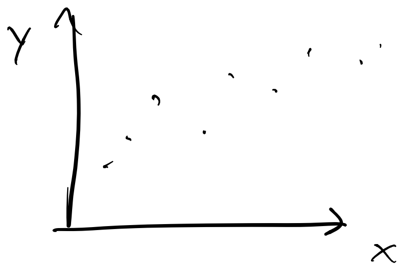


  - Rough sense of distribution

- Q-Q plot



- Scatter plot:



MAXIMUM LIKELIHOOD FUNCTIONS

- Definitions:
  - Target population: pop. you want to find statistic for.
    - Variate: property of population you are studying
    - Attribute: f(variate)
  - Study population: $\subseteq$ target from which you are drawing sample
    - Sample: $\subseteq$ study population. Random.
  - Errors:
    - Study error: target pop. param. − study pop. param ($\mu - \mu'$)
    - Sampling error: ″ − sample pop. statistic ($\mu - \bar{\mu}$)
    - Measurement error: data collection problem
- Goal: given a model + data, find the most likely params. for model
- Also:
  1. $L(\theta) = P(\text{observations}) = \prod f(y_i)$
  2. Log of likelihood
  3. Maximize via derivatives.
- Ex:// Coin tossed 100 times. P(H)?. $y = 70$ heads.

  1. Construct likelihood function

     Coin tosses $\longrightarrow$ binomial distribution

     $$L(\theta) = {}^{100}C_{70}\ \theta^{70}\ (1-\theta)^{30}$$

② Log:

$$\ln L(\theta) = \ln\left({}^{100}C_{70}\right) + 70\ln\theta + 30\ln(1-\theta)$$

③ Derivative:

$$(\ln L(\theta))' = \frac{70}{\theta} - \frac{30}{1-\theta} = 0$$

④ Solve:

$$\theta = 70$$

$\Downarrow$

Maximum likelihood distribution

- Ex:// Poisson model. Data: $\{y_1, \ldots, y_n\}$. MLE for $\hat{\mu}$

① Likehihood:

$$L(\hat{\mu}) = P(Y=y_1) \cdot P(Y=y_2) \cdot \ldots \cdot P(Y=y_n)$$

$$= \frac{e^{-\hat{\mu}}\hat{\mu}^{y_1}}{y_1!} \cdot \ldots \cdot \frac{e^{-\hat{\mu}}\hat{\mu}^{y_n}}{y_n!}$$

$$= \frac{e^{-n\hat{\mu}}\hat{\mu}^{\Sigma y_i}}{y_1! \ldots y_n!} \Rightarrow k$$

② Log:

$$\ln(L(\hat{\mu})) = -n\hat{\mu} + \Sigma y_i \ln(\hat{\mu}) - \ln(k)$$

③ Derivative:

$$(\ln(L(\hat{\mu})))' = -n + \frac{\Sigma y_i}{\hat{\mu}} = 0$$

④ Solve

$$\hat{\mu} = \frac{\Sigma y_i}{n} = \bar{\mu}$$

- If continuous: use density rather than probability

- Ex:// Exponential. Data: $\{y_1, \ldots, y_n\}$. $\hat{\lambda}$?

① Likelihood function:

$$L(\theta) = \prod \hat{\lambda} e^{-\hat{\lambda} y_i} = \hat{\lambda}^n e^{-\hat{\lambda}\Sigma y_i}$$

② Maximize:

$$\ln(L(\theta)) = n\ln(\hat{\lambda}) - \hat{\lambda}\Sigma y_i$$

$$(\ln(L(\theta)))' = \frac{n}{\hat{\lambda}} - \Sigma y_i = 0$$

$$\therefore \quad \hat{\lambda} = \frac{1}{\bar{y}}$$

- Ex:// Normal. Data: $\{y_1, \dots, y_n\}$. $\hat{\mu}$ + $\hat{\sigma}^2$?

① Likelihood function.

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\hat{\sigma}} e^{-\frac{1}{2\hat{\sigma}^2}(y_i - \mu)^2} = \frac{1}{(2\pi)^{n/2}\hat{\sigma}} e^{-\frac{1}{2\sigma^2}\sum(y_i - \hat{\mu})^2}$$

② Maximize:

$$\ln(L(\hat{\mu}, \hat{\sigma}^2)) = -\frac{n}{2}\ln(2\pi) + n\ln(\sigma) - \frac{1}{2\sigma^2}\sum(y_i - \mu)^2$$

Multiple params!

$$\left\{ \begin{array}{l} \frac{\partial}{\partial \mu} = \frac{-2\sum(y_i - \mu)}{2\sigma^2} = 0 \\ \frac{\partial}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2}{2\sigma^3}\sum(y_i - \mu)^2 = 0 \end{array} \right\} \begin{array}{l} \hat{\mu} = \bar{y} \\ \hat{\sigma}^2 = \frac{\sum(y_i - \bar{y})^2}{n} \end{array}$$

Not sample variance

- Ex:// $Y_i \sim U[0, \alpha]$. Data: $\{y_1, \dots, y_n\}$. $\hat{\alpha}$?

① Likelihood function:

$$L(\theta) = \begin{cases} \frac{1}{\alpha^n} & 0 \le y_i \le \alpha \ \forall i \\ 0 & \text{o.w.} \end{cases}$$

$\hookrightarrow \alpha \ge \max(y_1, \dots, y_n)$

② Graph:

$$\therefore \text{MLE for } \hat{\alpha} = \max(\{y_1, \dots, y_n\})$$

χ^2 & T DISTRIBUTION

- χ^2: W is a continuous r.v. takes non-neg. values.

$$W = Z_1^2 + \dots + Z_n^2 \Rightarrow W \sim \chi^2 \ (d.f. = n) \quad (Z_i \sim N(0,1))$$

∘ Properties:

1. $\forall n.\ E(W) = n,\quad Var(W) = 2n$

<u>Proof:</u> Let $W \sim \chi^2(n)$

$$\therefore\ W = Z_1^2 + \dots + Z_n^2$$

$$E(W) = E(Z_1^2 + \dots + Z_n^2)$$
$$= E(Z_1^2) + \dots + E(Z_n^2)$$
$$= 1 \times n$$

Given $X \sim N(0,1)$

$$Var(X) = E(X^2) - E(X)^2$$
$$1 = E(X^2) - 0^2$$
$$\therefore E(X^2) = 0$$

2. $W_1 \sim \chi^2(n_1)$, $W_2 \sim \chi^2(n_2)$ $\wedge$ independent.

$$W_1 + W_2 \sim \chi^2(n_1 + n_2)$$

- Effect of d.f:

i) d.f. = 1 $\Rightarrow$ 1 squared normal variable

Ex:// $W \sim \chi^2(1)$. Find $P(W \leq 1.69)$

$$P(W \leq 1.69) = P(Z^2 \leq 1.69) = P(-1.3 \leq Z \leq 1.3)$$

ii) d.f. = 2 $\Rightarrow W \sim Exp(\lambda = 1/2)$

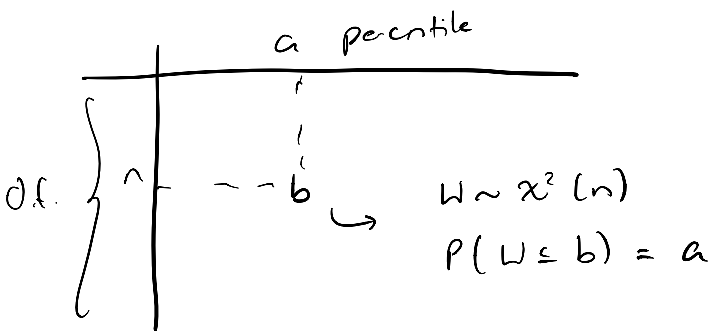
Ex:// $W \sim \chi^2(2)$. Find $P(W \geq 2.5)$

$$P(W \geq 2.5) = 1 - F(2.5)$$
$$= 1 - (1 - e^{-1/2 \cdot 2.5})$$
$$= e^{1.25}$$

Exp cumulative distribution

iii) d.f. $\geq 50 \Rightarrow W \sim N(n, 2n)$

iv) d.f. not in above: χ^2 table (CDF)



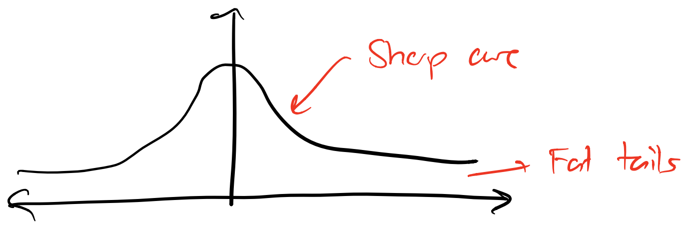
* Symmetry cannot be used *

- Strat:
  1. Find d.f. from data
  2. Use above categories

- T distribution: T is r.v. $\in (-\infty, \infty)$. $T \sim T(\text{d.f.} = n)$ if T is ratio of 2 independent variables.

$$T = \frac{Z \rightarrow N(0,1)}{\sqrt{\frac{W}{n}} \rightarrow \chi^2(n)}$$

- Properties:
  1) Symmetric about 0

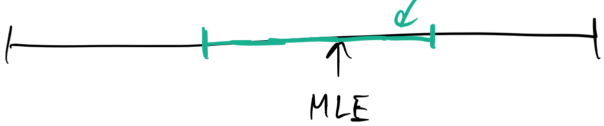


  2) $n \to \infty \Rightarrow T \sim Z$

- Strategy:
  1. Find degrees of freedom
  2. If degrees of freedom $\geq 100 \Rightarrow Z$
  3. Else: use T table (just like χ^2)

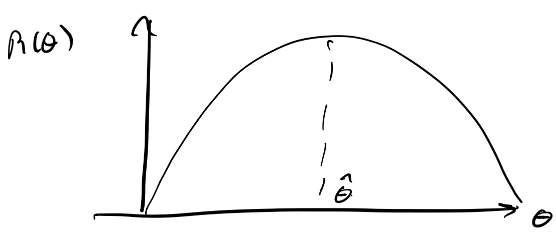
RELATIVE LIKELIHOOD FUNCTION

Estimate interval: x% confident that param. will be in this range.

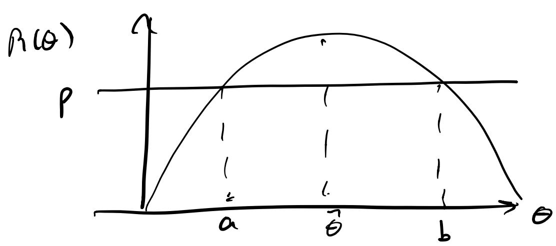


- Relative likelihood function:

$$R(\theta) = \frac{L(\theta) \rightarrow \text{input}}{L(\hat{\theta}) \rightarrow \text{MLE}}$$



- Likelihood interval: p% likelihood interval $\Rightarrow \{\theta : R(\theta) \geq p\}$



$\Rightarrow$ Interval: $[a, b]$

- Classifications:
  1. $R(\theta) \geq 0.5 \Rightarrow \theta$ is v. plausible
  2. $R(\theta) \geq 0.1 \Rightarrow \theta$ is plausible
  3. $0.1 > R(\theta) \geq 0.01 \Rightarrow \theta$ is implausible
  4. $0.01 > R(\theta) \Rightarrow \theta$ is v. implausible

  Subjective + hard to understand

- Ex:// Toss coin 100 times and we get 70 heads. $p = 0.75 \Rightarrow$ plausible?

  ① MLE

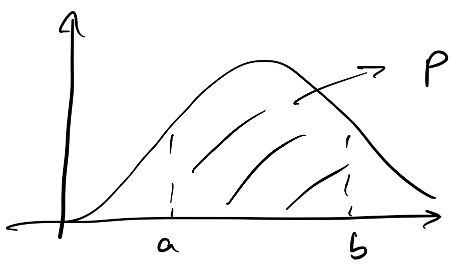
  Binomial MLE = 0.7

  ② Relative likelihood

$$R(0.75) = \frac{L(0.75)}{0.7} = \frac{{}^{100}C_{70}\ 0.75^{70}\ (1-0.75)^{30}}{0.7}$$

CONFIDENCE INTERVAL

- Goal: $[a, b]$ where we are $p\%$ confident that param. in interval



- General strategy: $Y_1, \ldots, Y_n \sim D(\theta)$. We know MLE of θ ($\hat{\theta}$).

  ① MLE for unknown param

  ② Estimate + estimator function

  Estimate: MLE for param + $\hat{\sigma}^2$ for variance

  Estimator function: distribution that outputs estimate

  ↳ To figure this out: CLT

  If $n \to \infty \Rightarrow$ Estimator function $(\bar{Y}) \sim N\left(\hat{\theta}, \frac{\text{var. est.}}{n}\right)$

  ③ Pivotal quantity:

  Standardize: $\dfrac{\bar{Y} - \hat{\theta}}{\sqrt{\frac{\text{var est.}}{n}}}$

④ Probability endpoints.

$$P\left(a \le \frac{\bar{Y} - \hat{\theta}}{\sqrt{\frac{\text{var est.}}{n}}} \le b\right) = p$$ $\Rightarrow$ I want to be p% confident

⑤ Use Z-table to find a and b

⑥ Manipulate prob. expression to get $\hat{\theta}$

$$P\left(\bar{Y} - a\sqrt{\frac{\text{var. est.}}{n}} \le \hat{\theta} \le \bar{Y} + b\sqrt{\frac{\text{var. est.}}{n}}\right)$$

⑦ Use the estimate:

$$C.I. = \hat{\theta} \pm z^* \sqrt{\frac{\text{var est.}}{n}}$$

estimate of our C.I.

- Poisson: param of interest: μ

We want to be p% confidence interval for μ. Data: $\{y_1, \ldots, y_n\}$.

① Model:

$$Y_i \sim \text{Poi}(\mu)$$

② MLE of μ:

$$\hat{\mu} = \bar{y}$$

$$\text{Var}(\text{Poi}(\mu)) = \mu$$

③ Estimate + estimator function

Estimate: $\text{MLE} = \bar{y}$ Estimator: $\bar{y} \in \bar{Y} \sim N\left(\mu, \frac{\mu}{n}\right)$

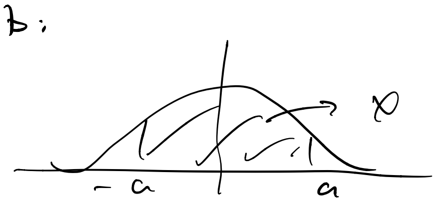
④ Find pivotal quantity:

$$\frac{\bar{Y} - \mu}{\sqrt{\mu/n}} \Rightarrow N(0,1)$$

OR

$$\frac{\bar{Y} - \mu}{\sqrt{\bar{Y}/n}} \Rightarrow N(0,1)$$

⑤ Find endpoints:



⑥ Probability:

$$P\left(-a \le \frac{\bar{Y}-\mu}{\sqrt{\bar{Y}/n}} \le a\right) = p$$

⑦ Solve for estimate:

$$P\left(\bar{Y} - a\sqrt{\bar{Y}/n} \le \mu \le \bar{Y} + a\sqrt{\bar{Y}/n}\right) = p$$

⑧ Estimate + conclude:

$$C.I. = \bar{y} \pm a\sqrt{\frac{\bar{y}}{n}}$$

- Exponential: $f(y) = \frac{1}{\mu} e^{-y/\mu}$

95% C.I. for μ given $\{y_1, \ldots, y_n\}$

① Estimate + estimator:

$$\hat{\mu} = \bar{y} \in \bar{Y} \sim N(\mu, \mu^2/n)$$ $Var(Exp(\mu)) = \mu^2$

② Pivotal quantity:

$$\frac{\bar{Y}-\mu}{\mu/\sqrt{n}} \sim N(0,1)$$

③ Probability + endpoints:

$$P\left(-1.96 \le \frac{\bar{Y}-\mu}{\mu/\sqrt{n}} \le 1.96\right) = 0.95$$

④ Isolate μ:

$$P\left(\frac{\bar{Y}}{1+\frac{1.96}{\sqrt{n}}} \le \mu \le \frac{\bar{Y}}{1-\frac{1.96}{\sqrt{n}}}\right) = 0.95$$

⑤ Estimate + conclude:

$$C.I. = \frac{\bar{y}}{1 \mp \frac{1.96}{\sqrt{n}}}$$

- Normal distribution:

C. I.

We know pop. variance

Use Z distribution for probabilities

We know sample. variance.

1) $$\frac{\bar{Y} - \mu}{S/\sqrt{n}} \sim T_{(n-1)}$$

2) $$\frac{(n-1)\,S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

random var! ↳ Estimate w/ sample var.

- Likelihood ratio test statistic: used if unable to create pivotal
  - $Y_1, \dots, Y_n$ w/ param θ w/ large n:

$$\Delta(\theta) = -2\log\frac{L(\theta)}{L(\hat{\theta})} \sim \chi^2(1)$$

Use this to find endpoints with LTS

$$P\left(a < -2\log\frac{L(\theta)}{L(\hat{\theta})} < b\right) = p$$

- Ex:// 10% likelihood interval $(R(\theta) \geq 0.1)$

$$\frac{L(\theta)}{L(\hat{\theta})} \geq 0.1$$

$$-2\log\frac{L(\theta)}{L(\hat{\theta})} \leq -2\log(0.1)$$

$$P(W \leq -2\log(0.1)) \quad \text{where } W \sim \chi^2(1)$$

$$P(Z^2 \leq -2\log(0.1))$$

$$P(-\sqrt{-2\log(0.1)} \leq Z \leq \sqrt{-2\log(0.1)})$$

- Likelihood → confidence interval

$$P(-1.96 \leq Z \leq 1.96) = 0.95$$

$$P(Z^2 \leq 1.96^2) = 0.95$$

Work backwards

$$P\left(-2\log\frac{L(\theta)}{L(\hat{\theta})} \le 1.96^2\right) = 0.95$$

$$P\left(\frac{L(\theta)}{L(\hat{\theta})} \ge e^{-1.96^2/2}\right) = 0.95$$

Convert to this form

Hypothesis Testing

- Two hypotheses:

  H_0 : null hypothesis (status quo)

  H_1 : alternative hypothesis (different status quo)

- Data ⇒ p-value : probability of observing data given H_0 is true
  - NOT a conditional probability: no cond. prob. needed
  - p-value is low ⇒ evidence against H_0 (0.05)

- Errors:

	Don't reject H_0	Reject H_0
H_0 true	✓	Type 1
H_0 false	Type 2	✓

Normal Distribution Hypothesis Testing

- Algorithm:
  1. State hypotheses
  2. Determine the distribution from which parameter comes from
  3. Construct test statistic : disagreement between data and H_0
  4. Calculate test statistic data (d)
  5. Construct p-value:

$$P(D \ge d)$$

- Ex:// 16 observations $\{Y_1, \ldots Y_{16}\} \sim N(\mu, \sigma^2)$. $\sigma^2 = 49$. $\bar{y} = 77$.

  H_0 : $\mu = 75$

$H_1: \mu \neq 75$

① Distribution of parameter:

$$\bar{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \text{Distribution from which } \bar{y} \text{ is drawn from}$$

$$\bar{Y} \sim N\left(\mu, \frac{\sigma^2}{16}\right)$$

$$\frac{\bar{Y} - \mu}{7/4} \sim Z = N(0, 1)$$

② Construct test statistic:

H_0 is true $\Rightarrow \mu = 75$

$$D = \left| \frac{\bar{Y} - 75}{7/4} \right| \Rightarrow d = \frac{77 - 75}{7/4} = 8/7$$

③ P-value:

$$P(D \geq d) = P(|Z| \geq 8/7)$$

$$= P(Z \leq -8/7 \wedge Z \geq 8/7)$$

$$= 0.26$$

Normal (Sample Variance), Binomial, Poisson

- Ex:// Data: $\{y_1, \ldots, y_{16}\}$. $\bar{y} = 12$, $s^2 = 25$. Test H_0 at 5% significance.

① Hypotheses:

$$H_0: \mu = 10$$

$$H_1: \mu \neq 10$$

② Distribution of parameter:

$$\bar{Y} \sim N\left(\mu, \frac{25}{16}\right) \leftarrow \text{sample}$$

$$\frac{\bar{Y} - \mu}{5/4} \sim T(\text{d.f.} = 15) \leftarrow n-1$$

③ Test statistic:

$$D = \left| \frac{\bar{Y} - 10}{5/4} \right|, \quad d = \frac{12 - 10}{5/4}$$

④ P-value:

$$P(D \geq d) = P(D \leq -8/3,\ D \geq 8/3)$$

$D \sim T(d.f. = 15)$ use that to find probabilities

- Ex:// 1000 votes. A: 475 votes. B: 525 votes. $\theta = P(A \text{ winning})$.

① Hypotheses:

$$H_0: \theta = 0.5$$

$$H_1: \theta \neq 0.5$$

② Distribution:

$$Y \sim Bin(1000, \theta)$$

$$\frac{Y - 1000\theta}{\sqrt{1000\theta(1-\theta)}} = Z \sim N(0,1)$$

If n large, $Bin \sim N$

③ Test statistic:

$$D = \left| \frac{Y - 500}{\sqrt{250}} \right| \sim N(0,1). \qquad d = \left| \frac{575 - 500}{\sqrt{250}} \right|$$

④ P-value:

$$P\left(D \geq \frac{25}{\sqrt{250}}\right) = \ldots$$

- Ex:// Let $Y_1, \ldots, Y_{50} \sim Poi(\mu)$. $\bar{y} = 11.7$

① Hypotheses:

$$H_0: \mu = 10$$

$$H_1: \mu \neq 10$$

② Distribution.

$\bar{y}$ is MLE for μ is Poisson.

By CLT $\Rightarrow$ $\bar{Y} \sim N\left(\mu, \frac{\mu}{n}\right)$

$$\frac{\bar{Y} - \mu}{\sqrt{\mu/n}} \sim N(0,1)$$

③ Test statistic:

$$D = \left| \frac{\bar{Y} - 10}{\sqrt{1/5}} \right| \sim N(0,1) \Rightarrow d = 1.4/\sqrt{1/5}$$

④ Calculate p-value:

$$P\left(-\frac{1.9}{\sqrt{1/5}} \le Z,\ Z \le \frac{1.9}{\sqrt{1/5}}\right)$$

- Ex:// $\{Y_1, \ldots, Y_n\} \sim N(\mu, \sigma^2)$. s_1^2 is sample variance

① Hypotheses:

$H_0: \quad \sigma^2 = \sigma_0^2$

$H_1: \quad \sigma^2 \neq \sigma_0^2$

② Distribution:

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

③ Test statistic:

$$d = \frac{(n-1)s_1^2}{\sigma_0^2}$$

④ P-value:

1. If $d >$ median $(\chi^2_{n-1}) \Rightarrow$ p-val $= 2 \cdot P(D \ge d)$
2. If $d <$ median $(\chi^2_{n-1}) \Rightarrow$ p-val $= 2 \cdot P(D \le d)$

} Check one side of χ^2 and double

LRTS

- LRTS: likelihood test ratio statistic
  - Used if test statistic is hard to construct

$$\Lambda(\theta) = -2\ln\frac{L(\theta_0)}{L(\hat{\theta})} \sim \chi^2(1) \text{ as } n \to \infty$$

- Ex:// $Y \sim Bin(200, \theta)$. $y = 110$ successes.

$H_0: \quad \theta = 0.5$

$H_1: \quad \theta \neq 0.5$

① Calculate MLE for parameter given data.

$$L(\theta) = {}^{200}C_{110}\, \theta^{110} (1-\theta)^{90}$$

⋮

$\hat{\theta} = 0.55$

② Test statistic:

$$\delta(\theta) = -2 \ln \frac{L(0.5)}{L(0.55)} = 2.003$$

③ P-value:

$$
\begin{aligned}
P(\chi^2(1) \geq 2.003) &= P(Z^2 \geq 2.003) \\
&= P(Z \geq \sqrt{2.003},\ Z \leq -\sqrt{2.003}) \\
&= 0.15
\end{aligned}
$$

COMPARING 2 POPULATIONS

- Setup:

$$Y_{11}, \ldots, Y_{1n} \sim N(\mu_1, \sigma_1^2) \qquad Y_{21}, \ldots, Y_{2n} \sim N(\mu_2, \sigma_2^2)$$

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

2 Mean Problems

- Matched Pair
- Unmatched, unequal variances
- Unmatched, Equal variances.

- Matched pair problems:
  - Setup: 1:1 mapping between samples in 2 pop.
    - Twins, before/after, husband/wife

$$Y_{11}, \ldots, Y_{1n} \sim N(\mu_1, \sigma_1^2),$$
$$Y_{21}, \ldots, Y_{2n} \sim N(\mu_2, \sigma_2^2)$$

- Strategy:

1. Declare hypotheses:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

2. Create new random variable with difference between pop.

$$Y_{i1} - Y_{i2} \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2 + 2\,\mathrm{cov}(Y_{i1}, Y_{i2}))$$

Data: $\{Y_{11} - Y_{21}, Y_{12} - Y_{22}, \ldots, Y_{1n} - Y_{2n}\}$

3. Create new hypotheses:

$$H_0: \mu_1 - \mu_2 = 0$$
$$H_1: \mu_1 - \mu_2 \neq 0$$

4. Construct same test statistic for mean test

$$D = \left| \frac{\bar{Y}_{i1} - \bar{Y}_{i2} - \mu_D}{S/\sqrt{n}} \right| \sim T(n-1)$$

Calculate from diff. data

$$d = \left| \frac{\bar{y}_D - 0}{s/\sqrt{n}} \right|$$

5. Construct p-val: $P(|T| \geq d)$

- Unmatched, unequal variances with large sample
  - Strategy:

1. State null hypotheses:

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

2. Construct r.v. for means of 2 populations

$$\bar{Y}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$$

$$\bar{Y}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

3. Construct r.v. for diff. of means:

$$\bar{Y}_1 - \bar{Y}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

4. Construct test statistic:

$$\frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim Z$$

n_1, n_2 are large

$$\frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim Z$$

5. Calculate test statistic

- Unmatched, equal variances.
  - Strategy:

1. State null hypotheses:

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

2. Construct r.v. for means of 2 populations

$$\bar{Y}_1 \sim N\left(\mu_1, \frac{\sigma^2}{n_1}\right)$$

$$\bar{Y}_2 \sim N\left(\mu_2, \frac{\sigma^2}{n_2}\right)$$

3. Construct r.v. for diff. of means:

$$\bar{Y}_1 - \bar{Y}_2 \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

4. Construct test statistic:

$$\frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim Z$$

$$\frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim T_{n_1 + n_2 - 2}$$

5. Calculate sample variance

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

6. Calculate test statistic + p-value.

ONE-SIDED TEST

$$\left.\begin{array}{l} H_0: \theta > \alpha \\ H_1: \theta \leq \alpha \end{array}\right\} \rightarrow$$ No longer about equality

- Strategy: Don't use absolute values when calculating p-values... only looking at 1 tail

Goodness Of Fit Tests

- Objective: does our data match a distribution?
- Setup:

  Data: $Y_1 \dots Y_n$. $Y_i \sim f(y_i, \theta)$, $i = 1 \dots n$

① Hypotheses:

$$H_0: Y_i \sim f(y_i, \theta)$$
$$H_1: !(Y_i \sim f(y_i, \theta))$$

② Contingency table:

Construct m categories

Categories	Observed Freq.	Expected Freq.
Cat. 1		
Cat. 2	$\frac{\text{\# of obs. in cat.}}{\text{total \# of obs.}}$	$n \times P(\text{category } i \mid f(y_i, \theta))$
Cat. 3		
⋮	⋮	
Cat. m		

Figuring out expected probability:

1. MLE for θ / given param. from problem
2. Use that in distribution to get probabilities of each category.

③ Test statistic:

observed freq.

p-val. dep. on categories.

$$\Delta(\theta) = 2 \sum Y_i \ln\left(\frac{Y_i}{E_i}\right) \sim \chi^2_{m-k-1}$$

expected freq.

\# of params. we're guessing

- Ex:// Data: $X_1 \ldots X_n$. H_0: $X_i \sim Poi(\mu)$

① Contingency table:

Expected probability:

1. MLE: $\bar{x} = \hat{\mu}$

2. Probability:

$$P(X = i) = \frac{e^{-\bar{x}} \bar{x}^i}{i!}$$

Categories	Observed freq.	Expected freq.
0	y_0	$\vdots$
1	$\vdots$	$\vdots$
2	$\vdots$	$n \cdot \frac{e^{-\bar{x}} \bar{x}^i}{i!}$
3	$\vdots$	$\vdots$
≥ 4	y_4	$\vdots$

② Test statistic:

$$\lambda = 2 \sum_{i=0}^{4} y_i \ln\left(\frac{y_i}{e_i}\right)$$

③ P-value:

$$P(\Lambda \geq \lambda) = P(\chi^2_{n-k-1} \geq \lambda)$$

$$= P(\chi^2_{5-1-1} \geq \lambda)$$

$$= P(\chi^2_3 \geq \lambda)$$

- Continuous distributions: Construct own categories
  - Probabilities: $P(a \leq f(x) \leq b) = F(b) - F(a)$