WEEK 1

PROBABILITY DEFINITIONS

- probability is math analyzing uncertainty
- random experiment: experiment whose outcomes are unknown
  - ↳ e.g. dice rolls, card draws, measuring IQs, measuring heights
- sample space: set of all possible outcomes in an experiment
  - ↳ e.g. toss a coin twice
    - $S = \{HH, HT, TH, TT\}$
    - # of heads: $S = \{0, 1, 2\}$
- event: any subset of sample space
  - ↳ e.g. event in coin toss: A = at least 1 head

- classical def of probability: $P(A) = \frac{\text{\# of elements in } A}{\text{\# of elements in } S}$
  - ↳ all outcomes in S are equally likely
  - ↳ e.g. fair coin is tossed twice, find probability that there's at least 1 head

    $S = \{HH, HT, TH, TT\}$

    $P(A) = \frac{3}{4}$
  - ↳ e.g. roll 6-sided dice, find P(sum = 8)

    $S = \{(1,1), (1,2), \dots, (1,6)$

    $\vdots \qquad\qquad \vdots \qquad \longrightarrow 36$

    $(6,1), \quad \dots, (6,6)\}$

    $A = \{(2,6), (3,5), (4,4), (5,3), (6,2)\}$

    $P(A) = \frac{5}{36}$
- problems w/ classical def:
  1) logically inconsistent
  2) elements in S may be difficult to count
  3) S needs to be finite

- relative frequency def of probability: P(A) = long-term relative freq of an event
  - ↳ problems include:
    - need infinite # of experiments
- subjective probability (Bayesian): guessing probability of event

COUNTING RULES

- addition rule: if exp can be done n_1 or n_2 diff ways, # of total possibilities is $n_1 + n_2$
  - ↳ OR rule
- multiplication rule: if exp A can be done n_1 ways & exp B can be done n_2 ways, # of total possibilities is $n_1 \cdot n_2$
  - ↳ AND rule
- n^k rule (sampling w/ replacement): k objects can be sampled w/ replacement from n objects in n^k total ways
  - ↳ e.g 4-digit ATM codes

    $\underline{10} \times \underline{10} \times \underline{10} \times \underline{10} = 10^4$ possibilities
  - ↳ e.g. ATM code start w/ 8 & end w/ 2 (assuming pwd bias doesn't exist)

    S: 10^4 total $\qquad P(A) = \frac{10^2}{10^4} = \frac{1}{10^2} = \frac{1}{100}$

    A: $1 \times \underline{10} \times \underline{10} \times 1 = 10^2$

↳ e.g. ontario license plates (4 letters 3 #s)

$26^4 \times 10^3$

SAMPLING WITHOUT REPLACEMENT

- **permutations**: suppose k things are selected from n objects w/o replacement & order does matter
  - ↳ # of ways is $\frac{n!}{(n-k)!} = n^{(k)} = {}_nP_k$
- **factorial**: $n! = n(n-1)(n-2)\ldots 3 \cdot 2 \cdot 1$
  - ↳ $0! = 1$
- e.g. create a 4-letter word w/ A, B, C, D; how many ways can all of letters be arranged w/o replacement?

  $4P4 = 4!$

  $= 24$

- **combinations**: suppose k things are selected from n objects w/o replacement & order does not matter
  - ↳ # of ways is $\frac{n!}{k!(n-k)!} = \binom{n}{k} = {}_nC_k$
  - ↳ # of possible subsets of size k that can be selected from set of n objects
- proof of combination formula ${}_nC_k = \frac{n!}{(n-k)!k!}$
  - ↳ ${}_nC_k \cdot k! = {}_nP_k$

    ${}_nC_k = \frac{{}_nP_k}{k!}$

    $= \frac{n!}{k!(n-k)!}$
- e.g. suppose 7 ppl were to form committee of 3 ppl : {A, B, C, D, E, F, G} ; find P(A is excluded)

  Total : ${}_7C_3$

  $= \frac{7!}{(7-3)!\,3!}$

  $= 35$

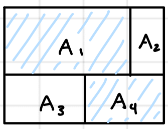
  # of ways to exclude A : ${}_6C_3$

  $= 20$

  $P = \frac{20}{35}$

  $= \frac{4}{7}$

LAWS OF PROBABILITY

- **probability function P**
  - ↳ domain : all possible events
  - ↳ range : $\mathbb{R}$
- P satisfies these **axioms**:
  1) $P(S) = 1$; $P(\emptyset) = 0$
  2) $A_1, A_2, \ldots$ are disjoint events then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$
     - 

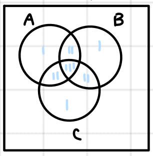
       ← entire square = S
     - $P(A_1 \cup A_4) = P(A_1) + P(A_4)$
- **subsets**: suppose there's events A & B, if $A \subseteq B$ then $P(A) \leq P(B)$
  - ↳ proof: if $A \subseteq B$, then $B = A \cup (B \setminus A)$ ← in B, not A

    $P(B) = P(A) + P(B \setminus A)$ by axiom 2
- **complements**: $P(A^c) = 1 - P(A)$
  - ↳ e.g. black cards in stack (52 cards in total) ; red cards are complement
  - ↳ proof: $P(A \cup A^c) = P(A) + P(A^c)$ ← by axiom 2

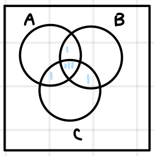
$$P(A \cup A^c) = P(S) = 1$$
$$1 = P(A) + P(A^c)$$
$$P(A^c) = 1 - P(A)$$

- Inclusion - Exclusion Principle : given 2 events, A & B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - ↳ if $A \cap B = \emptyset$, then I-E Principle is restating axiom 1
  - ↳ 3 events : $P(A \cup B \cup C) = P(A) + P(B) + P(C) - (P(A \cap B) + P(A \cap C) + P(B \cap C)) + P(A \cap B \cap C)$

    - $P(A) + P(B) + P(C)$



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



- e.g. 2 fair die are rolled, what's P(at least one is a 6)?

A_1 = 1st roll is 6 $\quad P = P(A_1) + P(A_2) - P(A_1 \cap A_2)$ $\quad$ Alt : $P = 1 - P(\text{no } 6)$ ← complement

A_2 = 2nd roll is 6 $\quad = \frac{1}{6} + \frac{1}{6} - \frac{1}{6} \cdot \frac{1}{6}$ $\quad = 1 - \frac{5}{6} \cdot \frac{5}{6}$

$= \frac{11}{36}$ $\quad = \frac{11}{36}$

- e.g. let S = 1, 2, 3, ..., 150; find # of sets A, where A is set of elements in S that are relatively prime to 70

$70 = 2 \cdot 5 \cdot 7$ $\quad (A + B + C) - (A \cap B + B \cap C + A \cap C) + (A \cap B \cap C)$

A = divisible by 2 = 75 $\quad = (75 + 30 + 21) - (15 + 4 + 10) + 2$

B = divisible by 5 = 30 $\quad = 126 - 29 + 2$

C = divisible by 7 = 21 $\quad = 99$

- e.g. random 13 card hand is dealt from deck of 52; find P(at least 1 suit isn't in dealt hand)

C = no clubs $\quad P = P(C) + P(H) + P(D) + P(S) - (P(C \cap H) + \ldots) +$

H = no hearts $\quad (P(C \cap H \cap D) + \ldots) - (P(C \cap H \cap D \cap S))$

D = no diamonds $\quad = 4 \cdot \frac{39C13}{52C13} - 6 \cdot \frac{26C13}{52C13} + 4 \cdot \frac{13C13}{52C13} - 0$

S = no spades $\quad \simeq 0.051066$

week 2

CONDITIONAL PROBABILITY

· how to update probabilities as new info is gained

↳ e.g. suppose A = stock price

$P(A) = 0.3$

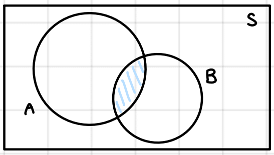
B = company A missed all its targets in a quarter

$P(A|B) = 0.1$ (A given B)

· conditional probability : let A & B be 2 events st $P(B) > 0$, then $P(A|B) = \frac{P(A\cap B)}{P(B)}$

↳ multiply rule: $P(A\cap B) = P(A|B) \cdot P(B)$

↳ A given B is the event A happens given B occurred

↳ 

- $P(A)$ is unconditional so $P(A) = \frac{\text{area } A}{\text{area } S}$
- if B happens, there's a new universe so $P(A\cap B) = \frac{\text{area}(A\cap B)}{\text{area}(B)}$

·

Example
Two cards are drawn from a 52-card deck one at a time without replacement. Find $P(A\|B)$ and $P(B\|A)$ if events A and B are defined as: $A =$ First card is a heart $B =$ Second card is red

$P(A|B) = \frac{P(AB)}{P(B)}$

$P(AB) = \frac{13 \cdot 25}{52 \cdot 51} = \frac{25}{4 \cdot 51}$

$P(A) = \frac{1}{4}$

$P(B) = \frac{1}{2}$ ← imagine looking at 2^{nd} card first

$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{25/4 \cdot 51}{1/2} = \frac{25}{102}$

$P(B|A) = \frac{P(A\cap B)}{P(A)} = \frac{25/4 \cdot 51}{1/4} = \frac{25}{51}$

· prosecutor's fallacy: $P(A|B) \neq P(B|A)$

chronological order is irrelevant

STATISTICAL INDEPENDENCE

· 2 events A & B are independent iff $P(A\cap B) = P(A) \cdot P(B)$

↳ e.g. A: grade > 90% in course

B = will rain in Waterloo tomorrow

· if A & B are independent, then these are independent

↳ A^c & B

↳ A & B^c

↳ A^c & B^c

- proof :
$$\begin{aligned} P(A^c \cap B^c) &= P((A\cup B)^c) \\ &= 1 - P(A\cup B) \\ &= 1 - (P(A) + P(B) - P(A\cap B)) \\ &= 1 - P(A) - P(B) + P(A) \cdot P(B) \\ &= 1 - P(A) - P(B)(1 - P(A)) \\ &= (1 - P(A))(1 - P(B)) \\ &= P(A^c)P(B^c) \end{aligned}$$

- 2 events A & B are mutually exclusive iff $P(A \cap B) = 0$
  - ↳ independent events cannot be mutually exclusive & vice versa
- independence of n-events: events A, B, & C are independent if:
  - ↳ $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$
  - ↳ each set of pairwise events is independent
- e.g.

Example

Consider two fair coin tosses with events A, B, and C defined as:
$A =$ First toss is a head
$B =$ Second toss is a head
$C =$ both tosses are the same

A, B are indep $\quad \{HH, HT, TH, TT\}$

A, C are indep $\rightarrow P(A \cap C) = \frac{1}{4}$

$P(A) \cdot P(C) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

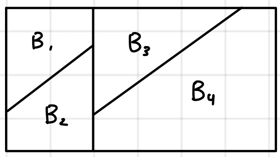
B, C are indep

$P(A \cap B \cap C) = \frac{1}{4} \qquad P(A) \cdot P(B) \cdot P(C) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$

3 events are dependent

LAW OF TOTAL PROBABILITY

- for $n = 3$, $P(A_1 A_2 A_3) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_2 A_1)$
- $P((A \cup B) | C) = P(A|C) + P(B|C) - P((A \cap B) | C)$
- Law of Total Probability theorem: let $S = \sum_{k=1}^{n} B_k$, where $B_i \cap B_j = \emptyset$ if $i \neq j$, then for any event A, $P(A) = \sum_{k=1}^{n} P(A | B_k) \cdot P(B_k)$



- ↳ i.e. if all mutually exclusive probabilities are added up, total probability can be calculated
- ↳ proof: observe that we can write any set $A = \cup_{i=1}^{n} (A \cap B_i)$

$$P(A) = P\left(\bigcup_{i=1}^{n} (A \cap B_i)\right)$$
$$= \sum_{i=1}^{n} P(A \cap B_i) \rightarrow \text{axiom 2}$$
$$= \sum_{i=1}^{n} P(A | B_i) \cdot P(B_i)$$

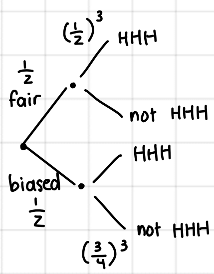
- e.g.

Example

You have one fair coin and a biased coin that lands on heads with a probability of $\frac{3}{4}$. If you pick a random coin and toss it three times, find the probability that all three tosses are heads.

A = all 3 coin tosses are heads

B_1 = fair coin is picked
B_2 = biased coin is picked
} mutually exclusive



$$P(A) = P(A | B_1) \cdot P(B_1) + P(A | B_2) \cdot P(B_2)$$
$$= \left(\tfrac{1}{2}\right)^3 \cdot \tfrac{1}{2} + \left(\tfrac{3}{4}\right)^3 \cdot \tfrac{1}{2}$$
$$= \frac{35}{128}$$

BAYES' RULE

- Bayes' Theorem: let $B_1, \ldots, B_n$ be partition of S & A be any event, then

$$P(B_i | A) = \frac{P(A|B_i) \cdot P(B_i)}{P(A)} = \frac{P(A|B_i) \cdot P(B_i)}{\sum_{k=1}^{n} P(A|B_k) \cdot P(B_k)}$$

↳ $P(B_i) \rightarrow$ prior probability

↳ $P(B_i \mid A) \rightarrow$ posterior probability

proof:
$$P(B_i \mid A) = \frac{P(B_i A)}{P(A)} = \frac{P(A B_i)}{P(A)} = \frac{P(A \mid B_i) P(B_i)}{P(A)} = \frac{P(A \mid B_i) P(B_i)}{\sum_{k=1}^{n} P(A \mid B_k) \cdot P(B_k)}$$

cond probability

LOTP

· e.g.

Example

You have one fair coin and a biased coin that lands on heads with a probability of $\frac{3}{4}$. A coin is chosen at random and tossed three times. If we observe three heads in a row, what is the probability that the fair coin was chosen?

$B_1 =$ The fair coin was chosen.
$B_2 =$ The biased coin was chosen.
$A =$ 3 heads observed in 3 tosses.

$P(B_1) = P(B_2) = \frac{1}{2}$ $\qquad$ $P(A \mid B_1) = (\frac{1}{2})^3$

$P(B_1 \mid A) = ?$ $\qquad$ $P(A \mid B_2) = (\frac{3}{4})^3$

$$P(B_1 \mid A) = \frac{P(A \mid B_1) P(B_1)}{P(A \mid B_1) P(B_1) + P(A \mid B_2) P(B_2)} = \frac{(\frac{1}{2})^3 (\frac{1}{2})}{(\frac{1}{2})^3 (\frac{1}{2}) + (\frac{3}{4})^3 (\frac{1}{2})} = \frac{8}{35}$$

· e.g.

Example

A person is getting tested for a disease that affects 1% of the population.

A test is 95% accurate in true positivity. (sensitivity)
A test is 95% accurate in true negativity. (specificity)

If a person tests positive, then what is the probability that they have the disease?

D = has disease $\qquad$ T = test pos

D^c = doesn't have disease $\qquad$ T^c = test neg

$P(D) = 0.01$ $\qquad$ $P(T) = 0.95$

$P(D^c) = 0.99$ $\qquad$ $P(T^c) = 0.95$

$P(D \mid T) = ?$

$$P(D \mid T) = \frac{P(T \mid D) \cdot P(D)}{P(T \mid D) \cdot P(D) + P(T \mid D^c) \cdot P(D^c)} = \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.05 \cdot 0.99} \simeq 0.161$$

WEEK 3

RANDOM VARIABLES

- X is a random variable (rv) if it's a function that maps each outcome of a random experiment to an element in $\mathbb{R}$
  - ↳ domain: set of outcomes in S
  - ↳ range: real line
  - ↳ e.g. roll die twice, X = sum of faces, is X rv?

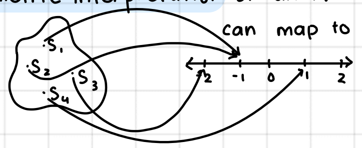
maps 2 ↖ $(1,1)$ . . $(6,1)$ ↗ maps 7

$\vdots \qquad \vdots$ } each element in S is mapped to $\in \mathbb{R}$ so X is rv

$(1,6)$ . . . $(6,6)$

maps 7 ↙ ↘ maps 12

  - ↳ notation: $P(X = x) \rightarrow P(X = 12)$
- experiment is random but X is not
  - ↳ input is random but output behaviour is known
- geometric interpretation of an rv
  - ↳ can map to same value



- discrete rv take on integer values
- continuous rv take on real values
  - ↳ e.g measuring heights of people ($\mathbb{Z}$ is not sufficient)
- e.g. flip coin 3 times

$X = |\#\text{ heads} - \#\text{ tails}|$

$\{HHH, HHT, HTT, HTH, \ldots, TTT\}$

X: 3, 1, 1, 1, 3

PROBABILITY MASS FUNCTION

- Probability Mass Function (pmf): $f(x)$ is pmf of X if $f(x) = P(X = x)$
  - ↳ e.g. roll a die twice, X = sum of faces

X	f
2	$\frac{1}{36}$
3	$\frac{2}{36}$
4	$\frac{3}{36}$
$\vdots$	$\vdots$
12	$\frac{1}{36}$

} pmf

- suppose X is rv w/ pmf f, then:
  - ↳ $f(x) \geq 0 \ \forall x$
    - $f(x) \geq 0$ for all values of x
  - ↳ $\sum_{x \in X} f(x) = 1$
    - probabilities add up to 1
- e.g. random bridge hand (13 cards), X = # of Aces, find pmf of X

X	f
0	$\binom{4}{0}\binom{48}{13}/\binom{52}{13} = \ldots$
1	$\binom{4}{1}\binom{48}{12}/\binom{52}{13} = \ldots$
2	.
3	.
4	

} sum = 1

- rv is function that maps outcomes of random experiment to real #s
- pmf specifies how often each value of rv occurs
  - ↳ e.g. suppose X has pmf f, $X^2 \to f^2$ does not work
- e.g. toss coin twice w/ X = # heads & Y = # tails

X	f
0	$\frac{1}{4}$
1	$\frac{1}{2}$
2	$\frac{1}{4}$

Y	g
0	$\frac{1}{4}$
1	$\frac{1}{2}$
2	$\frac{1}{4}$

  - ↳ 2 diff rvs can have same pmf

CUMULATIVE DISTRIBUTION FUNCTION

- cdf $F(x) = P(X \leq x)$
  - ↳ e.g.

X	f	F
0	0.2	0.2
1	0.5	0.7
2	0.3	1

  - ↳ sum of all probabilities for $X \leq$ current x
- e.g.

X	F	f
0	0.2	0.2
1	0.7	0.5
2	1	0.3

  - ↳ pmf → cdf : $\sum_{\text{if } y \leq x} f(y) = F(x)$
  - ↳ cdf → pmf: $f(x) = F(x) - F(x-1)$
- sometimes, w/ continuous functions, F exists & f does not
- properties of cdf:
  - ↳ $F(X) \geq 0$, non-dec
  - ↳ $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$
  - ↳ F is step function in discrete case
- e.g. 4 die rolled simultaneously, suppose rv $X = \max\{X_1, X_2, X_3, X_4\}$, find pmf of X

X_1	f
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

$X_1, X_2, X_3,$ & X_4 are independent & identically distributed

$$F(1) = P(X \leq 1)$$
$$= P(\max(X_1, X_2, X_3, X_4) \leq 1)$$
$$= P(X_1 \leq 1) \cdot P(X_2 \leq 1) \cdot P(X_3 \leq 1) \cdot P(X_4 \leq 1)$$
$$= (\tfrac{1}{6})^4$$
$$f(1) = (\tfrac{1}{6})^4$$
$$F(2) = P(X \leq 2)$$
$$= (\tfrac{1}{3})^4$$
$$f(2) = (\tfrac{1}{3})^4 - (\tfrac{1}{6})^4$$

EXPECTATION OF RANDOM VARIABLES

- expectation: let X be rv w/ pmf, f(x), then $\mu = E[X] = \sum_{x \in X} x \cdot f(x)$
  - ↳ e.g.

X	f
0	0.2
100	0.8

$$\mu = E[X]$$
$$= 0 \cdot 0.2 + 100 \cdot 0.8$$
$$= 80$$

  - ↳ avg of an rv
- expectations have linearity
  - ↳ $E(X+Y) = E(X) + E(Y)$
  - ↳ for any constant α, $E(\alpha X) = \alpha E(X$

↳ $E(aX+bY) = aE(X) + bE(Y)$

avg of sum of rvs

sum of avgs

- X & Y don't need to be related

- $E[g(X)] = \sum_{x \in X} g(x) \cdot f_X(x)$
  - ↳ $f_X(x)$ is pmf of X
  - ↳ e.g.

X	f
0	0.2
1	0.5
2	0.3

$E(X^2) = 0^2 \cdot 0.2 + 1^2 \cdot 0.5 + 2^2 \cdot 0.3$

$= 1.7$

VARIANCE OF RANDOM VARIABLES

- e.g. The St. Petersburg Paradox

Example

A fair coin is tossed until a head appears. The casino will pay:

$2 if a head appears on the first toss.

$4 if a head appears on the second toss.

$8 if a head appears on the third toss.

⋮

a) How much will you be willing to pay to play this game?

b) What is the expected payoff of this game?

X = winnings

X	f
2	$\frac{1}{2}$
4	$\frac{1}{4}$
8	$\frac{1}{8}$
16	$\frac{1}{16}$
⋮	

$E(X) = 2\left(\frac{1}{2}\right) + 4\left(\frac{1}{4}\right) + 8\left(\frac{1}{8}\right) + \ldots$

$= 1 + 1 + 1 + \ldots$

$E(X) \to \infty$ (avg earnings)

- $Var(X) = E[X - E(X)]^2 = E(X-\mu)^2 = E(X^2) - \mu^2$
  - ↳ symbol is σ^2
  - ↳ std deviation is σ
  - ↳ dispersion of X

if X is constant, $Var(X) = 0$

- std dev is $\sigma = \sqrt{Var(X)}$
- $Var(aX) = a^2 Var(X)$
- $Var(aX+b) = a^2 Var(X)$

BERNOULLI AND INDICATOR VARIABLES

- Bernoulli distribution: $X \sim Ber(p)$ if X takes values $\{0,1\}$ w/ $P(X=1) = p$
  - ↳ acts as on/off switch
  - ↳ $X \sim Ber(p)$
    - X is rv that follows Bernoulli dist w/ parameter p
  - ↳ e.g. one coin toss

	X	f	
T →	0	1-p	← $P(T) = 1-p$
H →	1	p	← $P(H) = p$

  - ↳ define "success" & "failure" states

- suppose $X_1, X_2, \ldots, X_n$ is sequence of independent $Ber(p)$
  - ↳ aka Bernoulli process

Bernoulli:

↳ pmf:

X	f	F
0	$1-p$	$1-p$
1	p	$1-p+p=1$

↳ $E(x) = 0\,(1-p) + 1\cdot p$

$= p$

- $E(X)$ is equal to probability

↳ $Var(X) = E(X^2) - \mu^2$

$= 0^2\cdot(1-p)^2 + 1^2\cdot p^2 - p^2$

$= p(1-p)$

↳ $\sigma(X) = \sqrt{Var(X)}$

$= \sqrt{p(1-p)}$

- given event A, there's indicator variables corresponding to A

↳ $I_A = \begin{cases} 1 & : \text{ if } A \text{ occurs} \\ 0 & : \text{ if } A^c \text{ occurs} \end{cases}$

↳

I_A	f
0	$1-P(A)$
1	$P(A)$

↳ $I_A \sim Ber(P(A))$

- e.g.

Example

Suppose we have n envelopes, each addressed to a different person. We also have n letters written to each of those different people. Ideally, we would send the i^{th} letter to the i^{th} person, but unfortunately, we are just stuffing each random letter into a random envelope. So, a letter is in the right envelope if the name on the letter matches the name on the envelope.

If $X =$ the number of matches, then what is $E(X)$?

I_1 = indicator var for 1^{st} letter being match to 1^{st} envelope

⋮

I_n = n^{th} letter match

$$X = \overbrace{I_1 + I_2 + \ldots + I_n}^{\text{total matches, } 0 = \text{no match, } 1 = \text{match}}$$

↑
matches

$E(X) = E(I_1 + I_2 + \ldots + I_n)$

$= E(I_1) + E(I_2) + \ldots + E(I_n)$

$= \frac{1}{n} + \frac{1}{n} + \ldots + \frac{1}{n}$ ← E(X) of indicator var is success rate

$= \frac{n}{n}$

$= 1$

I_i	f
0	$1-\frac{1}{n}$
1	$\frac{1}{n}$

e.g. **Example**

The Blue Jays play the Red Sox 10 times a season. Suppose that the probability that the Jays win a game against the Sox is 0.6, and that this is independent of the outcome of any other game.

Let T be the number of wins followed by losses and the number of losses followed by wins. Find $E(T)$.

T : WL or LW

10 wins $\rightarrow T = 0$

WLW ... W $\rightarrow T = 2$

WW ... WL $\rightarrow T = 1$

$P(T=0) = P(10\text{Ws}) + P(10\text{ Ls})$
$= 0.6^{10} + 0.4^{10}$

Let $I_i = \begin{cases} 1 & \text{if } i^{th} \text{ game was T game} \\ 0 & \text{if not} \end{cases}$

$i = 2, \ldots, 10$

$E(T) = E(I_2) + E(I_3) + \ldots + E(I_{10})$

$E(I_i) = 0.4 \cdot 0.6 + 0.6 \cdot 0.4 \quad \leftarrow L \cdot W \text{ or } W \cdot L$

$= 0.48$

$E(T) = 9 \quad 0\ 48$

$= 4.32$

WEEK 4

BINOMIAL DISTRIBUTION

- binomial distribution: suppose there's n independent trials, w/ 2 possible outcomes
  - ↳ success: probability p
  - ↳ failure: probability 1-p
  - ↳ if X = # of successes in n trials, then X follows BD w/ parameters n & p
  - ↳ e.g. toss a coin 10 times

    H - success $\quad X = \#H$

    T - failure $\quad X \sim Bin(10, \frac{1}{2})$
  - ↳ must be w/ replacement
  - ↳ i.e. if $X_1, X_2, \ldots, X_n$ are independent Ber(p) & $X = X_1, X_2, \ldots, X_n$, then $X \sim Bin(n, p)$
- e.g. Example - find the p.m.f.

  Suppose that a coin is tossed 10 times with a biased coin where the probability of getting a head is 60%.

  X = # heads, find $P(X=7)$

  $X \sim Bin(10, 0.6)$

  10 times w/ 7H 3T

  $(0.6)^7(0.4)^3 \leftarrow$ 10C7 ways to arrange

  $$P(X=7) = \underset{\uparrow \ \#\text{ possible ways}}{10C7}\ \underbrace{(0.6)^7(0.4)^3}_{\text{probability of 1 particular string of 10}}$$

- suppose $X \sim Bin(n, p)$
  - ↳ support of X: $\{0, 1, 2, \ldots, n\} \rightarrow n+1$ values
  - ↳ pmf of X: $P(X=k) = f(k) = \binom{n}{k} p^k (1-p)^{n-k}$
    -

X	f
0	$nC_0 p^0 (1-p)^n$
1	$nC_1 p^1 (1-p)^{n-1}$
2	⋮
⋮	⋮
n	

- if $X \sim Bin(n, p)$, then $E(X) = np$ & $Var(X) = np(1-p)$
  - ↳ proof using Bernoulli:

    $E(X) = E(X_1) + E(X_2) + \ldots + E(X_n)$
    $= p + p + \ldots + p$
    $= np$

    $Var(X) = Var(X_1) + Var(X_2) + \ldots + Var(X_n)$
    $= p(1-p) + p(1-p) + \ldots + p(1-p)$
    $= np(1-p)$

- e.g **Example**

  Pick up 10 cards with replacement from a well-shuffled 52-card deck. If $X =$ the number of hearts, then:

  a. Find the CDF of X at 3

  b. Find $P(1 < X \leq 4)$

  c. Find $E(X)$

a) Success: getting a heart

Trial: picking a card, $n = 10$

$p = \frac{1}{4}$

$F_X(3) = P(X \leq 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$

$= \binom{10}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^{10} + \binom{10}{1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^9 + \binom{10}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^8 + \binom{10}{3}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^7$

b) $P(1 < X \leq 4) = P(X=2) + P(X=3) + P(X=4)$

c) $E(X) = np$

$= 10\left(\frac{1}{4}\right)$

$= 2.5$

· e.g.

Example

A fair coin is tossed n times. What is the probability the 1st toss is a head given we know we had r heads in n trials?

A = 1st toss is head

B = r Hs in n coin tosses

X = # heads, $X \sim Bin(n, \frac{1}{2})$

$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{2}P(X = r-1)}{P(X=r)}$ ← 1st toss is alr H so r-1 Hs remain

probability of A happening

$= \frac{\frac{1}{2}\binom{n-1}{r-1}\left(\frac{1}{2}\right)^{r-1}\left(\frac{1}{2}\right)^{n-1-(r-1)}}{\binom{n}{r}\left(\frac{1}{2}\right)^r\left(\frac{1}{2}\right)^{n-r}}$

$= \frac{\frac{1}{2}\binom{n-1}{r-1}\left(\frac{1}{2}\right)^{r-1}\left(\frac{1}{2}\right)^{n-r}}{\binom{n}{r}\left(\frac{1}{2}\right)^r\left(\frac{1}{2}\right)^{n-r}}$

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

$= \frac{r}{n}$

· e.g.

Example

You can take steps only to the left or to the right and do so randomly. If the probability of stepping to the left or the right is 0.5, then what is the probability after 20 steps you end where you started?

$P(R) = \frac{1}{2} = P(L)$

X = # steps to right, $X \sim Bin(20, \frac{1}{2})$

$P(X = 10) = \binom{20}{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right)^{10}$

GEOMETRIC DISTRIBUTION

- in sequence of indep Bernoulli trials: success → p, failure → p
  - ↳ binomial: given n & p, asked to find $P(X = k)$
    - k is # successes
  - ↳ geometric: # successes are fixed
- geometric distribution: $X \sim Geo(p)$ where X is # trials required to observe 1st success in seq of independent Bernoulli experiments

suppose $X \sim Geo(p)$

↳ support of X: $\{1, 2, 3, \dots\}$ ← infinite trials

↳ pmf of X:

X	f
1	p
2	$(1-p)p$
3	$(1-p)^2 p$
⋮	⋮

- $P(X = k) = (1-p)^{k-1} p$

- alt formulation:

↳ X = # trials to observe 1st success

- $X \sim Geo(p)$, $X = 1, 2, 3, \ldots$
  - ↳ Y = # fails before 1^{st} success
    - $Y \sim Geo(p)$, $Y = 0, 1, 2, \ldots$
- if $X \sim Geo(p)$, then $E(X) = \frac{1}{p}$ & $Var(X) = \frac{(1-p)}{p^2}$
  - ↳ proof:

X	f
1	p
2	$(1-p)p$
3	$(1-p)^2 p$
⋮	⋮

$$E(X) = p + 2p(1-p) + 3p(1-p)^2 + \ldots$$
$$-(1-p)E(X) = p(1-p) + 2p(1-p)^2 + \ldots$$
$$pE(X) = p + p(1-p) + p(1-p)^2 + \ldots$$
$$E(X) = 1 + (1-p) + (1-p)^2 + \ldots$$
$$= \frac{1}{p}$$

- if X is rv, then it's said to be memoryless if for any integers m & n, $P(X > m+n \mid X > n) = P(X > m)$
  - ↳ geometric dist is only discrete dist w/ this property
  - ↳ e.g. 50 ppl come into coffee shop & P(ppl wearing hat)
    - person A observes first 25 ppl & person B observes all 50 ppl
    - probability models for A & B are same
  - ↳ e.g. suppose fair coin is tossed, X = # trials required for 1^{st} success (i.e. H)
    $X \sim Geo(\frac{1}{2})$
    Suppose 1^{st} 6 tosses: $\underbrace{T\,T\,T\,T\,T\,T}_{\text{event } A}\ T\ T\ T\ H$

success on 4^{th} flip

$$P(X = 10 \mid A) = \frac{P(X=10 \cap A)}{P(A)} = \frac{P(X=10)}{P(A)} = \frac{(1-p)^9\, p}{(1-p)^6} = (1-p)^3 p = P(X = 4)$$

← fact that event A occurred doesn't matter b/c we know it's happened

- e.g.

> **Example**
>
> What is the probability that an average Jeopardy player will appear in 75 shows?

74 → won $\qquad P(\text{win}) = \frac{1}{3}$ ← failure

1 → lost

$P(X = 75)$, $X \sim Geo(\frac{2}{3})$

$$P(X = 75) = \left(\frac{1}{3}\right)^{74}\left(\frac{2}{3}\right)^1 \simeq e^{-36} \simeq 3.288 \cdot 10^{-36}$$

NEGATIVE BINOMIAL DISTRIBUTION

- negative binomial distribution is rep as $X \sim NB(r, p)$
  - ↳ X is # trials required to observe r^{th} success
  - ↳ each trial is part of sequence of independent Bernoulli experiments each w/ probability of success p
- suppose $X \sim NB(r, p)$
  - ↳ support of X: $\{r, r+1, r+2, \ldots\}$
  - ↳ pmf: $P(X = k) = \binom{k-1}{r-1} p^{r-1} (1-p)^{k-r}\, p = \binom{k-1}{r-1} p^r (1-p)^{k-r}$
    - e.g. suppose $r = 3$, $k = 7$

$\underbrace{\_\ \_\ \_\ \_\ \_\ \_}_{2\,S,\ 4\,F}\ \ \underset{\substack{\uparrow \\ 3^{rd} \text{ success}}}{S}$

$P(X = 7) = \binom{6}{2} p^2 (1-p)^4 p$

arrangement of 2S & 4F; 2S; 4F; 3^{rd} success

- NB vs binomial:
  - ↳ NB: we sell k vacuums & stop as soon as we've sold k vacuums
    - $k = 10 \leftarrow$ sales
      - → how many houses?
  - ↳ Bin: we sell k vacuums to n houses
- NB vs geometric:
  - ↳ X_1 = waiting time for 1^{st} success $\quad X_1 \sim Geo(p)$

    X_2 = waiting time for 2^{nd} success after 1^{st} $\longrightarrow$ $X_2 \sim Geo(p)$

    $\vdots$

    X_r = waiting time for r^{th} success after $(r-1)^{th}$ $\quad X_r \sim Geo(p)$
  - ↳ $X = X_1 + X_2 + \dots + X_r$ where each $X_i \sim Geo(p)$ & X_i s are indep
  - ↳ $X \sim NB(r,p)$ is sum of $X \sim Geo(p)$ for X_1 to X_r
- if $X \sim NB(r,p)$, then $E(X) = \frac{r}{p}$ & $Var(X) = \frac{r(1-p)}{p^2}$
  - ↳ proof: $X = X_1 + X_2 + \dots X_r \leftarrow$ iid (independent, identically distributed) Geo(p)

$$E(X) = E(X_1) + \dots E(X_r) = \frac{1}{p} + \dots + \frac{1}{p} = \frac{r}{p}$$

$$Var(X) = Var(X_1) + Var(X_2) + \dots + Var(X_r) = \frac{1-p}{p^2} + \frac{1-p}{p^2} + \dots + \frac{1-p}{p^2} = \frac{r(1-p)}{p^2}$$

- e.g.

Example

A mathematician carries two matchboxes in his pockets. One in a pocket on the left and another in a pocket on the right. Both matchboxes contain 40 matches. Being somewhat eccentric, this mathematician flips a coin to determine which matchbox he will get a match from whenever he wants to light a cigarette.

What is the probability that the box on the left has 6 matchsticks remaining if the mathematician opened the right matchbox to discover it was empty?

Suppose flip T → right, H → left

$\underset{\text{empty right box}}{40} + \underset{\text{6 remain in left box}}{34} + \underset{\text{realize right box is empty}}{1}$

$X \sim NB(41, \frac{1}{2})$

$P(X = 75) = \binom{74}{40}\left(\frac{1}{2}\right)^{40}\left(\frac{1}{2}\right)^{34} \cdot \frac{1}{2}$ ← probability of last right

- e.g.

Example

In a large city, 10% of people have an O-positive blood type. If 100 people are picked at random:

a. Find P(10 have an O-positive blood type)

b. Find P(10^{th} O-positive person is tested on the 100^{th} sample)

a) $p = 0.1$ $\quad X = \#\ O^+$ ppl

$n = 100$ $\quad X \sim Bin(100, 0.1)$

$k = 10$ $\quad P(X = 10) = \binom{100}{10}(0.1)^{10}(0.9)^{90}$

b) $r = 10$ = # successes

$p = 0.1$

y = # trials until 10^{th} 0'

$Y \sim NB(10, 0.1)$

$P(Y = 100) = \binom{99}{9}(0.1)^9(0.9)^{90} \cdot 0.1$

↳ note that a) → A , b) → B , & $B \subset A$

- i.e. NB is special case of Bin

↳ large city is necessary to keep draws indep

· e.g.

Example

The Lakers are playing the Celtics in a 7-game series. Suppose the Lakers have a 60% chance of winning each game.

a. Find P(Lakers win in exactly 6 games)

b. Find P(Celtics win the series)

a) $\underbrace{\_\ \_\ \_\ \_\ \_}_{3 \text{ Ws } \& \text{ 2Ls}}\ \underline{W}$

$P = \binom{5}{3}(0.6)^3(0.4)^2(0.6)$

b) P(Celtics win) = P(Celtics on 4) + P(Celtics on 5) + P(C on 6) + P(C on 7)

$= \binom{3}{3}(0.4)^3(0.6)^0\,0.4 + \binom{4}{3}(0.4)^4(0.6) + \binom{5}{3}(0.4)^4(0.6)^2 + \binom{6}{3}(0.4)^4(0.6)^3$

week 5

HYPERGEOMETRIC DISTRIBUTION

- $X \sim HGeo(N, r, n)$ where:
  - ↳ X is # observed successes
  - ↳ n objects are randomly selected from N total objects w/o replacement
  - ↳ only 2 types of objects: success or failure
    - r total successes
  - ↳ e.g. 52 cards w/ red being fail & black being success; draw 10 cards w/o replacement & look for 6 black cards

    $X \sim HGeo(52, 26, 10)$

    $P(X = 6) = ?$

suppose $X \sim NB(r, p)$
- ↳ support of X:
  - $k \leq r$ b/c can't draw more successes than are available
  - $k \leq n$ b/c can't draw more successes than total objects drawn
  - $x \leq \min(r, n)$
  - $x \geq n - (N - r)$

    total objects drawn (n) — # total failures ($N-r$)
  - $x \geq \max(0, n - (N-r))$
    - → e.g. 2 aces in 13-card hand

      $n - (N-r) = 13 - (52 - 4)$

      $= -35$
- ↳ pmf of X: $P(X = k) = \dfrac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}}$
  - $\binom{r}{k}$ is choose k successes from pool of r successes
  - $\binom{N-r}{n-k}$ is choose $n-k$ failures from pool of $N-r$ total failures
  - $\binom{N}{n}$ is choose n objects from N total objects
  - e.g. $P(2 \text{ aces in } 13\text{-card hand}) = \dfrac{\binom{4}{2}\binom{48}{11}}{\binom{52}{13}}$

- e.g.

Example

Suppose that we have 15 marbles in a bag, 6 red and 9 blue. We will randomly draw 8 of those marbles without replacement.

What is the probability that we will draw 3 red marbles?

X = # red marbles

$X \sim HGeo(15, 6, 8)$

$P(X = 3) = \dfrac{\binom{6}{3}\binom{9}{5}}{\binom{15}{8}}$

$= 0.391608$

With replacement, $X \sim Bin(8, \frac{6}{15})$

$P(X = 3) = \binom{8}{3}\left(\frac{2}{5}\right)^3\left(\frac{3}{5}\right)^5 = 0.278692$

· e.g.

> **Example**
>
> Suppose that we have 1500 marbles in a bag, 600 red and 900 blue. We will randomly draw 8 of those marbles without replacement.
>
> What is the probability that we will draw 3 red marbles?

X = # red marbles

$X \sim HGeo(15, 600, 8)$

$P(X=3) = \frac{600C3 \cdot 900C5}{1500C8}$ ← calculator error

Since pool of options is huge, can use binomial

$X \sim Bin(8, \frac{6}{15})$

· if N is large & n is small, then $HGeo(N, r, n) \approx Bin(n, \frac{r}{N})$

POISSON DISTRIBUTION

· $X \sim Poi(\lambda)$ where:

↳ support of X is $\{0, 1, 2, \dots\}$

↳ pmf is $P(X=x) = f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

- $x = 0, 1, 2, \dots$
- $\lambda > 0$

↳ λ: parameter for rate/mean

- avg # successes/failures per unit time/area/volume, etc.

· pmf: $f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

X	f
0	$e^{-\lambda}\lambda^0/0!$
1	$e^{-\lambda}\lambda/1!$
2	$e^{-\lambda}\lambda^2/2!$
3	$e^{-\lambda}\lambda^3/3!$
⋮	⋮

$$\text{sum} = e^{-\lambda}\left(\frac{\lambda^0}{0!} + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \dots\right)$$ ← Taylor series expansion of e^λ

$$= e^{-\lambda} \cdot e^{\lambda}$$

$$= 1$$

· $E(X) = \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda}\lambda^x}{x!}$

$$= \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^x}{(x-1)!}$$

$$= e^{-\lambda}\lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

$$= e^{-\lambda}\lambda\left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \dots\right)$$ ← Taylor series

$$= e^{-\lambda} \cdot \lambda \cdot e^{\lambda}$$

$E(X) = \lambda = \mu$

· $Var(X) = E(x^2) - \mu^2$

$$= \sum_{x=0}^{\infty} x^2 \cdot \frac{e^{-\lambda}\lambda^x}{\lambda!} - \lambda^2$$

$$= \lambda^2 + \lambda - \lambda^2$$

$Var(X) = \lambda$

· Poisson models are useful in modelling events where N is large & p is small, & we're interested in # successes/arrivals

↳ e.g. # texts you receive in 1 hr

- N = # people who text you
- p = prob a person out of these N texts you

↳ e.g. # choco chips in a cookie

↳ e.g. # earthquakes felt in S Cali in a year

} no reason these follow Poisson model but Poisson is good starting point

· if $n \to \infty$, $p \to 0$, & $np = \lambda$ then $Bin(n,p) \to Poi(\lambda)$ where $\lambda = np$

- i.e. if limits approach at apx same rate, then np should remain constant (λ)
- proof: (assume $p = \frac{\lambda}{n}$)

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$
$$= \frac{n!}{x!(n-x)!} \cdot \left(\frac{\lambda}{n}\right)^x \cdot \frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^x}$$
$$= \frac{n(n-1)\cdots(n-x)!}{x!(n-x)!} \cdot \frac{\lambda^x}{n^x} \cdot (1-\tfrac{\lambda}{n})^n (1-\tfrac{\lambda}{n})^{-x}$$
$$= \frac{n}{n} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdots \frac{(n-x+1)}{n} \cdot \frac{\lambda^x}{x!} \cdot (1-\tfrac{\lambda}{n})^n (1-\tfrac{\lambda}{n})^{-x}$$
as $n \to \infty$ $= 1 \cdot 1 \cdot 1 \cdots 1 \cdot e^{-\lambda} \cdot 1$
$$= \frac{\lambda^x e^{-\lambda}}{x!}$$

· e.g.

Example

Suppose there are 200 people at a party and the date of each person's birthday is equally likely for all 365 days. What is the probability that two people at the party were born on January 1st?

Bin:

X = # ppl born Jan 1

$X \sim Bin(200, \frac{1}{365})$

$P(X=2) = \binom{200}{2}\left(\frac{1}{365}\right)^2\left(\frac{364}{365}\right)^{198}$

$\simeq 0.086767$

Poi:

$n = 200$

$p = \frac{1}{365}$

$\lambda = \frac{200}{365}$

$X \overset{apx}{\sim} Poi(\lambda = \frac{200}{365})$

$P(X=2) = \frac{e^{-200/365}\left(\frac{200}{365}\right)^2}{2!}$

$\simeq 0.086791$

e.g.

Example

Suppose a plane has 120 seats and the airline has sold 122 tickets. If the probability of a passenger showing up for the flight is 0.97, what is the probability that there will not be enough available seats?

$E(X) = np = \lambda = 122 \cdot 0.03$

P(1 or 0 ppl don't show up) $= P(0) + P(1)$

$= \frac{e^{-0.03\cdot 122}(0.03\cdot 122)^0}{0!} + \frac{e^{-0.03\cdot122}(0.03\cdot122)^1}{1!}$

$\simeq 0.119914$

POISSON PROCESS

· Poisson process occurs if:
- ↳ # of events occurring in non-overlapping intervals are independent
- ↳ events occur singly & not in clusters
- ↳ events occur w/a constant rate (λ)

· e.g.

Example

Suppose 911 calls arrive at a rate of 3 per minute and follow a Poisson process.

a) Find the probability that exactly 2 calls arrive in 30 seconds.

b) Find the probability of getting 6 calls in 2.5 minutes.

c) Find the probability of 2 calls in the 1st minute given there were 6 calls in 1st 2.5 minutes.

a) X = # calls received in 30s

$X \sim Poi(1.5) \rightarrow \frac{3 \text{ calls}}{1 \text{ min}} \cdot 0.5 \text{ min}$

$= 1.5$ calls

$P(X=2) = \dfrac{e^{-1.5}\, 1.5^2}{2!}$

b) $Y = $ # calls in 2.5 mins

$\lambda = \dfrac{3 \text{ calls}}{1 \text{ min}} \cdot 2.5 \text{ min}$

$= 7.5$

$P(Y=6) = \dfrac{e^{-7.5}\, 7.5^6}{6!}$

c) $A = 2$ calls in 1^{st} min

$B = 6$ call in 1^{st} 2.5 min

$P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

$= \dfrac{P(2 \text{ calls in } 1^{st} \text{ min}) \cdot P(4 \text{ calls in last } 1.5 \text{ min})}{P(B)}$

$= \left(\dfrac{e^{-3}\, 3^2}{2!}\right)\left(\dfrac{e^{-4.5}\, 4.5^4}{4!}\right) \Big/ \left(\dfrac{e^{-7.5}\, 7.5^6}{6!}\right)$

- random variables X & Y are independent if $\underbrace{P(X \le x, Y \le y)}_{\text{joint PDF}} = P(X \le x) \cdot P(Y \le y), \forall x, y \in \mathbb{R}$
  - ↳ discrete case: $P(X=x, Y=y) = P(X=x) \cdot P(Y=y), \ \forall x \in X, \ \forall y \in Y$
  - ↳ 3+ : $P(X \le x, Y \le y, W \le w) = P(X \le x) \cdot P(Y \le y) \cdot P(W \le w), \ \forall x, y, w \in \mathbb{R}$
- e.g.

Example

Suppose two dice are rolled and we define rv's:

$X = \#$ on the first roll

$Y = \#$ on the second roll

Are $X + Y$ and $X - Y$ independent?

$P(X+Y=12, X-Y=1) = 0$ impossible

↓
$X = 6$
$Y = 6$

$P(X+Y=12) = \dfrac{1}{36}$

$P(X-Y=1) = \dfrac{5}{36}$

$P(X+Y=12, X-Y=1) \neq P(X+Y=12) \cdot P(X-Y=1)$

$X+Y$ & $X-Y$ are not independent

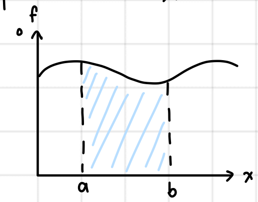
- if X & Y are independent
  - ↳ $g(X)$ & $h(Y)$ are independent
  - ↳ $E(XY) = E(X) \cdot E(Y)$
- iid rvs are independent & identically distributed
-

Independent?	Identically distributed?	Example
Yes	Yes	2 coins are flipped $X = \begin{cases} 1 & \text{H on } 1^{st} \text{ flip} \\ 0 & \text{otherwise} \end{cases}$ $Y = \begin{cases} 1 & \text{H on } 2^{nd} \text{ flip} \\ 0 & \text{otherwise} \end{cases}$ $\Big\rangle$ Bin$(2, \frac{1}{2})$
Yes	No	X = flip of coin (H) Y = take card from deck (ace)
No	Yes $X \sim \text{Bin}(2, \frac{1}{2})$ $Y \sim \text{Bin}(2, \frac{1}{2})$	Fair coin is tossed twice $X = \#$ H $Y = 2 - X$ (# T)
No	No	Fair coin is tossed 3 times $X = \#H$ $Y = \|\#H - \#T\|$ $= \|X - \#T\|$

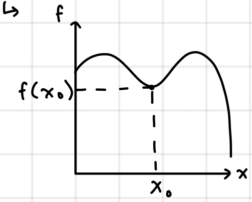
WEEK 7

CONTINUOUS DISTRIBUTIONS

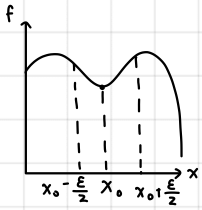
- X is continuous rv if its support is subset of $\mathbb{R}$
  - ↳ cdf of X : $F(x) = P(X = x)$
    - $P(X = a) = 0$ if X is cont rv
      - → i.e. pmfs don't exist for cont rv
  - ↳ pdf of X : X is rv w/ density function f if $P(a \leq X \leq b) = \int_a^b f(x)\,dx$



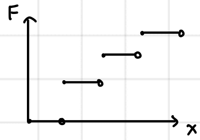
    - f can be > 1 but area under ≤ 1
- $f(x_0)$ is single point
  - ↳



  - ↳ take $\varepsilon > 0$, $x \pm \frac{\varepsilon}{2}$



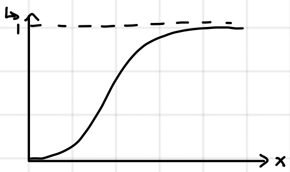
    - $P(x_0 - \frac{\varepsilon}{2} < x < x + \frac{\varepsilon}{2}) \simeq \varepsilon f(x_0)$
  - ↳ e.g. if $f(x_0) = 0.5$ & $f(y_0) = 0.25$, then $f(x_0)$ is roughly twice as likely to occur as $f(y_0)$
- discrete CDF . F is non-dec & right-cont
  - ↳ $\lim_{x \to -\infty} F(x) = 0$
  - ↳ $\lim_{x \to \infty} F(x) = 1$
  - ↳



discrete pmf : $f(x) \geq 0$ & $\sum f(x) = 1$

continuous CDF: properties are same as discrete

↳



- continuous pdf : $f(x) \geq 0$ & $\int_{-\infty}^{\infty} f(x) = 1$

Discrete	Continuous
$E(X) = \sum x f(x)$	$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$
$Var(X) = E(X^2) - (E(X))^2$	$Var(X) = \int_{-\infty}^{\infty} x^2 f(x)\,dx - [\int_{-\infty}^{\infty} x f(x)\,dx]^2$
pmf to cdf : $F(x) = \sum f(x)$	pdf to cdf: $F(x) = \int_{-\infty}^{x} f(y)\,dy$
cdf to pmf: $f(x) = F(x) - F(x-1)$	cdf to pdf: $\frac{dF}{dx} = f(x)$

· e.g.

Example

Let X be a r.v. with the Following pdf:

$$f(x) = \begin{cases} 3x^2 & : 0 \le x \le 1 \\ 0 & : otherwise \end{cases}$$

Is it legit pdf? Yes

↳ $f(x) \ge 0$ ✓

↳ add to 1? ✓

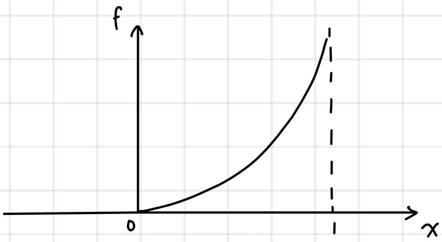
$\int_{-\infty}^{\infty} f(x)\,dx$

$= 0 + \int_0^1 3x^2 dx + 0$

$= [x^3]_0^1$

$= 1 - 0$

$= 1$



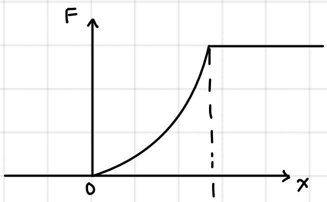
Find CDF:

$F(x) \rightarrow P(X \le x) = \int_0^x 3y^2\,dy$

$= [y^3]_0^x$

$= x^3$

$$F(x) = \begin{cases} 0 & , x < 0 \\ x^3 & , 0 \le x \le 1 \\ 0 & , x > 1 \end{cases}$$



$P(\frac{1}{3} < x < \frac{3}{4}) = P(\frac{1}{3} \le x \le \frac{3}{4}) = \int_{\frac{1}{3}}^{\frac{3}{4}} 3x^2\,dx$

$= [x^3]_{1/3}^{3/4}$

$= \frac{27}{64} - \frac{1}{27}$

$= \frac{665}{1728}$

$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^1 x\,3x^2\,dx$

$= \int_0^1 3x^3\,dx$

$= [\frac{3}{4}x^4]_0^1$

$= \frac{3}{4}$

$P(x > \frac{1}{2}) = \int_{\frac{1}{2}}^1 3x^2\,dx$

$= [x^3]_{\frac{1}{2}}^1$

$= 1 - \frac{1}{8}$

$= \frac{7}{8}$

$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx$

$= \int_0^1 3x^4\,dx$

$= [\frac{3}{5}x^5]_0^1$

$= \frac{3}{5}$

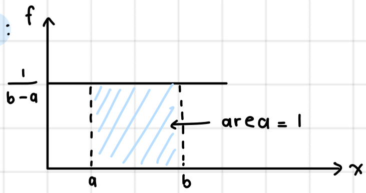
UNIFORM CONTINUOUS DISTRIBUTION

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

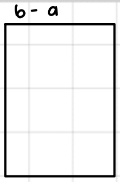
↳ i.e. U, Uni

↳ most often used $X \sim U(0,1)$

· pdf:



↳ $b-a$



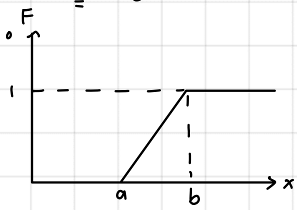
$\frac{1}{b-a}$

- $A = \frac{b-a}{b-a} = 1$

· CDF:

$$F(x) = \begin{cases} 0 & , \quad x < a \\ \frac{x-a}{b-a} & , \quad a \le x \le b \\ 1 & , \quad x > b \end{cases}$$

$$↳ F(x) = \int_0^x \frac{1}{b-a} \, dy = \left[\frac{y}{b-a} \right]_a^x = \frac{x-a}{b-a}$$



$$\cdot \; E(X) = \int_{-\infty}^{\infty} x f(x) \, dx = \int_a^b x \frac{1}{b-a} \, dx = \frac{1}{b-a} \int_a^b x \, dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{1}{b-a} \left(\frac{b^2 - a^2}{2} \right)$$

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

$$\cdot \; Var(X) = E(X^2) - (E(X))^2 = E(X^2) - \left(\frac{a+b}{2} \right)^2$$

$$Var(X) = \frac{1}{12} (b-a)^2$$

· e.g.

Example

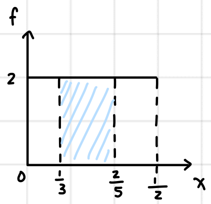
Let $X \sim \text{Uniform}[a, b]$. Find $P(\frac{1}{3} \le x \le \frac{2}{5})$.

↓

Uni $[0, \frac{1}{2}]$

$$f(x) = \frac{1}{b-a} = \frac{1}{\frac{1}{2} - 0} = 2$$

$$P\left(\frac{1}{3} \le x \le \frac{2}{5}\right) = \int_{1/3}^{2/5} 2 \, dx = 2 [x]_{1/3}^{2/5} = 2 \left(\frac{2}{5} - \frac{1}{3} \right) = \frac{2}{15}$$



$$P\left(\frac{1}{3} \le x \le \frac{2}{5}\right) = F\left(\frac{2}{5}\right) - F\left(\frac{1}{3}\right)$$

UNIVERSALITY OF UNIFORM DISTRIBUTION

· suppose X is rv w/cdf $F(X)$; want to generate n simulations of X (i.e. generate n outcomes $\{x_1, x_2, \dots, x_n\}$)

↳ solution: $F^{-1}(u_1), \dots, F^{-1}(u_n)$

- must be easy to generate $u_1, \dots, u_n$
- F must be invertible

· e.g.

Example

Suppose X has a pdf:

$$f(x) = \begin{cases} e^{-x} & : x \ge 0 \\ 0 & : \textit{otherwise} \end{cases}$$

Generate n outcomes from this r.v.

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

$= \int_0^x e^{-y}\,dy$

$= [-e^{-y}]_0^x$

$= 1 - e^{-x}$

$$F(x) = \begin{cases} 1 - e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$u_1, \dots, u_n \sim U(0,1)$ is generated

$F(x_i) = u_i \Rightarrow x_1 = F^{-1}(u_i)$

$1 - e^{-x_i} = u_i$

$-x_i = \ln(1 - u_i)$

$x_i = -\ln(1 - u_i)$

↓ computer generates simulations

$-\ln(1-u_1), -\ln(1-u_2), \dots, -\ln(1-u_n)$

- theorem: let X be cont rv w/ invertible cdf F, then $F(X) \sim U \sim U(0,1)$
  - ↳ i.e. $X = F^{-1}(U)$
  - ↳ proof: suppose $F(X)$ is cdf of rv X; let $Y = F(X)$

$P(Y \leq y)$ ← $F(Y) = F(F(X))$ since $Y = F(X)$

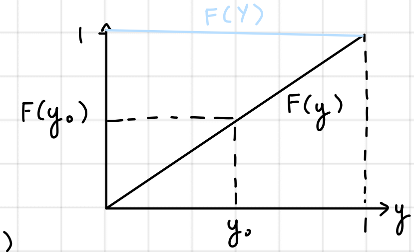
$= P(F(X) \leq y)$

$= P(X \leq F^{-1}(y))$

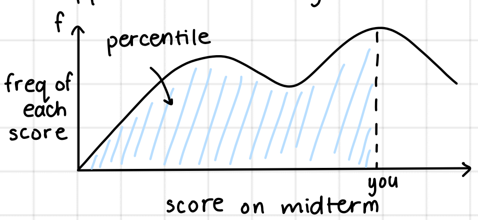
$= F(F^{-1}(y))$

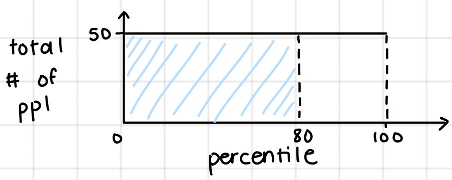
$= y$

So, $P(Y \leq y) = y$ & Y must be $U(0,1)$ since $F(Y) = y$ for Uniform(0,1)



- an illustration:
  1) suppose we look at grades on midterm (we'll assume they're cont)



  2) to understand how you did, comp your score against the class
     - look at your percentile (proportion of class that you did better than)
     - $P(X \leq x) = F(X)$ → cdf
  3) percentiles are uniform in their spread
     - i.e. 70-79 vs 80-89 (both contain top 10% of ppl)
     - 

  4) take random test score, then convert to percentile
     - gives uniformly random percentile btwn 0 & 100
  - ↳ source of distribution doesn't matter
     - $F(X) \sim U$ (CDF of rv X is uniform)
     - $F^{-1}(U) \sim X$ (inverse of CDF of uniform gives us rv of that distribution)

EXPONENTIAL DISTRIBUTION

- X is Exponential rv w/ λ: $X \sim Exp(\lambda)$ if $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0 & x < 0 \end{cases}$

↳ λ is rate parameter

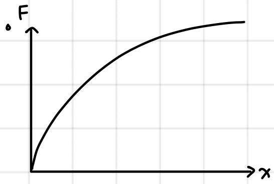
↳ $\int_0^\infty \lambda e^{-\lambda x}\,dx = 1$

↳ for $\lambda = 1$,

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

$$= \int_0^x \lambda e^{-\lambda y}\,dy$$
$$= \left[\frac{\lambda e^{-\lambda y}}{-\lambda}\right]_0^x$$
$$= 1 - e^{-\lambda x}$$

↳ $F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$



↳ alt notation using θ & μ: $F(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}}$

- $\lambda = \frac{1}{\mu} = \frac{1}{\theta}$

- $E(X) = \int_0^\infty x f(x)\,dx$

$$= \int_0^\infty x \lambda e^{-\lambda x}\,dx$$

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

↳ this is why $\lambda = \frac{1}{\mu}$

- $Var(X) = E(X^2) - (E(X))^2$

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

- e.g.

Example

Suppose $X \sim \text{Exp}(\lambda = \frac{1}{2})$. Find the median of X.

Let m rep median.

$P(X \le m) = 0.5 = F(m)$ ← CDF of m

$$0.5 = 1 - e^{-\frac{1}{2}m}$$
$$e^{-\frac{1}{2}m} = 0.5$$
$$m = -2\ln\left(\tfrac{1}{2}\right)$$

- exponential distribution is memoryless

↳ only cont dist w/memorylessness

↳ if $X \sim Exp(\lambda)$, then $\forall s, t \ge 0$ $P(X > s+t \mid X > s) = P(X > t)$

- proof: $P(X > s+t \mid X > s) = \frac{P(X > s+t \cap X > s)}{P(X > s)}$

$$= \frac{P(X > s+t)}{P(X > s)}$$
$$= \frac{1 - P(X \le s+t)}{1 - P(X \le s)}$$
$$= \frac{1 - F(s+t)}{1 - F(s)}$$
$$= \frac{1 - (1 - e^{-\lambda(s+t)})}{1 - (1 - e^{-\lambda s})}$$
$$= \frac{e^{-\lambda s} e^{-\lambda t}}{e^{-\lambda s}}$$
$$= e^{-\lambda t}$$
$$= P(X > t)$$

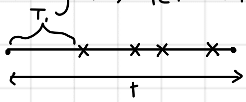
- can use waiting time interpretation when thinking abt memorylessness

↳ Bernoulli process: seq of 0s & 1s

- waiting time for 1st success ~ Geo

↳ Poisson process: cont version of Bernoulli

- waiting time for 1^{st} success ~ Exp

- let N_t be # arrivals in time interval of length t of Poisson process w/ intensity λ; let T_1 be waiting time for 1^{st} arrival; then, $T_1 \sim Exp(\lambda)$
  - ↳ 
    λ: rate
  - ↳ proof:
    Let CDF of T_1 be $F_{T_1}(t_1) = P(T_1 \le t_1)$
    Look at $P(T_1 > t_1) = P(0 \text{ successes in } [0, t_1]) \rightarrow Poi(\lambda t_1)$

    prob you had to wait more than t_1 for 1^{st} success

    $$= \frac{e^{-\lambda t_1}(\lambda t_1)^0}{0!}$$
    $$= e^{-\lambda t_1}$$

    $$P(T_1 \le t_1) = 1 - P(T_1 > t_1)$$
    $$= 1 - e^{-\lambda t}$$

    So, $T_1 \sim Exp(\lambda)$

- note that for discrete, NB was seq of Geo dist & for continuous, Gamma dist is analagous to NB

- e.g.

Example (City A)

City A has buses that appear every 10 minutes, but you arrive at bus stops at random times (uniform).

a. What is the expected waiting time?
b. Find $P(X > 6 + 2 | X > 6)$.

a) X = time spent waiting (mins)
$$E(X) = \frac{0+10}{2}$$
$$= 5 \text{ mins}$$

b) $$P(X > 8 | X > 6) = \frac{P(X > 8, X > 6)}{P(X > 6)}$$
$$= \frac{P(X > 8)}{P(X > 6)}$$
$$= \frac{2/10}{4/10}$$
$$= \frac{1}{2}$$

Not memoryless: $P(X > 2) = \frac{8}{10}$

- e.g

Example (City B)

City B has buses that appear $\sim Exp(\lambda = \frac{1}{10}$ every 10 minutes, but you arrive at bus stops at random times (uniform).

a. What is the expected waiting time?
b. Find $P(X > 6 + 2 | X > 6)$.

a) $\mu = \frac{1}{\lambda} = 10$
$E(X) = 10$

b) $$P(X > 8 | X > 6) = P(X > 2)$$
$$= 1 - F(2)$$
$$= 1 - (1 - e^{-\frac{1}{10}(2)})$$
$$= e^{-\frac{1}{5}}$$

WEEK 8

MOMENT GENERATING FUNCTIONS

- moment generating function (mgf) of rv X is $M_X(t) = E(e^{tX})$
  - ↳ e^{tX} is function of t (dummy var)
  - ↳ $E(e^{tX})$ needs to be finite on open interval containing 0 or mgf DNE
  - ↳ $M_X(0) = E(e^{0X})$
    $= E(1)$
    $= 1$
  - ↳ $M_X(t) = E(e^{tX})$
    $= \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$
    - in discrete case, $M_X(t) = \sum_{x \in X} e^{tx} \cdot f(x)$
- e.g

Example (Bernoulli mgf)

Find the mgf of a Bernoulli distribution.

X	f
0	1-p
1	p

$M_X(t) = E(e^{tX})$
$= \sum e^{tx} \cdot f(x)$
$= e^{t(0)} \cdot (1-p) + e^{t(1)}(p)$
$= 1 - p + pe^t$

- e.g.

Example (Uniform mgf)

Find the mgf of a Uniform distribution.

$M_U(t) = E(e^{tU})$
$= \int_a^b e^{tu} \cdot \frac{1}{b-a}\, du$
$= \frac{e^{tb} - e^{ta}}{t(b-a)}$

↳ although not defined at $t=0$, still exists on open interval around 0

- can calc n^{th} moment of rv X by evaluating n^{th} derivative of mgf at $t=0$
  - ↳ $M_X(t) = E(e^{tX})$
    $= E(1 + \frac{tx}{1!} + \frac{t^2x^2}{2!} + \dots)$
    $= E(1) + \frac{t}{1!}E(x) + \frac{t^2}{2!}E(x^2) + \dots$

    $M'_X(t) = 0 + E(x) + \frac{2t}{2!}E(x^2) + \frac{3t^2}{3!}E(x^3) + \dots$

    At $t=0$, $M'_X(t) = E(x)$

    $M''_X(t) = E(x^2) + \frac{6t}{3!}E(x^3) + \dots \longrightarrow M_X''(0) = E(x^2)$
- mgfs allow us to replace messy integration w/cleaner derivatives
- if 2 RVs have same mgf, then they have same distribution
- if RVs X & Y are independent, then $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$
  - ↳ useful b/c often interested in sums & avgs in modelling
- e.g.

Example

Given $X \sim Bin(n,p)$, find the mgf of X.

$M_X(t) = E(e^{tX})$
$= e^{t \cdot 0}\binom{n}{0}p^0(1-p)^x + e^{t \cdot 1}\binom{n}{1}p^1(1-p)^{n-1} + \dots$

Since $X = X_1 + X_2 + \dots + X_n$ (indep Bernoulli processes),

$M_X(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdots M_{X_n}(t)$
$= (pe^t + 1 - p)^n$

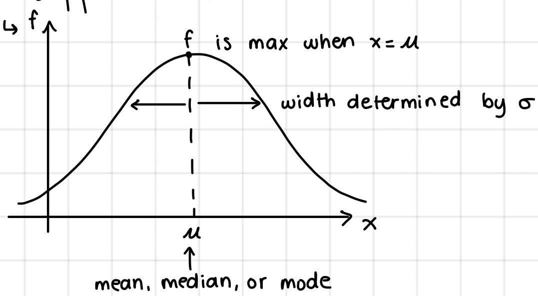
$M'_X(t) = n(pe^t + 1 - p)^{n-1} \cdot pe^t$

↳ $M'_X(0) = np \leftarrow E(X)$

- moments rep:
  - 1^{st} : $E(X) = \mu$
  - 2^{nd} : $E(X^2)$
    - w/ 1^{st} moment $\Rightarrow \sigma^2$ ($Var(X)$)
  - 3^{rd} : $E(X^3)$
    - skewness
  - 4^{th} : $E(X^4)$
    - kurtosis (i.e. how "fat" tail is)

NORMAL DISTRIBUTION

- rv X is normally distributed if its pdf $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
  - denoted as $X \sim N(\mu, \sigma^2)$
  - $f(x)$ is symmetrical around μ
    - i.e. $f(x+\mu) = f(\mu - x)$
  - support of X is $(-\infty, \infty)$
  -

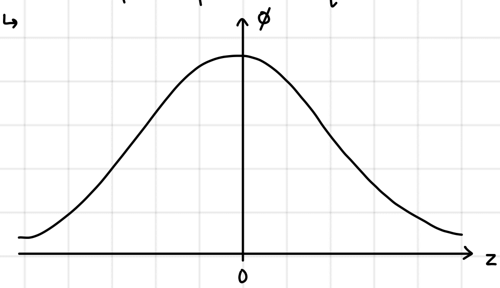


- cdf is $F(x) = P(X \leq x)$

  $F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy$
  - no closed-form soln
- expectation & variance:
  - $E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$

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

    $E(X) = \mu$
  - $Var(X) = \sigma^2$
- standard normal is $Z \sim N(0, 1)$, $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$
  - $\mu = 0$
  - $\sigma^2 = 1$
  - ϕ (phi) is equiv to f
  - Φ (capital phi) is equiv to F
  -

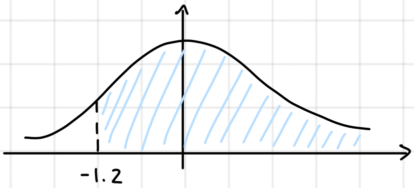


    - ϕ is symmetric around 0
      - → i.e. $\phi(z) = \phi(-z)$
  - to find probabilities, use z-table
- e.g. calculate:

  a) $P(Z \geq 1) = 1 - 0.84134$

  $= 0.15866$

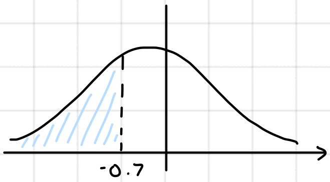
b) $P(Z \geq -1.2) = P(Z \leq 1.2) = 0.88493$



c) $P(Z \leq -0.7) = 1 - P(Z \leq 0.7)$

$= 1 - 0.75804$

$= 0.24196$



d) $P(1 \leq Z \leq 2) = P(Z \leq 2) - P(Z \leq 1)$

$= 0.97725 - 0.84134$

$= 0.13591$

· standardization of normal distribution theorem: if $X \sim N(\mu, \sigma^2)$ & $Z = \frac{X-\mu}{\sigma}$, then $Z \sim N(0,1)$

↳ allows us to convert any normal distn into standard normal distn

↳ proof: suppose $X \sim N(\mu, \sigma^2)$

Let $Y = \frac{X-\mu}{\sigma}$

$F_Y(y) = P(Y \leq y)$

$= P(\frac{X-\mu}{\sigma} \leq y)$

$= P(X \leq \mu + y\sigma)$

$= P(X \leq x)$

$= F_X(x)$

· e.g.

Example

Suppose that final grades in this course follow a normal distribution with $\mu = 75$ and $\sigma^2 = 64$.

a) Find the probability that a student has a final grade of 83 or more.

b) Calculate the 95$^{\text{th}}$ percentile.

a) $X \sim N(75, 64)$

$P(X \geq 83) = P(\frac{X-\mu}{\sigma} \geq \frac{83-\mu}{\sigma})$

$= P(Z \geq \frac{83-75}{8})$

$= P(Z \geq 1)$

$= 1 - 0.84134$

$= 0.15866$

b) $P(Z \leq z) = 0.95$

$0.95 = P(Z \leq 1.6449)$

$Z = \frac{X-\mu}{\sigma} \leq 1.6449$

$\frac{X-75}{8} \leq 1.6449$

$X \leq 88.1592$

· to interpret z: $z = \frac{x-\mu}{\sigma}$

↳ z is z-score

↳ RHS is # std dev above/below that x is away from mean(μ)

- 68-95-99 rule: if $X \sim N(\mu, \sigma^2)$, then
  - ↳ ~68% of observations are btwn $\mu \pm \sigma$
  - ↳ ~95% of observations are btwn $\mu \pm 2\sigma$
  - ↳ ~99% of observations are btwn $\mu \pm 3\sigma$
- if $X \sim N(\mu, \sigma^2)$, then $M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$
  - ↳ mgf of $Z \sim N(0,1)$ is $M_Z(t) = e^{\frac{t^2}{2}}$
    - proof: $M_Z(t) = E(e^{tz})$

$$= \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$
$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2 + tz} dz$$

completing the square

$$= e^{\frac{1}{2}t^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2} dz$$

$N(t,1)$

$$= e^{\frac{1}{2}t^2}(1)$$
$$= e^{\frac{1}{2}t^2}$$

  - ↳ 1st moment: $M'_X(t) = e^{(\mu t + \frac{1}{2}\sigma^2 t^2)} \cdot (\mu + \sigma^2 t)$
    $M'_X(0) = \mu$
- if X has mgf $M_X(t)$ & $Y = aX + b$, then $M_Y(t) = e^{at} M_X(bt)$
- sampling distn of sample mean theorem: let $X_1, X_2, \dots, X_n$ be iid $N(\mu, \sigma^2)$, then:
  - ↳ $S_n = \sum X_i \sim N(n\mu, n\sigma^2)$
    - i.e. sum of normals is normal w/ $n\mu$ & $n\sigma^2$
  - ↳ $\bar{X}_n = \frac{1}{n} \sum X_i \sim N(\mu, \frac{\sigma^2}{n})$
    - i.e. avg of normals is normal w/ μ, $\frac{\sigma^2}{n}$
- e.g.

Example

Suppose that final grades in this course follow a normal distribution with $\mu = 75$ and $\sigma^2 = 64$. A sample of 9 students is taken

a) $P(S \geq 700)$

b) $P(\bar{X} < 73)$

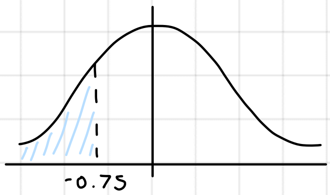
a) $S \sim N(75 \cdot 9, 64 \cdot 9)$

$= S \sim N(675, 576)$

$$P(S \geq 700) = P\left(\frac{S-\mu}{\sigma} \geq \frac{700-\mu}{\sigma}\right)$$
$$= P\left(Z \geq \frac{700-675}{24}\right)$$
$$= P(Z \geq 1.041667)$$
$$= 1 - 0.85083$$
$$= 0.14917$$

b) $\bar{X} \sim N(75, \frac{64}{9})$

$$P(\bar{X} < 73) = P\left(\frac{\bar{X}-\mu}{\sigma} < \frac{73-\mu}{\sigma}\right)$$
$$= P\left(Z < \frac{73-75}{8/3}\right)$$
$$= P(Z < -0.75)$$
$$= 1 - P(Z < 0.75)$$
$$= 1 - 0.77337$$
$$= 0.22663$$



CENTRAL LIMIT THEOREM

- law of large #s: consider sequence of n iid rv $X_1, X_2, \dots, X_n$ where $E(X_i) = \mu$ & $Var(X_i) = \sigma^2 \; \forall i$
  - ↳ let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$, then $\bar{X}_n \to \mu$ as $n \to \infty$
  - ↳ i.e. as a sample size grows, its means gets closer to μ of whole population
  - ↳ holds for any distn w/ finite mean & var
  - ↳ if we don't know μ, take a sample & as n gets bigger, $\bar{X}_n \to \mu$

- Central Limit Theorem (CLT): as $n \to \infty$, $\bar{X}_n \underset{\substack{\uparrow \\ \text{approaches}}}{\sim} N(\mu, \frac{\sigma^2}{n})$ or $Z = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$
  - ↳ e.g. if $X \sim Bin(n, p)$, then $X \sim N(np, np(1-p))$
-

Normal distn	Other distns
$\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$	$\bar{X}_n \underset{app}{\sim} N(\mu, \frac{\sigma^2}{n})$
$S_n \sim N(n\mu, n\sigma^2)$	$S_n \underset{app}{\sim} N(n\mu, n\sigma^2)$

$\underbrace{\hspace{4em}}_{\text{as } n \to \infty}$ (applies to the Other distns column)

e.g

Example

Suppose $X \sim Bin(10000, 0.4)$, then calculate $P(3800 < X < 4200)$.

$X \sim N(4000, 2400)$ by CLT

$$P(3800 < X < 4200) = P\left(\frac{3800 - 4000}{\sqrt{2400}} < \frac{X - \mu}{\sigma} < \frac{4200 - 4000}{\sqrt{2400}}\right)$$
$$= P(-4.082483 < Z < 4.082483)$$
$$= P(Z < 4.082483) - P(Z < -4.082483)$$
$$= \ldots$$

- have to use continuity correction when using normal distn (continuous) to apx binomial distn (discrete)
  - ↳ if $P(X = n)$, use $P(n - 0.5 < X < n + 0.5)$
- if $X_1, X_2, \ldots, X_n \sim Poi(\lambda)$, then $\bar{X}_n \sim N(\lambda, \frac{\lambda}{n})$
  - ↳ $\mu = \lambda$
  - ↳ $\frac{\sigma^2}{n} = \frac{\lambda}{n}$

week 9

DISCRETE JOINT DISTRIBUTIONS

- discrete joint distribution function: joint pmf of discrete rvs X & Y is $f(x,y) = P(X=x, Y=y)$
  - ↳ $x \in$ support of X
  - ↳ $y \in$ support of Y
  - ↳ properties:
    - $f(x,y) \geq 0$
    - $\sum_{\text{all } x} \sum_{\text{all } y} f(x,y) = 1$
- e.g.

X \ Y	0	1
0	0.3	0.1
1	0.2	0.4

} adds to 1

  - ↳ $P(X=1, Y=1) = 0.4$
  - ↳ $f(0,0) = 0.3$
- marginal distn is distn of values for one rv that ignores other rvs in dataset
  - ↳ given $f(x)$ & $f(y)$, we can find $f(x,y)$ when X & Y are indep
  - ↳ e.g.

X \ Y	0	1
0	0.3	0.1
1	0.2	0.4

→

X	f_x
0	0.3+0.1=0.4
1	0.6

Y	f_y
0	0.5
1	0.5

  - ↳ to find marginal distn, fix specified var at each value & add up column/row
- e.g.

Example

Given that f_X, f_Y and X, Y are independent with:

X	f_X
0	0.4
1	0.6

Y	f_Y
0	0.5
1	0.5

Find the joint distribution.

X \ Y	0	1
0	0.4·0.5 = 0.2	0.2
1	0.3	0.3

- covariance of X & Y is $Cov(X,Y) = E((X-E(X))(Y-E(Y)))$
  - ↳ another form is $Cov(X,Y) = E(XY) - E(X) \cdot E(Y)$
    - $E(XY)$ is avg of product
    - $E(X) \cdot E(Y)$ is product of avgs
  - ↳ unit dependent & assumes unit of product
  - ↳ +ve means rvs change in same dir & -ve means they change inversely
  - ↳ properties:
    - $Cov(X,k) = 0$
      - → k is constant
    - $Cov(X,X) = Var(X)$
    - $Cov(X,Y) = Cov(Y,X)$
    - $Cov(X, a+bX) = bVar(X)$
    - if X & Y are indep $\Rightarrow Cov(X,Y) = 0$

→ implication does not hold other way around

- $Cov(X, Y+Z) = Cov(X,Y) + Cov(X,Z)$

- correlation coeff : $\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}$
  - ↳ unit indep (i.e. dimensionless)
  - ↳ indicates linear rltnship btwn 2 rvs
  - ↳ properties:
    - $|\rho_{X,Y}| \le 1$
      - → $|\rho| \simeq 1$ indicates strong linear rltnship
      - → $|\rho| \simeq 1$ indicates no linear rltnship
    - if $Y = a + bX$, then $\rho_{X,Y} = 1$ or -1
    - unit indep
    - if X & Y are indep $\Rightarrow \rho_{X,Y} = 0$
      - → $\rho_{X,Y} = 0 \not\Rightarrow$ X & Y are indep b/c there may be a non-linear rltnship
    - if $\rho_{X,Y} = 0$, we say X & Y are not correlated
    - correlation $\neq$ causation

$Var(aX+bY)$
$= a^2 Var X + b^2 Var Y + 2ab\,Cov(X,Y)$

- e.g.

Example

Suppose $f(x,y) = kq^{x+y-2}p^2$, $q = 1-p$, $x = 1,2,\ldots$, $y = 1,2,\ldots$

a) Calculcate k

b) Find $f_X(x)$

a)
$$1 = \sum_{x=1}^{\infty}\sum_{y=1}^{\infty} kq^{x+y-2}p^2$$
$$1 = \sum_{x=1}^{\infty}\sum_{y=1}^{\infty} kq^x q^y \frac{p^2}{q^2}$$
$$1 = \frac{kp^2}{q^2}\sum_{x=1}^{\infty} q^x \sum_{y=1}^{\infty} q^y$$
$$1 = \frac{kp^2}{q^2}\left(\frac{q}{1-q}\right)\left(\frac{q}{1-q}\right)$$
$$1 = \frac{k(1-q)^2}{q^2}\left(\frac{q}{1-q}\right)^2$$
$$1 = k$$

b) To find $f_X(x)$, keep x fixed & add all y
$$f(x,y) = q^{x+y-2}p^2 \quad \rightarrow k = 1$$
$$f_X(x) = \sum_{y=1}^{\infty} q^{x+y-2}p^2$$
$$= \frac{q^x p^2}{q^2}\sum_{y=1}^{\infty} q^y$$
$$= \frac{q^x p^2}{q^2}\left(\frac{q}{1-q}\right)$$
$$= \frac{q^{x+1}p^2}{q^2 p}$$
$$= q^{x-1}(p)$$
$$= (1-p)^{x-1}p$$

MULTINOMIAL DISTRIBUTION

- vector $(X_1, X_2, \ldots, X_k) \sim Multi(n, p_1, p_2, \ldots, p_k)$ follows multinomial distn if
  - ↳ there's n trials where each trial has k possible outcomes w/ probabilities p_i
    - $i = 1, 2, \ldots, k$ & $\sum_{i=1}^{k} p_k = 1$
  - ↳ trials are indep
  - ↳ X_i is # of successes of type i in n total trials so $\sum_{i=1}^{k} X_i = n$

- e.g.

Example

You are officiating a race between your friends One, Two and Three. Suppose One, Two, Three the probabilities of winning a race are $p_1 = 0.2$, $p_2 = 0.3$, and $p_3 = 0.5$ respectively. Your friends will run 10 total races.

$(X_1, X_2, X_3) \sim \text{Multi}(10, 0.2, 0.3, 0.5)$

$P(X_1 = 2, X_2 = 4, X_3 = 4) = f(2, 4, 4)$

$= \frac{10!}{2!4!4!} (0.2)^2 (0.3)^4 (0.5)^4$

↳ suppose $X_1 = 2$; what's conditional distn of $X_2 | X_1 = 2$?

Let $Y = X_2 | X_1 = 2$

$Y = \{0, 1, 2, ..., 8\}$

↓

only Two/Three can win these 8 games

$$P(\text{Two win} | \text{Two/Three win}) = \frac{P(\text{Two win})}{P(\text{Two/Three win})} = \frac{0.3}{0.3 + 0.5} = \frac{3}{8}$$

$Y \sim \text{Bin}(8, \frac{3}{8})$

- joint pmf is $f(x_1, x_2, ..., x_k) = \frac{n!}{x_1! x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$
- marginals are $f_{X_1}(x_1) = \binom{n}{x_1} p_1^{x_1} (p_2 + ... + p_k)^{x_2 + ... + x_k}$

INTRO TO DATA ANALYSIS AND STATISTICAL INFERENCE

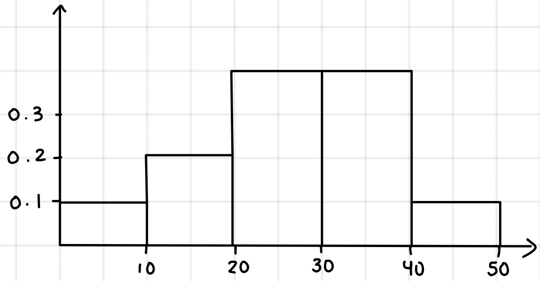
- 2 main types of data:
  - ↳ categorical
    - one subtype is ordinal where there's implied order
  - ↳ numerical
    - subtypes are discrete & cont
- when collecting data, goal is to ensure it's representative of entire pop
  - ↳ biased data comes from systematic errors in data collection
- measures of central tendency:
  - ↳ sample mean is $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$
  - ↳ sample median is middle value
    - not affected by extreme values
  - ↳ quartiles:
    - Q_1 : 25^{th} percentile
    - Q_2 : median / 50^{th} percentile
    - Q_3 : 75^{th} percentile
- 5-number summary is $(\min, Q_1, Q_2, Q_3, \max)$
- measures of dispersion & symmetry:
  - ↳ range = max − min
  - ↳ $IQR = Q_3 - Q_1$
    - inter-quartile range
  - ↳ variance
    - $s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$
    - std dev is s
  - ↳ skewness measures bias twd left/right side
  - ↳ kurtosis is measure of normality
- diff symbol usage:

Population	Sample	Meaning
μ	$\bar{x}$	mean
σ^2	s^2	variance
σ	s	std dev
$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	$s_{\bar{x}} = \frac{s}{\sqrt{n}}$	std error of mean

· in density histogram, area of each bar is proportional to relative freq

↳ e.g.

	freq	rel freq
0-10	10	0.1
10-20	20	0.2
20-30	30	0.3
30-40	30	0.3
40-50	10	0.1

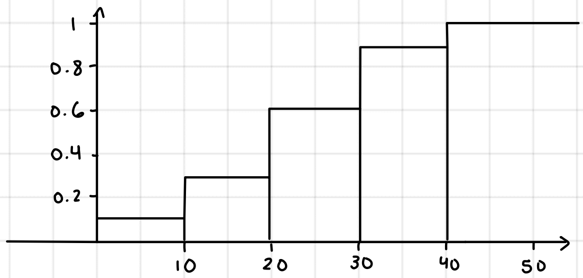


· empirical CDF is used to estimate CDF

↳ for data $\{y_1, \dots, y_n\}$, $\hat{F}(y) = \#$ observations $\leq y$

↳ e.g.

	freq	rel freq	cumulative rel freq
< 0	0	0	0
0-10	10	0.1	0.1
10-20	20	0.2	0.3
20-30	30	0.3	0.6
30-40	30	0.3	0.9
40-50	10	0.1	1

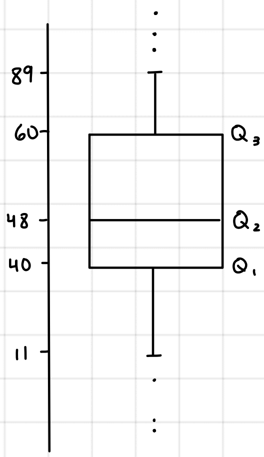


· box plots use 5# summary

↳ e.g. $\underset{\uparrow \text{ data}}{\{y_1, \dots, y_n\}} = \{\underbrace{1, 3, 7}_{\text{outliers}}, 11, \dots, 85, 89, \underbrace{110, 120, 145}_{\text{outliers}}\}$

Suppose $Q_1 = 40$, $Q_2 = 48$, $Q_3 = 60$

$IQR = 60 - 40 = 20$



↳ upper bound is biggest # in data $\leq Q_3 + 1.5IQR$

↳ lower bound is smallest # in data $\geq Q_1 - 1.5IQR$

WEEK 10

week 10

STATISTICAL MODELLING

- statistical model is identification of distribution from which data is drawn
  - ↳ data is $\{x_1, x_2, \ldots, x_n\}$ independently collected from some experiment
    - θ is parameter that rep unknown attribute of pop & try to find/estimate this using data
  - ↳ e.g.

Example

Suppose that we toss a coin 100 times and we are interested in finding $\theta = P(H)$

getting a head

$Y_i = \begin{cases} 1 & \text{if } H \\ 0 & \text{if } T \end{cases}$

$\{y_1, y_2, \ldots, y_n\}$ is data

Statistical model:

$Y_i \sim Ber(\theta)$, Y_i's independent, $i = 1, \ldots, 100$

  - ↳ e.g.

Example

$y_1, y_2, \ldots y_n$, where y_i = the starting salary of a SE grad from UW.

Suppose we look at data & it looks normal

Model: $Y_i \sim N(\mu, \sigma^2)$

- μ & σ^2 are unknown

- when finding a statistical model, note that:
  - ↳ attributes of pop we're interested in are typically parameters of model
  - ↳ this is an empirical question, not theoretical
    - i.e. based on data gathered by acc experiments ($P \simeq \frac{\text{\# observed successes}}{\text{total \# trials}}$)
- estimation is best "guess" for unknown parameters
  - ↳ given $X_1, X_2, \ldots, X_n$ is seq of iid rvs w/pmf $f(\theta)$, goal is to construct $\hat{\theta}(x_1, x_2, \ldots, x_n)$ = est of θ
    - θ is unknown parameter
    - data is $\{x_1, x_2, \ldots, x_n\}$ which are outcomes from rvs that we observe
      - → known #s
- when estimating, note that:
  - ↳ θ is unknown constant & will never be known
  - ↳ multiple methods exist
  - ↳ assume class of distribution has been properly identified
  - ↳ assume we have iid datasets
- e.g.

Example

A coin is tossed 100 times with $\theta = P(H)$. Suppose that we know the coin is biased:

$$\theta = \frac{1}{3} \text{ or } \theta = \frac{2}{3}$$

We run exp & observe 60 Hs so we choose $\theta = \frac{2}{3}$

Suppose $\theta = \frac{1}{3}$:

$P(\text{observed our data}) = \binom{100}{60}\left(\frac{1}{3}\right)^{60}\left(\frac{2}{3}\right)^{40}$

Suppose $\theta = \frac{2}{3}$:

$P(\text{observed our data}) = \binom{100}{60}\left(\frac{2}{3}\right)^{60}\left(\frac{1}{3}\right)^{40}$

Use likelihood → P(observing our data as factor of θ)

$L(\theta) = \binom{100}{60}\theta^{60}(1-\theta)^{40}$

$\hat{\theta} = \frac{2}{3}$ would maximize $L(\theta)$

if $Y_i \sim f(y_i, \theta)$, $i = 1, 2, \dots, n$, where Y_i are iid rvs w/ observations $\{y_1, y_2, \dots, y_n\}$, then likelihood function is

$$L(\theta; y_1, y_2, \dots, y_n) = P(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n)$$
$$= P(Y_1 = y_1) \cdot P(Y_2 = y_2) \cdots P(Y_n = y_n)$$
$$= f(y_1) \cdot f(y_2) \cdots f(y_n)$$
$$= \prod_{i=1}^{n} f(y_i, \theta)$$

↳ i.e. probability of observing our dataset as function of θ

- $\hat{\theta}$ is max likelihood estimate (MLE) if $\hat{\theta}$ maximizes $L(\theta; y_1, y_2, \dots, y_n)$
  ↳ i.e. uses value of θ that's most likely to have generated our data

- e.g.

Example

Suppose $Y \sim Bin(n, \theta)$, with y observed successes. Then what is $\hat{\theta}(y)$?

Suppose $n = 100$ & $y = 70$

$L(y; \theta) = P(Y = y) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y}$

To find MLE, take natural log of $L(\theta)$ to get $\ell(\theta) = \ln L(\theta)$

↳ $\hat{\theta}$ maximizes $L(\theta) \iff \hat{\theta}$ maximizes $\underset{\uparrow \text{ easier}}{\ell(\theta)}$

$L(\theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y}$ ← $k = \binom{n}{y}$

$$\ell(\theta) = \ln k + y\ln\theta + (n-y)\ln(1-\theta)$$
$$\frac{d\ell}{d\theta} = \frac{y}{\hat{\theta}} + \frac{n-y}{1-\hat{\theta}}(-1)$$
$$0 = \frac{y}{\hat{\theta}} - \frac{n-y}{1-\hat{\theta}}$$
$$y - y\hat{\theta} = \hat{\theta}n - y\hat{\theta}$$
$$y = \hat{\theta}n$$
$$\hat{\theta} = \frac{y}{n}$$
$$\hat{\theta} = \frac{70}{100} = 0.7$$

- e.g.

Example

Suppose $Y_1, \dots, Y_n \sim Poi(\theta)$, with observations/data of $\{y_1, \dots, y_n\}$. What is the MLE of θ?

$$L(\theta; y_1, \dots, y_n) = \prod_{i=1}^{n} f(y_i, \theta)$$
$$= P(Y_1 = y_1) \cdots P(Y_n = y_n)$$
$$= \frac{e^{-\theta}\theta^{y_1}}{y_1!} \cdots \frac{e^{-\theta}\theta^{y_n}}{y_n!}$$
$$= \frac{e^{-n\theta}\theta^{\sum y_i}}{y_1! y_2! \cdots y_n!}$$ ← let denom be k

$$\ell(\theta) = -n\theta + \sum y_i \ln\theta - \ln k$$
$$\frac{d\ell}{d\theta} = -n + \frac{\sum y_i}{\theta}$$
$$0 = -n + \frac{\sum y_i}{\hat{\theta}}$$
$$\hat{\theta} = \frac{\sum y_i}{n}$$
$$\hat{\theta} = \bar{y}$$ ← for Poisson, parameter is λ & $E(Y) = \lambda$ so parameter is mean

- for continuous models, $Y_1, \dots, Y_n \sim f(y_i; \theta)$ & objective is to find $\hat{\theta}$
  ↳ f is pdf of Y_i s, which are iid

↳ recall pmf gives prob directly but pdf $f(y_i)$ is not a prob

- e.g

Example

Suppose $Y_1, \ldots, Y_n \sim Exp(\lambda)$, with observations/data of $\{y_1, \ldots, y_n\}$.

observed waiting times

What is the MLE of λ?

$$L(\lambda) = \lambda e^{-\lambda y_1} \cdots \lambda e^{-\lambda y_n}$$
$$= \lambda^n e^{-\lambda \Sigma y_i}$$
$$\ell(\lambda) = n \ln \lambda - \lambda \Sigma y_i$$
$$\frac{d\ell}{d\lambda} = \frac{n}{\lambda} - \Sigma y_i$$
$$0 = \frac{n}{\hat{\lambda}} - \Sigma y_i$$
$$\hat{\lambda} = \frac{n}{\Sigma y_i}$$
$$\hat{\lambda} = \frac{1}{\bar{y}}$$

← recall for $X \sim Exp(\lambda)$, $E(X) = \frac{1}{\lambda} = \mu$ so $\lambda = \frac{1}{\mu}$

- e.g

Example

Suppose $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$, with observations/data of $\{y_1, \ldots, y_n\}$.

What is the MLE of $\underbrace{\mu \text{ and } \sigma^2}_{\text{unknown parameters}}$?

$$L(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y_1 - \mu}{\sigma}\right)^2} \cdots \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y_n - \mu}{\sigma}\right)^2}$$
$$= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} e^{-\frac{1}{2\sigma^2}\Sigma(y_i - \mu)^2}$$
$$\ell(\mu, \sigma^2) = -\frac{n}{2}\ln 2\pi - n\ln\sigma - \frac{1}{2\sigma^2}\Sigma(y_i - \mu)^2$$
$$\frac{d\ell}{d\mu} = \frac{1}{2\hat{\sigma}^2} \cdot 2 \cdot \Sigma(y_i - \hat{\mu})$$
$$0 = \frac{1}{\hat{\sigma}^2} \cdot \Sigma(y_i - \hat{\mu})$$
$$0 = \sum_{i=1}^{n}(y_i - \hat{\mu})$$
$$n\hat{\mu} = \Sigma y_i$$
$$\hat{\mu} = \bar{y}$$
$$\frac{d\ell}{d\hat{\sigma}} = -\frac{n}{\hat{\sigma}} + \frac{1}{\sigma^3}\Sigma(y_i - \hat{\mu})^2$$
$$0 = -\frac{n}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^3} \cdot \Sigma(y_i - \hat{\mu})^2$$
$$\sigma^2 = \frac{1}{n}\Sigma(y_i - \hat{\mu})^2$$
$$\sigma^2 = \frac{1}{n}\Sigma(y_i - \bar{y})^2$$

- properties of MLE:
  - ↳ consistency: as $n \to \infty$, $\hat{\theta} \to \theta$
    - i.e. the more data we collect, estimate converges to true value
  - ↳ efficiency: want min variance when finding $\hat{\theta}_i$s
  - ↳ invariance: if $\hat{\theta}$ is MLE of θ, then $g(\hat{\theta})$ is MLE of $g(\theta)$

- e.g.

Example

$Y_1, \ldots, Y_n \sim Exp(\lambda)$. Find the MLE for the median of Y.

Find median:

$$F(m) = \frac{1}{2}$$

→ $F(m) = P(Y \leq m)$, which is cdf of Exp

$$1 - e^{-\lambda m} = \frac{1}{2}$$
$$m = \frac{1}{\lambda}\ln\frac{1}{2}$$

Since $\hat{\lambda} = \frac{1}{\bar{y}}$, then $\hat{m} = -\bar{y}\ln 2$

relative likelihood function is $R(\theta) : \frac{L(\theta)}{L(\hat{\theta})}$

↳ since $L(\theta) = \prod_{i=1}^{n} f \cdot$, for large enough n, $L(\theta)$ is extremely small

↳ $R(\theta)$ is way to normalize $L(\theta)$

↳ $0 \leq R(\theta) \leq 1$ since $L(\hat{\theta})$ is maximized

- e.g. **Example**

  Suppose a coin is tossed 200 times and we observe $y = 80$ heads and $\theta = P(Heads)$

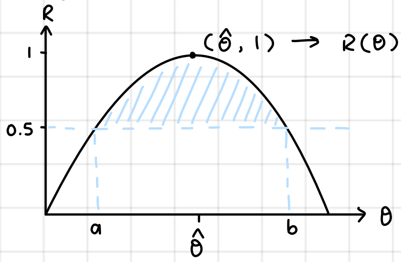
$$R(0.7) = \frac{L(0.7)}{L(\hat{\theta})} = \frac{L(0.7)}{L(80/200)} = \frac{\binom{200}{80} 0.7^{80}\, 0.3^{120}}{\binom{200}{80} 0.4^{80}\, 0.6^{120}}$$

← relative likelihood of θ being 0.7 given $\hat{\theta} = 0.4$

WEEK 11

INTRO TO INTERVAL ESTIMATION

- a $100p\%$ likelihood interval is $\{\theta : R(\theta) \geq p\}$
  - ↳ graph of relative likelihood function:



$(\hat{\theta}, 1) \rightarrow R(\hat{\theta}) = \frac{L(\hat{\theta})}{L(\hat{\theta})} = 1$

  - ↳ e.g. we want $p = 0.5$ so we use $\{\theta : R(\theta) \geq 0.5\}$ & say 50% likelihood interval is $[a, b]$
    - $R(\theta) \geq 0.5 \Rightarrow L(\theta) \geq 0.5 L(\hat{\theta})$
    - any value in $[a, b]$ is at min 0.5 of likelihood of MLE
      → i.e. 50% as likely as best value (MLE)
- rankings of plausability:
  - ↳ $R(\theta) \geq 0.5$: very plausible
  - ↳ $R(\theta) \geq 0.1$: plausible
  - ↳ $R(\theta) < 0.1$: implausible
  - ↳ $R(\theta) < 0.01$: very implausible
- e.g.

Example

A coin is tossed 200 times with $\theta = P(H)$. Suppose that we observe $y = 80$ successes. Find the 10% Likelihood Interval for θ.

↑ $p = 0.1$

$\hat{\theta} = \frac{80}{200} = 0.4$

Find all θ st $R(\theta) \geq 0.1$:

$R(\theta) \geq 0.1$

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

$$\frac{\binom{200}{80}\theta^{80}(1-\theta)^{120}}{\binom{200}{80}0.4^{80}\,0.6^{120}} \geq 0.1$$

↓ computer

$[0.33, 0.47]$ ← at min 10% as likely as $\hat{\theta} = 0.4$

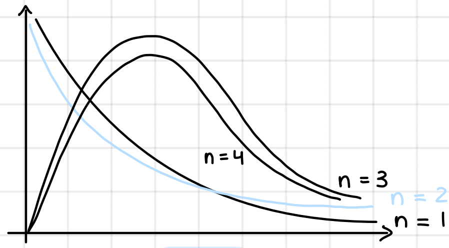
↳ Is $\theta = 0.36$ plausible?

Yes, $R(0.36) \geq 0.1$ so it's in plausible interval

CHI-SQUARED DISTRIBUTION

- Chi-squared (χ^2) distribution: let n be +ve int, then $W \sim \chi^2(n)$ if
  $W = Z_1^2 + Z_2^2 + \ldots + Z_n^2$
  - ↳ n is param that rep degrees of freedom (df)
  - ↳ $Z_i \sim N(0,1)$ & Z_i s are indep
- $f(w) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} w^{\frac{n}{2}-1} e^{-\frac{w}{2}}$
  - ↳ $w \geq 0$
  - ↳ $\Gamma(\frac{n}{2}) = \int_0^\infty x^{\frac{n}{2}-1} e^{-x}\,dx$

χ^2 distribution for special cases:



↳ if n=1 : $W = Z^2$

- e.g. let $W \sim \chi^2(1)$, what is $P(W \geq 1.44)$?

  $P(W \geq 1.44) = P(Z^2 \geq 1.44)$

  $= P(Z \leq -1.2) + P(Z \geq 1.2)$

↳ if n= 2 · $W \sim Exp(\lambda = \frac{1}{2})$

- $\frac{1}{2^{\frac{2}{2}}\Gamma(\frac{2}{2})} w^{\frac{2}{2}-1} e^{-\frac{w}{2}}$

  $= \frac{1}{2 \cdot 1} e^{-\frac{w}{2}}$

  $= \frac{1}{2} e^{-\frac{w}{2}}$

expectation is $E(X) = n = df$

↳ proof : $W \sim \chi^2(n)$

$W = Z_1^2 + Z_2^2 + \ldots + Z_n^2$

$E(W) = E(Z_1)^2 + \ldots + E(Z_n)^2$

$= 1 + \ldots + 1$

$= n$

$Z_i \sim N(0,1)$ so $E(Z_i) = 0$, $Var(Z_i) = 1$

$Var(Z_i) = E(Z_i^2) + (E(Z_i))^2$

$1 = E(Z_i^2) + 0^2$

$E(Z_i^2) = 1$

· variance is $Var(X) = 2n = 2df$

· applying CLT to χ^2, for a large n, $W \sim N(n, 2n)$

↳ e.g. $W \sim \chi^2(72)$, find $P(W > 96)$

Using CLT, $W \sim N(72, 144)$

$P(W > 96) = P(\frac{W-\mu}{\sigma} > \frac{96-72}{12})$

$= P(Z > 2)$

$= 1 - F(2)$

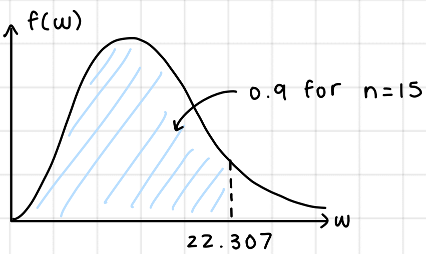
$= 1 - 0.97725$

$= 0.02275$

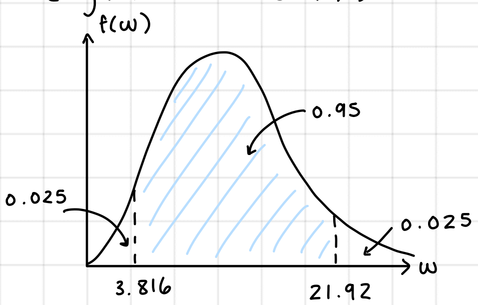
· use χ^2 table for intermediate n

↳ e.g. df = 15 , percentile = 0.9

$w = 22.307$



↳ e.g. let $W \sim \chi^2(11)$; find a & b st $P(a \leq W \leq b) = 0.95$



$a = 3.816$, $b = 21.92$

· rules of thumb :

↳ $n=1 \rightarrow W = Z^2$

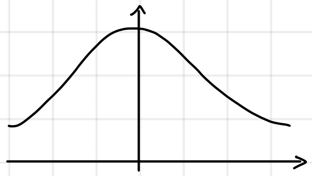
↳ $n=2 \rightarrow W \sim Exp(\frac{1}{2})$

↳ intermediate $n \rightarrow \chi^2$ table

↳ large $n \rightarrow W \sim N(n, 2n)$

STUDENT'S T-DISTRIBUTION

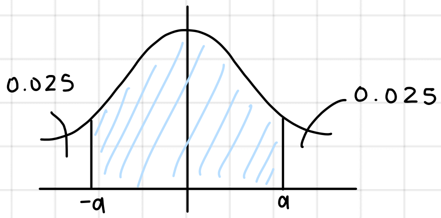
- T follows a Student's t-distr w/ n deg of freedom if T can be written as ratio of 2 indep rvs, Z & $\sqrt{\frac{W}{n}}$
  - ↳ written as $T \sim T(n)$
  - ↳ $T = \frac{Z}{\sqrt{\frac{W}{n}}}$
    - $Z \sim N(0,1)$
    - $W \sim \chi^2(n)$
- T is symmetric around 0
- properties:
  - ↳ support is $T: (-\infty, \infty)$
  - ↳ expectation is $E(T) = 0$
  - ↳ pdf is $\frac{\Gamma(\frac{n+1}{2})(1+\frac{t^2}{n})^{-\frac{n+1}{2}}}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})}$
- as $n \rightarrow \infty$, $T \rightarrow Z$
  - ↳



- "fatter" tails than normal distn
  - ↳ for large n, use z-table instead of t-table

e.g. suppose $T \sim T_{20}$, find a st $P(|T| \leq a) = 0.95$

↑ T w/ $n=20$ & $df=20$



$F(a) = 0.975$

$a = 2.0860$

- χ^2 distn is used for various tests & to define student's t-distn, which can be used as CLT sub when n is small

INTRO TO CONFIDENCE INTERVALS

- goal is to fix a prob (90%, 99%, etc.) & then find interval where 100p% of such intervals contain parameter of interest
- notation:
  - ↳ θ is unknown param
  - ↳ $\hat{\theta}$ is estimate to θ
    - # calc based data observed
  - ↳ $\tilde{\theta}$ is an rv that corresponds to $\hat{\theta}$
    - rv of which $\hat{\theta}$ is an outcome
- estimates are calc from samples
  - ↳ e.g. $\hat{\theta}$, $\bar{y}$, $\hat{u}$, etc.
- estimators are rvs

↳ e.g. $\tilde{\theta}$

↳ e.g. we take multiple samples & each have mean $\bar{y}_i$; each of $\bar{y}_i$ come from $\bar{Y}$

- $\bar{Y}$ is an estimator & y_is are estimates

· an $100p\%$ confidence interval for θ is an estimate of interval $[L, U]$ for rvs, $L(Y_1, \dots, Y_n)$ & $U(Y_1, \dots, Y_n)$ st $P(L(Y_1, \dots, Y_n) < \theta < U(Y_1, \dots, Y_n)) = p$

↳ interval is $[\ell(y_1, \dots, y_n), u(y_1, \dots, y_n)] = 100p\%$ CI

- ℓ is estimate from rv L
- u is estimate from rv U

· e.g. suppose we have 95% CI : $\hat{\theta} \pm a$, where $\hat{\theta} = 20$ & $a = 10$

$20 \pm 10 \rightarrow [10, 30]$ over 95% CI

↳ does not mean there's 95% this interval $[10, 30]$ contains param of interest

↳ instead, it means 95% of constructed CIs fall around (i.e contain) the param

· e.g.

Example: unrealistic to have σ^2 but not μ

Suppose that we have $Y_1, \dots Y_{25} \sim N(\mu, 144)$ and we have collected $\{y_1, \dots, y_{25}\}$ and $\bar{y} = 75$

Construct the 95% Confidence Interval for μ.

We will do this in 3 steps:

1. Construct the Pivotal Quantity
2. Using the Pivotal Distribution, construct the Coverage Interval
3. Estimate the Coverage Interval using your data

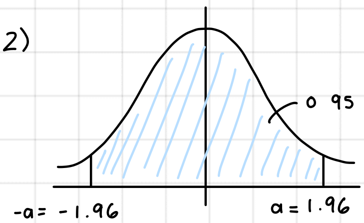
1) pivotal quantity is rv which is function of θ st distn is known w/o knowing value of θ

$Y_1, \dots, Y_{25} \sim N(\mu, 144)$ so $\bar{Y} = \frac{1}{25} \sum Y_i$

$\bar{Y} \sim N(\mu, \frac{144}{25}) \rightarrow \frac{\bar{Y} - \mu}{12/5} = Z \sim N(0,1)$

$\uparrow$ PQ

2)



$P(-1.96 < Z < 1.96) = 0.95$

$P(-1.96 < \frac{\bar{Y} - \mu}{12/5} < 1.96) = 0.95$

$P(\bar{Y} - 1.96(\frac{12}{5}) < \mu < \bar{Y} + 1.96(\frac{12}{5})) = 0.95$

$\uparrow$ random interval that contains μ w/ 95% probability

3) Best est of $\bar{Y}$ is $\bar{y}$

CI : $\bar{y} \pm a(\frac{\sigma}{\sqrt{n}})$

- $\frac{\sigma}{\sqrt{n}}$ is standard error

CI : $75 \pm 1.96 \left(\frac{12}{5}\right)$

NORMAL CONFIDENCE INTERVALS

· to get CI when σ^2 known :

↳ $Y_1, \dots, Y_n \sim N(\mu, \sigma_0^2)$ so $\bar{Y} \sim N(\mu, \frac{\sigma_0^2}{n})$

↳ $PQ = \frac{\bar{Y} - \mu}{\sigma_0 / \sqrt{n}} = Z \sim N(0,1)$

↳ find $-z^*$ & z^* st $P(-z^* < \frac{\bar{Y}-\mu}{\sigma_0/\sqrt{n}} < z^*) = p$

↳ CI: $\bar{y} \pm z^*\left(\frac{\sigma_0}{\sqrt{n}}\right)$

- to get CI when σ^2 is unknown, use theorem that if $Y_1, \dots, Y_n$ are iid $N(\mu, \sigma^2)$, $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, & $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$, then:
  1) $\frac{\bar{Y}-\mu}{S/\sqrt{n}} \sim T(n-1)$
  2) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

  ↳ $\bar{Y}$ & S are rvs while $\bar{y}$ & s are estimates

  ↳ recall for $N(\mu, \sigma^2)$ the MLE of σ^2 was $\hat{\sigma}^2 = \frac{1}{n}\sum(Y_i - \bar{Y})^2$ but also $S^2 = \frac{1}{n-1}\sum(Y_i - \bar{Y})^2$
  - S^2 is unbiased estimator of σ^2 but $\hat{\sigma}^2$ is biased b/c there's only $n-1$ deg of freedom
    → last value is dependent on other ones since there's a fixed mean/std dev
  - for estimate $s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$, use easier formula
    $s^2 = \frac{1}{n-1}\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right]$

  ↳ for $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$:
  - $E\left(\frac{(n-1)S^2}{\sigma^2}\right) = n-1$
    → since $E(W \sim \chi^2(n)) = n$
  - $E(S^2) = \sigma^2$
    → despite S^2 having $n-1$ as part of its defn, $E(S^2)$ is still σ^2

- e.g.

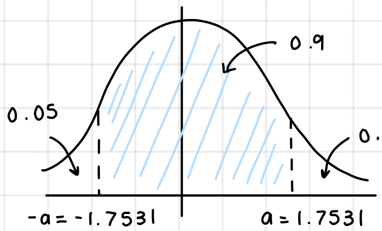
Example:

$Y_1, \dots, Y_n \sim N(\mu, \sigma^2)$, μ, σ are unknown.

A sample of 16 students were taken and it was calculated that $\bar{y} = 75$ an $s^2 = 100$. Find a 90% CI for μ.

PQ : $\frac{\bar{Y}-\mu}{\frac{S}{\sqrt{n}}} \sim T(n-1)$ ← prob distn (PD)

$\frac{\bar{Y}-\mu}{S/\sqrt{16}} \sim T(15)$



Cov interval:

$P(-1.75 < T(15) < 1.75) = 0.9$

$P(\bar{Y} - 1.75\frac{S}{\sqrt{n}} < \mu < \bar{Y} + 1.75\frac{S}{\sqrt{n}}) = 0.9$

CI: $\bar{y} \pm 1.75\left(\frac{10}{4}\right) = 75 \pm 4.375$

MORE CIS - VARIANCE, BINOMIAL, AND POISSON

CI to est σ^2:

↳ PQ: if $Y_1, \dots, Y_n \sim N(\mu, \sigma^2)$ & $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$, then $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

↳ cov interval: $\left[\frac{(n-1)S^2}{a}, \frac{(n-1)S^2}{b}\right]$

↳ CI: $\left[\frac{(n-1)s^2}{a}, \frac{(n-1)s^2}{b}\right]$
  - square root everything to est σ

- binomial CI:

  ↳ PQ: if $Y \sim Bin(n, \theta)$, then $Y \sim N(n\theta, n\theta(1-\theta))$
  - using CLT

  ↳ pivotal distn (PD): $\frac{Y - n\theta}{\sqrt{n\theta(1-\theta)}} = Z \sim N(0,1)$

  ↳ cov interval: $\tilde{\theta} \pm z^*\sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{n}}$

  ↳ CI: $\hat{\theta} \pm z^*\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$

  ↳ e.g. suppose we poll 1000 Americans & 400 approve of candidate A (600 disapprove); construct 95% CI for approval rating
  $Y \sim Bin(1000, \theta)$ but it's not quite binomial b/c we're sampling w/o replacement but

it's good apx since N is large & n is small

$\hat{\theta} \pm z^* \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$

$\hat{\theta} = \frac{400}{1000} = 0.4$

$z^* = 1.96$

$n = 1000$

CI : $0.4 \pm 1.96 \sqrt{\frac{0.4(0.6)}{1000}}$

- margin of error (ME) is $1.96 \sqrt{\frac{0.4(0.6)}{1000}}$
- to get ME < 0.03, we need $0.03 > 1.96 \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$
- we can control n but we're at mercy of $\hat{\theta}$
- ME is max when $\hat{\theta} = 0.5$

· poisson CI ·

↳ PQ : if $Y_1, \ldots, Y_n \sim Poi(\theta)$ then $\bar{Y} \sim N(\theta, \frac{\theta}{n})$

- using CLT

↳ PD : $\frac{\bar{Y} - \theta}{\sqrt{\bar{Y}/n}} = Z \sim N(0,1)$

↳ cov interval : $\bar{Y} \pm z^* \sqrt{\frac{\bar{Y}}{n}}$

↳ CI : $\bar{y} \pm z^* \sqrt{\frac{\bar{y}}{n}}$

Choice of distribution:

↳ σ^2 / σ is known : z - table

↳ σ^2 / σ is unknown :
t - table

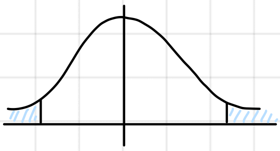
↳ estimate σ^2 / σ : χ^2 distn

week 12

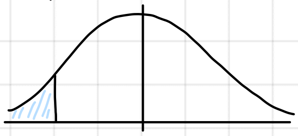
INTRO TO HYPOTHESIS TESTING

- hypothesis is some claim abt population
  - ↳ usually made abt param
- null hypothesis is current belief
  - ↳ denoted by H_0
  - ↳ conventional wisdom
- alternate hypothesis is challenge to H_0
- 3 types of tests: two-tailed, one-tailed on left, & one-tailed on right
  - ↳ we'll use two-tailed tests
  - ↳ e.g.

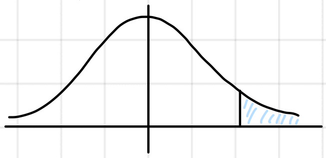
Two-tailed:
$H_0: \theta = \frac{1}{6}$
$H_1: \theta \neq \frac{1}{6}$



One-tailed on left:
$H_0: \theta \geq \frac{1}{6}$
$H_1: \theta < \frac{1}{6}$



One-tailed on right:
$H_0: \theta \leq \frac{1}{6}$
$H_1: \theta > \frac{1}{6}$



- method to hypothesis testing:
  1) construct test statistic
  2) calc value of test statistic
  3) compute p-value
  4) draw appropriate conclusions
- test statistic is an rv that allows us to calc probabilities given H_0 is true
- p-value is probability of observing our evidence given H_0 is true
  - ↳ convention is:
    - $p < 0.05$: strong support of H_1
    - $p < 0.01$: v strong support of H_1
    - $p \geq 0.05$: insufficient evidence against H_0
  - ↳ level of significance is probability that event occurred by chance (i.e. probability of rejecting H_0 when it's true)
    - if level is low, we say event is significant
  - ↳ e.g. decide on following p-values before conducting study
    - $p < 0.05$ → reject H_0, fail to reject H_1
    - $p \geq 0.05$ → fail to reject H_0, reject H_1
- 2 types of errors:
  - ↳ type I: P(rejecting H_0 if H_0 is true) = p-value
  - ↳ type II: P(failing to reject H_0 if H_0 is false) = $1 - \beta$
    - β is power of test
  - ↳

	H_0 true	H_0 false
reject H_0	✗ (type I)	✓
don't reject H_0	✓	✗ (type II)

NORMAL HYPOTHESIS TESTING

- rv D is test statistic if :
  - ↳ distn of D is known if H_0 is true
  - ↳ when $D = 0$, we have strongest possible evidence in favour of H_0
    - as $|D|$ becomes larger, more evidence against H_0
  - ↳ i.e. D is discrepancy measure, meaning how much data disagrees w/ H_0
- e.g.

Example

Suppose $Y_1, \ldots, Y_{25} \sim N(\mu, 144)$, Y_i's independent.
$n = 25$ and $\bar{y} = 50$

↓ σ^2 is known

$H_0 : \mu = 45$
$H_1 : \mu \neq 45$

Can we conclude that our sample data supports H_1?

1) $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$

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

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

If H_0 is true, $\frac{\bar{Y} - 45}{12/5} = Z \sim N(0,1)$

- $D = \left| \frac{\bar{Y} - 45}{12/5} \right| = |Z|$

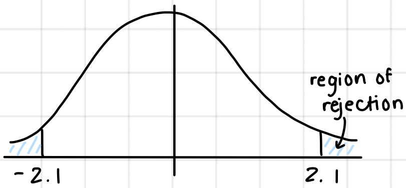
2) $d = \frac{\bar{y} - 45}{12/5}$

$= \frac{50 - 45}{12/5}$

$\simeq 2.1$

3) p-value $= P(D \geq d) < 0.05$

$= P(|Z| \geq 2.1) < 0.05$



$P(Z \leq 2.1)$

$= 2(1 - 0.98214)$

$= 0.03572$

4) Since $p < 0.05$, we have strong evidence against H_0 (reject H_0, fail to reject H_1)

- e.g.

Example

Suppose $Y_1, \ldots, Y_{25} \sim N(\mu, \sigma^2)$, Y_i's independent.
$n = 25$, $s^2 = 169$, and $\bar{y} = 54$

σ^2 is unknown

$H_0 : \mu \leq 50$
$H_1 : \mu > 50$

Can we conclude that our sample data supports H_1?

1) $\frac{\bar{Y} - \mu_0}{S/\sqrt{n}} \sim T_{n-1}$

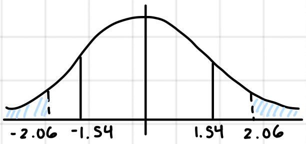
$D = \left| \frac{\bar{Y} - 50}{S/\sqrt{n}} \right|$

2) $d = \left| \frac{\bar{y} - 50}{13/5} \right|$

$= \left| \frac{54 - 50}{13/5} \right|$

$\simeq 1.54$

3) Use t-table w/ $df = 24$ → $t = 2.0639$ for $p = 0.975$



4) Since $p > 0.05$, we fail to reject H_0 & reject H_1 (not enough evidence to support H_1).

MORE HYPOTHESIS TESTING

- e.g.

Example (Binomial Hypothesis Test)

A survey of 1000 Americans asked who they would vote for; Biden (52%) or Trump (48%).

Is this election too close to call?

$H_0 : \theta = \frac{1}{2}$

$H_1 : \theta \neq \frac{1}{2}$

If Y = # ppl voting for Biden → $Y \sim Bin(1000, \theta)$

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

↑ PQ for Bin CI

$$Z = \frac{Y - 1000(\frac{1}{2})}{\sqrt{1000(\frac{1}{2})(\frac{1}{2})}}$$

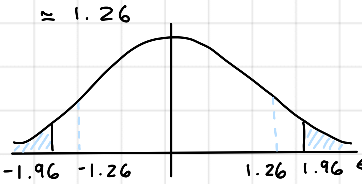
$$D = \left| \frac{\bar{Y} - 500}{\sqrt{250}} \right|$$

2) $d = \left| \frac{\bar{y} - 500}{\sqrt{250}} \right|$

$= \left| \frac{520 - 500}{\sqrt{250}} \right|$

$\simeq 1.26$

3)



← $F(1.96) = 0.975$ so there's 2.5% for each tail

$\underbrace{P(D > 1.26)}_{\text{calc area of tails}} = 2(1 - 0.89617) > 0.05$

4) Since $p > 0.05$, fail to reject H_0 & reject H_1 (too close to call).

- e.g.

Example (Poisson Hypothesis Test)

Suppose that we can model the number of Twitter employees that Elon Musk will fire in a given time frame using a Poisson distribution:

$$Y_1, \ldots, Y_{49} \sim Poi(\lambda)$$

$H_0 : \lambda = 16, H_1 : \lambda \neq 16$

If we collect our data, $\{y_1, \ldots y_{49}\}$ and find $\bar{y} = 15$, then what should we conclude?

1) $\bar{Y} \sim N(\lambda, \frac{\lambda}{n})$ by CLT if n is large ← $n \geq 30$ is large enough

$D = \left| \frac{\bar{Y} - \lambda}{\sqrt{\lambda / n}} \right| = Z$

$D = \left| \frac{\bar{Y} - 16}{\sqrt{16/49}} \right|$

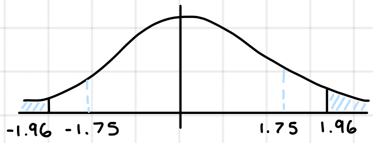
2) $d = \left| \frac{\bar{y} - 16}{4/7} \right|$

$= \left| \frac{15 - 16}{4/7} \right|$

$\simeq 1.75$

3) p-value $= P(D \geq d)$

$= P(|Z| \geq 1.75) > 0.05$



4) Not enough evidence to reject H_0

· e.g.

Example (Exponential Hypothesis Test)

$Y_1, \ldots, Y_{100} \sim Exp(\mu = \frac{1}{\lambda})$

$H_0 : \cancel{X}\,\mu = 75$

$H_1 : \cancel{X}\,\mu \neq 75$

Data: $\{y_1, \ldots y_{100}\}$ and find $\bar{y} = 60$, then what should we conclude?

1) $Y_1, \ldots, Y_n \sim Exp \Rightarrow \bar{Y} \sim N(\mu, \mu^2)$ where $\mu = \frac{1}{\lambda}$

PQ is $\frac{\bar{Y} - \mu}{\mu / \sqrt{n}} = Z$

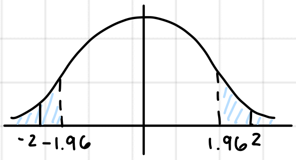
$D = \left| \frac{\bar{Y} - 75}{75 / \sqrt{100}} \right|$

2) $d = \left| \frac{60 - 75}{75 / 10} \right|$

$= 2$

3) $P(D \geq d) = P(|Z| \geq 2) < 0.05$

$z^* = 1.96$ for $F(z) = 0.975$



4) Since $p < 0.05$, we reject H_0 & have enough evidence to support H_1

· to illustrate rltnship btwn CIs & hypothesis tests, suppose we have unknown param θ & we want a 95% CI vs we conduct a hypothesis test where $H_0 : \theta = \theta_0$ & $H_1 \neq \theta_0$ (using 0.05 cutoff for p-value)

↳ for CI : $\theta_0 \pm a$

↳ for hypothesis test: if it's in $\theta_0 \pm a \rightarrow H_0$ & if not $\rightarrow H_1$

same idea

HYPOTHESIS TESTING WITH TWO MEANS

· matched pair : there's 1-1 map that could be drawn btwn individual data points in a pop

↳ i.e. twin studies, follow-up studies

↳ e.g. group 1 takes placebo & group 2 takes vitamin; after 3 months, switch so that group 1 takes vitamin & group 2 takes placebo

· e.g

Problem (Matched Pairs)

$A_1, \ldots, A_{n_1} \sim N(\mu_1, \sigma_1^2)$

$B_1, \ldots, B_{n_2} \sim N(\mu_2, \sigma_2^2)$

Pairs : (A_i, B_i), $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 \neq \mu_2$

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

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

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

$Z = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ $\longleftarrow$ assume H_0 $(\mu_1 = \mu_2)$ is true

$$D = \frac{|\bar{Y}_1 - \bar{Y}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$\downarrow$ if σ_1^2 & σ_2^2 unknown

$$D = \frac{|\bar{Y}_1 - \bar{Y}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1)$$

↳ only if n is large enough

· e.g.

Problem (Unmatched Pairs w/ Large Sample)

$Y_{11}, \ldots, Y_{1n_1} \sim N(\mu_1, \sigma_1^2)$
$Y_{21}, \ldots, Y_{2n_2} \sim N(\mu_2, \sigma_2^2)$

$$D = \frac{|\bar{Y}_1 - \bar{Y}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n}}} \sim N(0,1)$$

↳ when n is large, $\sigma^2 \to s^2$

· e.g.

Problem (Unmatched Pairs w/ Equal Variance)

$Y_{11}, \ldots, Y_{1n_1} \sim N(\mu_1, \sigma^2)$
$Y_{21}, \ldots, Y_{2n_2} \sim N(\mu_2, \sigma^2)$ $\longrightarrow$ indep

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

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

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

$$D = \frac{|\bar{Y}_1 - \bar{Y}_2|}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = |Z|$$

↳ if we don't know σ but n isn't large:

$$D = \frac{|\bar{Y}_1 - \bar{Y}_2|}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim T_{n_1+n_2-2}$$

- $d = \frac{|\bar{y}_1 - \bar{y}_2|}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$
- $s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$

WEEK 13

week 13

GOODNESS OF FIT

- theorem: $\Lambda = 2\sum Y_j \ln \frac{Y_j}{E_j} \sim \chi^2(k-1-p)$ is a test statistic for goodness of fit
  - ↳ value of test statistic is $\lambda = 2\sum y_j \ln \frac{y_j}{e_j} \sim \chi^2(k-1-p)$
  - ↳ p-value is $P(\Lambda \geq \lambda)$
  - ↳ k is # categories
  - ↳ p is # parameters estimated under H_0
  - ↳ y_j is observed freq
  - ↳ e_j is expected freq under H_0
- e.g.

Example

Suppose that we have a 6-sided fair die. How can we test that it is fair?

$H_0 : \pi_1 = \pi_2 = \ldots = \pi_6 = \frac{1}{6}$

Roll the die 300 times then calculate the frequency of each outcome.

X	y_j (observed freq)	e_j
1	30	50
2	70	50
3	40	50
4	60	50
5	50	50
6	50	50

Test statistic:

$\lambda = 2(30\ln\frac{30}{50} + 70\ln\frac{70}{50} + \ldots) \longrightarrow$ if $\forall j,\ y_j = e_j$ then $\lambda = 0$

For $H_0 = \pi_1 = \pi_2 = \ldots = \pi_6 = \frac{1}{6}$, all params are given so $p = 0$

$\Lambda \sim \chi^2(6-1-0)$

$= \Lambda \sim \chi^2(\underset{\substack{\uparrow \\ df = 5}}{5})$

- e.g. Poisson example:

Number of α-particles detected: j	Observed Frequency: f_j	Expected Frequency: e_j
0	57	54.3
1	203	210.3
2	383	407.1
3	525	525.3
4	532	508.4
5	408	393.7
6	273	254.0
7	139	140.5
8	45	68.0
9	27	29.2
10	10	11.3
≥ 11	6	5.8
Total	2608	2607.9

← expected freq $= np_i$

n = # observations = 2608

p_i = expected prob = $\frac{e^{-\theta}\theta^i}{i!}$

↳ e.g. $p_0 = P(X=0)$

$= e^{-\theta}\frac{\theta^0}{0!}$

MLE of θ:

$\hat{\theta} = \frac{1}{n}\sum x_i$

$= 3.8715$ (weighted avg of j & f_j)

$H_0: X_i \sim Poi(\theta)$

$\lambda = 2 \sum y_j \ln \frac{y_j}{e_j}$

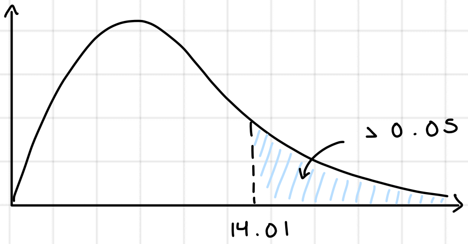
$= 2(57 \ln \frac{57}{54.3} + \ldots)$

$= 14.01$

$p = 1$ b/c we estimated θ

p-value $= P(\Lambda \geq \lambda = 14.01) \sim \chi^2(12-1-1) = \chi^2(10)$

$= 0.17$



> 0.05 so cannot reject H_0

· e.g. exponential example:

$X_1, \ldots, X_n$

$H_0: X_i \sim Exp(\theta)$

X	Y_i (# obs)	E_i
$[0,10]$	20	np_1
$(10,20]$	40	.
$(20,40]$	20	.
>40	10	np_4

$\rightarrow p_1 = P(0 \leq X \leq 10)$

$= F(10) - F(0)$

$= 1 - e^{-10\lambda} - (1 - e^0)$

for exp, $F(x) = 1 - e^{-\lambda x}$ &

$\theta = \frac{1}{\lambda}$ so we need to estimate $\hat{\theta} = \frac{1}{n} \underbrace{\sum x_i y_i}_{\text{weighted avg}}$

$\lambda = 2 \sum y_i \ln \frac{y_i}{e_i}$

p-value $= P(\Lambda \geq \lambda)$ using $\chi^2(k-1-p) = \chi^2(4-1-1) = \chi^2(2)$

EQUALITY OF PROPORTIONS AND INDEPENDENCE OF ATTRIBUTES

· motivating example: LH vs RH smokers

↳ $H_0: \pi_L = \pi_R$

- π_L is proportion of LH smokers
- π_R is proportion of RH smokers

↳ $H_1: \pi_L \neq \pi_R$

↳ contingency table (aka 2-way freq table):

	Smokers	Non-Smokers
LH	y_{11}	y_{12}
RH	y_{21}	y_{22}

- testing proportion of LH & RH smokers is same as testing for independence of attributes

→ 2 attributes are handedness & smoker status

· e.g.

	Covid	No Covid	
Moderna	64	176	240
Placebo	86	150	236
	150	326	476

π_1 = proportion of ppl catching Covid | had Moderna

π_2 = proportion of ppl catching Covid | had placebo

$H_0: \pi_1 = \pi_2$ (attributes are independent)

$H_1: \pi_1 \neq \pi_2$ (attributes are not independent)

To test if attributes 1 (vaccinated) & 2 (caught Covid) are indep, construct expected values table:

	Covid	No Covid	
Moderna	e_{11}	e_{12}	240
Placebo	e_{21}	e_{22}	236
	150	326	476

↳ e_{ij} are values if attributes are indep

↳ e.g. $e_{11} = n \; P(\text{Moderna}) \cdot P(\text{Covid})$

$= 476 \left(\frac{240}{476}\right)\left(\frac{150}{476}\right)$

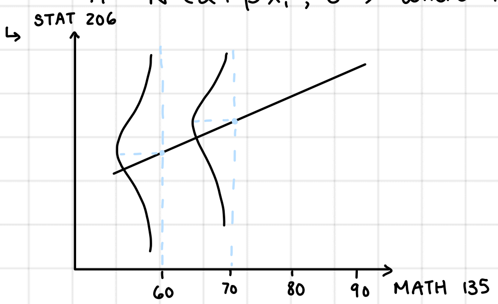
$\lambda = 2 \sum\sum y_{ij} \ln \frac{y_{ij}}{e_{ij}} \sim \chi^2((a-1)(b-1)) = \chi^2(1)$

- to test for independence of attributes:
  1) make expected values table
     - $e_{ij} = \frac{(i^{th} \text{ row total})(j^{th} \text{ column total})}{\text{total}}$
  2) calculate $\lambda = 2\sum\sum y_{ij} \ln \frac{y_{ij}}{e_{ij}}$
     - y_{ij} is observed freq
     - e_{ij} is expected freq if attributes are indep
  3) calculate p-value where $\Lambda \sim \chi^2((a-1)(b-1))$ & if p-value ≤ 0.02, can reject H_0
     - a : # rows
     - b: # columns

LINEAR REGRESSION

- linear regression attempts to model rltnship btwn 2 vars by fitting linear eqn to observed data (used for predictive analysis)
- e.g Y_i = grade in STAT 206, X_i = grade in MATH 135

  ↳ can use simple linear regression model to predict X_i based on Y_i
  - $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$ where $i = 1, \ldots, n$ & y_i s are indep

  ↳ 

  - assumes that Y_i follows normal distn no matter the value of X_i
  - have diff means but they all lie on straight line $y = \underbrace{\alpha + \beta x}_{\mu}$ w/variance σ^2
  - X : explanatory var
  - Y: response var

  ↳ use datapoints $(x_1, y_1), \ldots, (x_n, y_n)$ to estimate $\hat{\alpha}$, $\hat{\beta}$, & $\hat{\sigma}^2$
  - if $\beta = 0$, then X has no predictive power for Y
- make Gauss-Markov assumptions when doing linear regression:

  ↳ X_i & Y_i are normally distributed

  ↳ mean is $\mu = \alpha + \beta x$

  ↳ Y_i s are indep

  ↳ $Var(Y_i) = \sigma^2$
- formulation for simple linear regression model is $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$ where $i = 1, \ldots, n$ are indep

  ↳ alt formulation is $Y_i = \underbrace{\alpha + \beta x_i}_{\text{explained}} + \underbrace{R_i}_{\text{unexplained (i.e. residuals)}}$

- $R_i \sim N(0, \sigma^2)$ where R_is are indep

↳ likelihood function is $L(\alpha, \beta, \sigma) = \prod_{i=1}^{n} f(y_i)$

$= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y_i - (\alpha + \beta x_i))^2}$

$L(\alpha, \beta, \sigma) = \frac{1}{\sigma^n (2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2}\Sigma(y_i - (\alpha + \beta x_i))^2}$

- to get MLEs, set $\frac{dl}{d\sigma} = 0$, $\frac{dl}{d\alpha} = 0$, & $\frac{dl}{d\beta} = 0$

↳ $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$

↳ $\hat{\beta} = \frac{S_{xy}}{S_{yy}} = \frac{\Sigma((x_i - \bar{x})(y_i - \bar{y}))}{\Sigma(x_i - \bar{x})^2}$

↳ $\hat{\sigma}^2 = \frac{1}{n}\Sigma(y_i - (\hat{\alpha} - \hat{\beta}x_i))^2$

- $Y = \hat{\alpha} + \hat{\beta}X$ is called line of best fit

- e.g.

x	y	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$
1	2	-4	-8	16	32
3	7	-2	-3	4	6
5	11	0	1	0	0
7	15	2	5	4	10
9	15	4	5	16	20

$\bar{x} = \frac{25}{5} = 5$ $\qquad S_{xx} = 40$ $\qquad S_{xy} = 68$

$\bar{y} = \frac{50}{5} = 10$

$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$ $\qquad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$

$= \frac{68}{40}$ $\qquad = 10 - 1.7(5)$

$= 1.7$ $\qquad = 1.5$

Line of best fit is $Y = 1.5 + 1.7X$

↳ e.g. our best guess for Y when $X = 5$ is $Y = 1.5 + 1.7(5) = 10$

- for hypothesis testing regarding regression, $\tilde{\beta} \sim N(\beta, \frac{\sigma^2}{S_{xx}})$ so $\frac{\tilde{\beta} - \beta}{\sigma/\sqrt{S_{xx}}} \sim N(0,1)$

↳ null hypothesis is $\beta = 0$ so $\frac{\tilde{\beta}}{\sigma/\sqrt{S_{xx}}} \sim N(0,1)$

- use z-table

↳ use t-table for probability distn w/ $df = n-2$ (since we're est α & β) when σ is unknown

- $\frac{\tilde{\beta}}{s_e/\sqrt{S_{xx}}} \sim T_{n-2}$

↳ test statistic is $D = \left|\frac{\tilde{\beta}}{s_e/\sqrt{S_{xx}}}\right|$

- $s_e^2 = \frac{1}{n-2}(\Sigma(y_i - (\hat{\alpha} - \hat{\beta}x_i))^2)$