YES! "success" | **NO!!** "failure"

BINOMIAL DISTRIBUTION

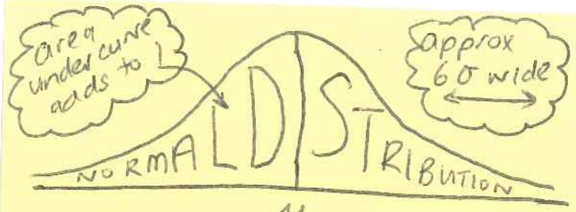
discrete data "count the successes"

n = number of trials
$\pi = P(\text{success})$

Assumptions

- **F**ixed no. of trials
- Each trial is **I**ndependent
- P(success) remains **C**onstant
- **T**wo outcomes in each trial: success & failure.

NORMAL DISTRIBUTION



⟵ continuous data ⟶
(but can model discrete data with correction)

μ = mean σ = standard deviation

Assumptions

- **B**ell shaped curve with tendency towards centre
- no **m**in / max (in theory!)
- **S**ymmetrical about the mean

POISSON DISTRIBUTION

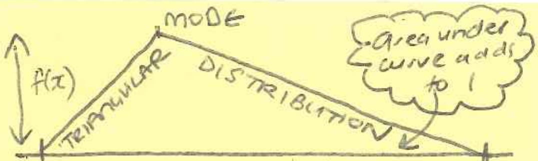
⟵ in a continuous interval ⟶

λ = average

Assumptions

- rare and **R**andom event
- each event is **I**ndependent
- probability is **P**roportional to the size of the interval
- 2 events cannot occur **S**imultaneously

TRIANGULAR DISTRIBUTION



min ⟵ continuous data ⟶ MAX

a = min
b = max *
c = mode *

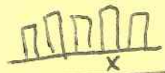
* WARNING!! Take care

Assumptions

- we know **m**in / MODE / max
- good if **S**kewed data but also ok if **S**ymmetrical
- **S**teady increase to mode and then decrease
- * pdf gives height of Δ not the area (use $\frac{1}{2} \times b \times h$)

UNIFORM DISTRIBUTION

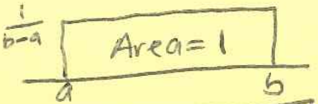
DISCRETE



n = no. of equally likely events

- $P(x) = \frac{1}{n}$

CONTINUOUS



a = min b = max

- pdf gives height of rectangle

Assumptions

- only know **m**in / max
- **E**qually likely events
- **E**ven distribution of values

CLUELESS DISTRIBUTION

our goal

TOP SECRET
"unknown"

TRUE PROBABILITY

so we estimate ↑ in 2 ways

EXPERIMENTAL ESTIMATE	MODEL ESTIMATE
• 'real life' data • simulation	• use prob. distribution • needs parameters
• sample size • variation	• assumptions made

MEANS OF PROBABILITY DISTRIBUTIONS

measures of centre

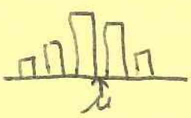
$\mu = E[X] = \sum x . P(X=x)$

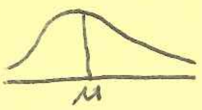
a.k.a. Expected value.

normal	Binomial	Poisson
μ	$n\pi$	λ

GRAPH

Estimate!!





TABLE

Find $E[X]$

x	1	2	3
P(X=x)	0·4	0·1	0·5

$E[X] = 1 \times 0{\cdot}4 + 2 \times 0{\cdot}1 + 3 \times 0{\cdot}5 = 2{\cdot}1$

STANDARD DEVIATION / VARIANCE OF PROBABILITY DISTRIBUTIONS

measures of spread

$SD = \sqrt{VAR}$

$VAR = (SD)^2$

OR

$\sigma = \sqrt{\sigma^2}$

$\sigma^2 = \sigma$

normal	Binomial	Poisson
σ	$\sqrt{n\pi(1-\pi)}$	$\sqrt{\lambda}$

GRAPH σ is about $\frac{1}{6}$ range

TABLE use $VAR[X] = E[X^2] - (E[X])^2$

x	1	2	3
x^2	1	4	9
P(X=x)	0·4	0·1	0·5

- $E[X] = 2{\cdot}1$
- $E[X^2] = 1 \times 0{\cdot}4 + 4 \times 0{\cdot}1 + 9 \times 0{\cdot}5 = 5{\cdot}3$
- $VAR[X] = 5{\cdot}3 - 2{\cdot}1^2 = 0{\cdot}89$
- $SD[X] = \sqrt{0{\cdot}89} = 0{\cdot}94$

COMBINING PROBABILITY DISTRIBUTIONS

$E[X+Y] = E[X] + E[Y]$; $E[X-Y] = E[X] - E[Y]$

$VAR[X \pm Y] = VAR[X] + VAR[Y]$

* always add variances (don't subtract)

$E[aX] = aE[X]$ | $VAR[aX] = a^2 VAR[X]$

Let F = Food weight   C = Can weight

T = Total weight food & can (F+C)

$E[F] = 450g$	$E[C] = 40g$
$VAR[F] = 100g$	$VAR[C] = 36g$

- $E[T] = 450 + 40 = 490g$
- $VAR[T] = 100 + 36 = 136g$
- $SD[T] = \sqrt{136} = 11.7g$
- $E[3T] = 3 \times 490 = 1470g$
- $VAR[3T] = 9 \times 136 = 1224g$

Assumes F and C are independent

KEY PROBABILITY VOCAB

RANDOM EXPERIMENT
- repeated trials of an
- unpredictable process

e.g. "rolling a die 60 times"

TRIAL: one repeat of experiment — e.g. "one die roll"

OUTCOME: result of 1 trial — "getting a 2"

SAMPLE SPACE: all possible outcomes

$\{1, 2, 3, 4, 5, 6\}$

EVENT: subset of sample space

"rolling an even number" $\{2, 4, 6\}$

RANDOM VARIABLE: a variable whose value is the outcome of a random expt.

"X = number of 6s"

TWO WAY TABLES

Finding probabilities for 2 events

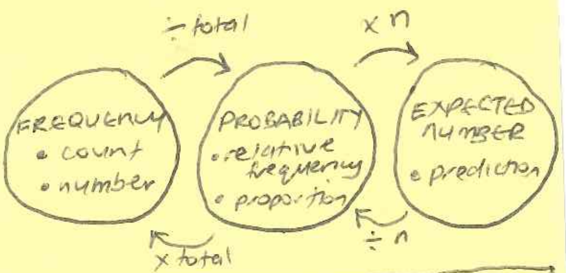
R = Rains

F = Forecast to rain

	F	F'	
R	18	4	22
R'	1	7	8
	19	11	30

- P(Rains) $= P(R) = \frac{22}{30}$
- P(Rains and Forecast) $= P(R \cap F) = \frac{18}{30}$
- P(Rains OR forecast) $= P(R \cup F) = \frac{23}{30}$
- P(Rains given is forecast) $= P(R|F) = \frac{18}{19}$

Linking Probability Info



Let X = colour of mms in bag of 32

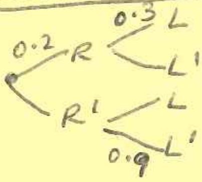
Number of red mm	P(red)	Expected reds in bag of 500
4	$= \frac{4}{32}$	$= \frac{4}{32} \times 500$
	$= 0.125$	$= 62.5$

TREE DIAGRAMS

Finding probabilities for 2 (or more) consecutive events

R = Rains

L = Late to school



- P(Does not rain) $= P(R') = 0.8$
- P(Late given that rains) $= P(L|R) = 0.3$
- P(rains and late) $= P(R \cap L) = 0.06$
- P(Late) $= P(R \cap L) + P(R' \cap L)$
  $= 0.06 + 0.08$
  $= 0.14$

PROBABILITY RULES TO KNOW (NOT GIVEN)

- $P(A) = 1$ "certain"
- $P(A) = 0$ "impossible"

$$\boxed{0 \leq P(A) \leq 1}$$

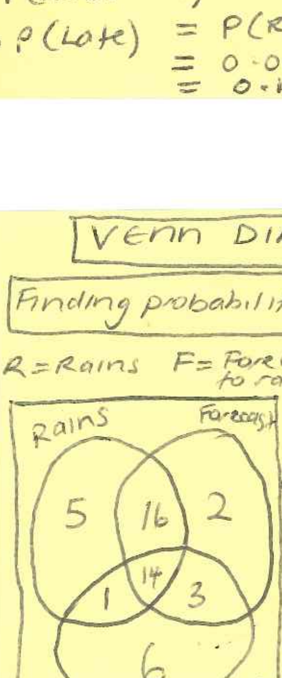
"probabilities always bet. 0 & 1"

- $P(A) + P(A') = 1$ "complementary"
- $P(A \cap B) = 0$ "mutually exclusive"
- $P(A) \times P(B) = P(A \cap B)$ "independent"
- $P(A|B) = P(A)$ "independent"

VENN DIAGRAMS

Finding probabilities for 2+ events

R = Rains F = Forecast to rain L = Late to school

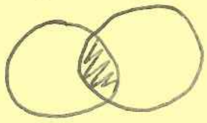


- P(Late) $= \frac{24}{50}$
- P(rains and forecast)
  $P(R \cap F) = \frac{30}{50}$
- P(rains or late)
  $P(R \cup L) = \frac{45}{50}$
- P(rains & forecast & late)
  $P(R \cap F \cap L) = \frac{14}{50}$
- P(rains given forecast)
  $P(R|F) = \frac{30}{35}$

INTERSECTION OF EVENTS

"both A AND B"

$\boxed{P(A \cap B)}$



UNION OF EVENTS

"either A OR B (or both)"

$\boxed{P(A \cup B)}$

INDEPENDENT EVENTS

Events are independent if the occurrence of Event A does not effect the probability of Event B

$\boxed{P(A) \times P(B) = P(A \cap B)}$

TRUE → Events A & B are independent

FALSE → Events A & B are NOT independent

MUTUALLY EXCLUSIVE EVENTS

A ○ ○ B

$P(A \cap B) = 0$

Events cannot both occur

"winning & losing a game of netball"

COMPLEMENTARY EVENTS

A A'

The complement of Event A is "not A"

$P(A) + P(A') = 1$

"A = winning"

"A' = not winning"

would include a draw..

CONDITIONAL PROBABILITY

Probability of Event A given that Event B has already happened.

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

aka: absolute risk of A for group B

and: If events are independent $P(A/B) = P(A)$

RISK

The probability of an event

e.g. P(injured in car accident)

ABSOLUTE RISK

The conditional probabilities of an event for certain groups

e.g. P(injured | wearing seatbelt)

RELATIVE RISK

Compares absolute risk for 2 groups

$$RR = \frac{P(\text{injured} \mid \text{wearing seatbelt})}{P(\text{injured} \mid \text{not wearing seatbelt})}$$