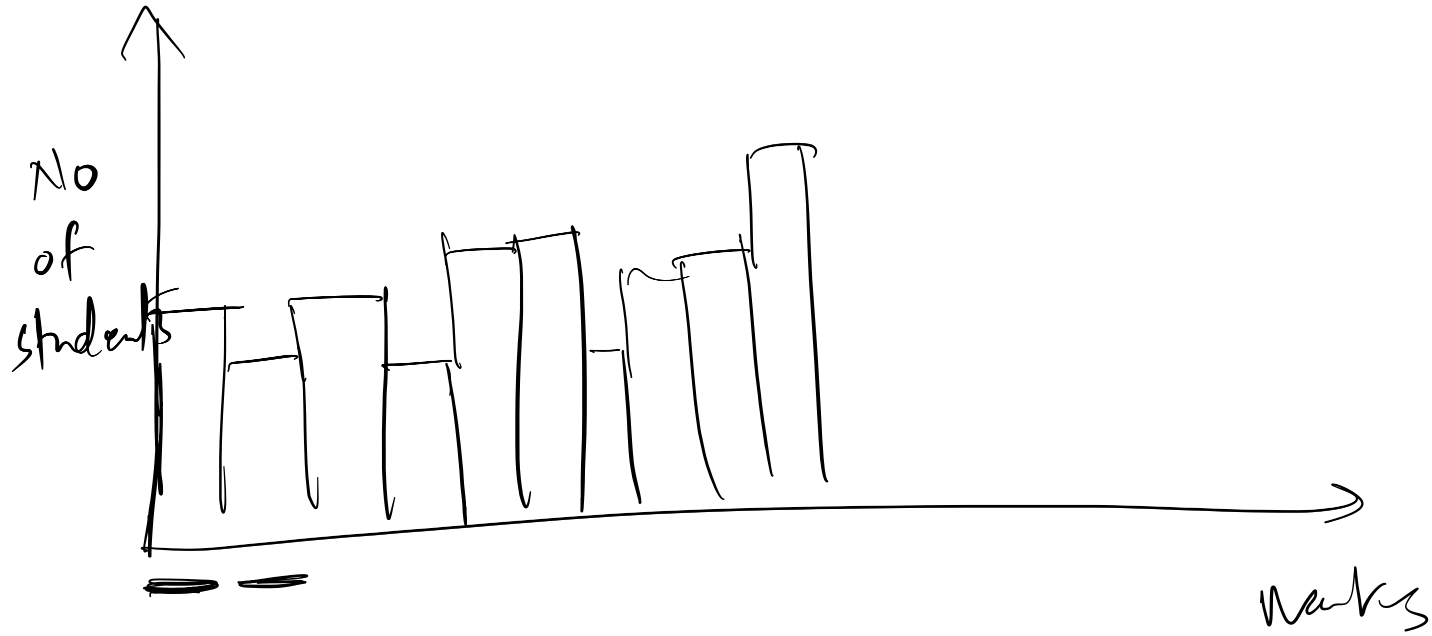
IC252 : Probability and Statistics

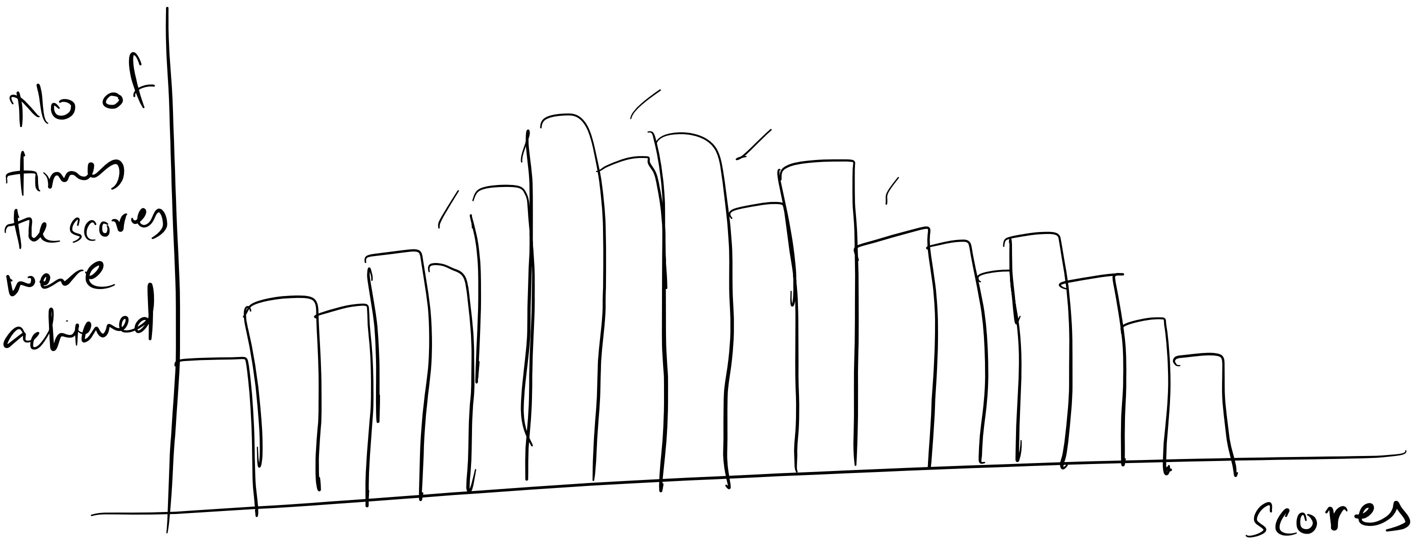
Data analysis → Statistics (Due to inherent variation) in the data

→ uncertainity of decisions

→ The need for 'enough' data to make decision under uncertainity

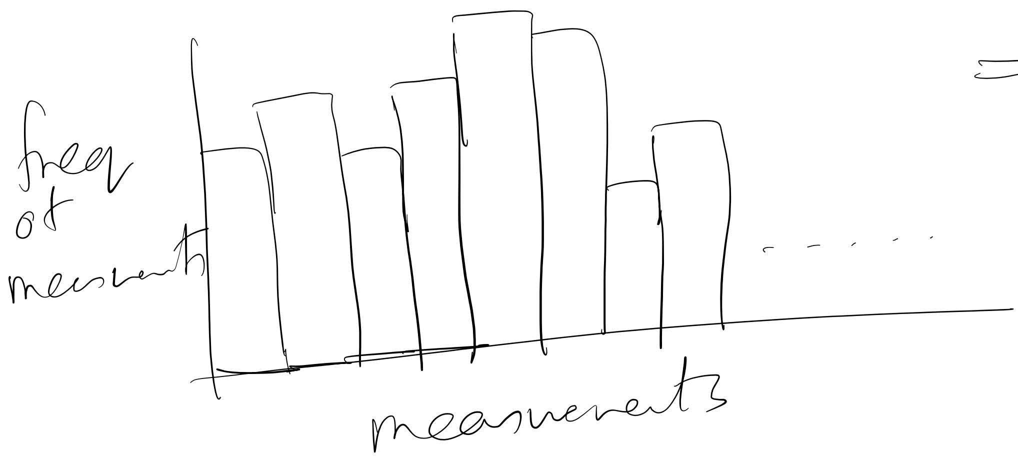
— formal mechanism to quantify the decision making under uncertainity





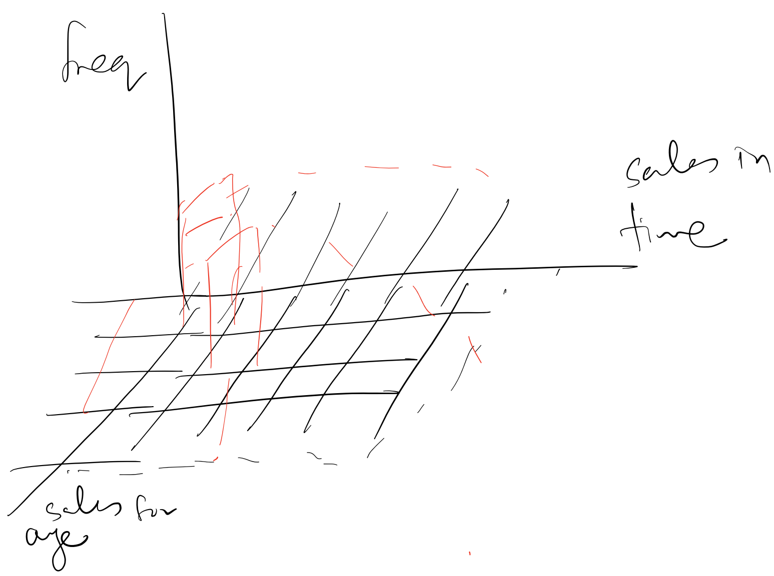
Examples of histograms

Descriptive statistics



$\Rightarrow$ 1D Histogram

Sample: a



Sample: $\begin{bmatrix} a \\ b \end{bmatrix}$

$\rightarrow$ lead to nd histogram

Sample: $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$

List of common descriptive statistics

Sample
- mean
- variance / std dev
- median
- mode
- percentile (quantile)

For DS based samples, we do not use distributions

⇓

we use only samples

I

mean: $\bar{u} = \frac{1}{N} \sum_{i=1}^{N} \bar{x}_i$ $\quad x_i$: i^{th} sample

$x_i \in \underline{\underline{\mathbb{R}^n}}$

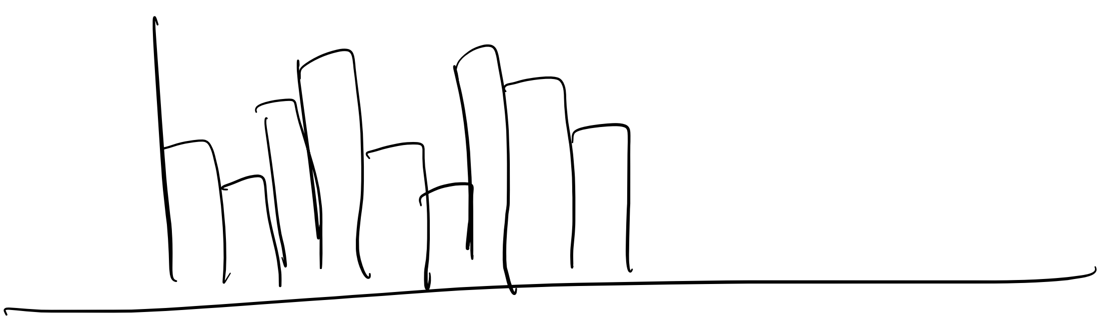
(first order statistics)

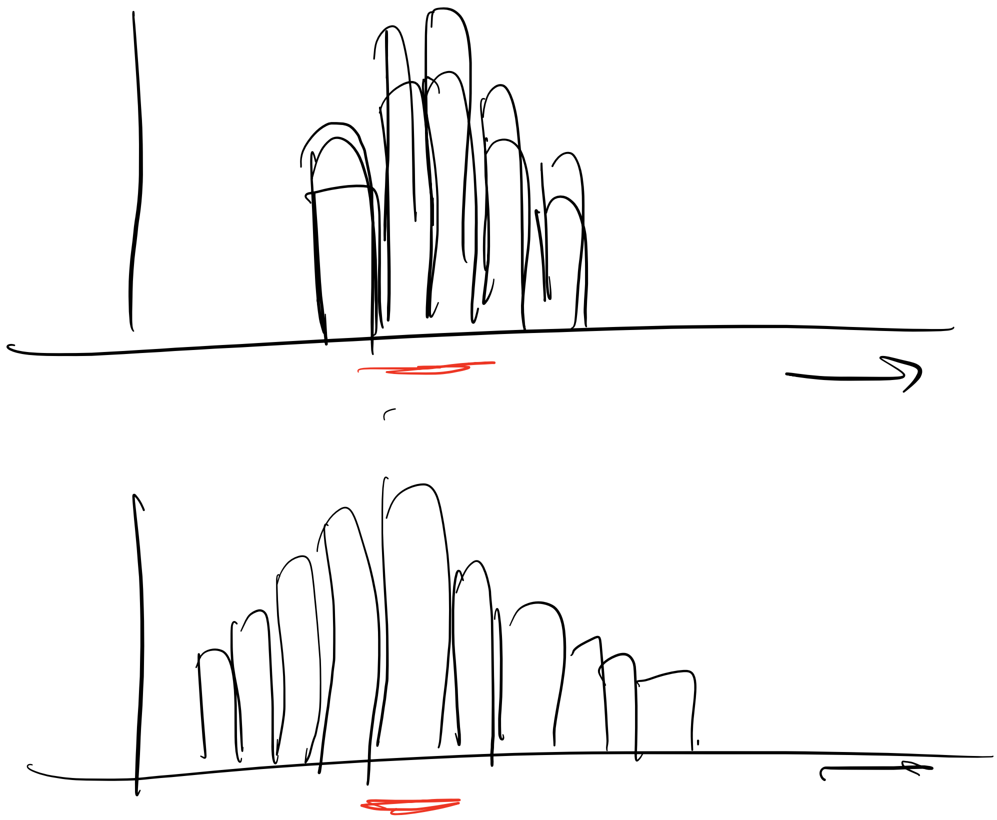
II

Variance: $\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - u)^2$

(second order statistics) $\rightarrow$ quadratic <u>form</u>

Avg of deviation of each value from the mean





examples of same mean but different variance

III median

1 2 3 4 5

$\mu = 3$

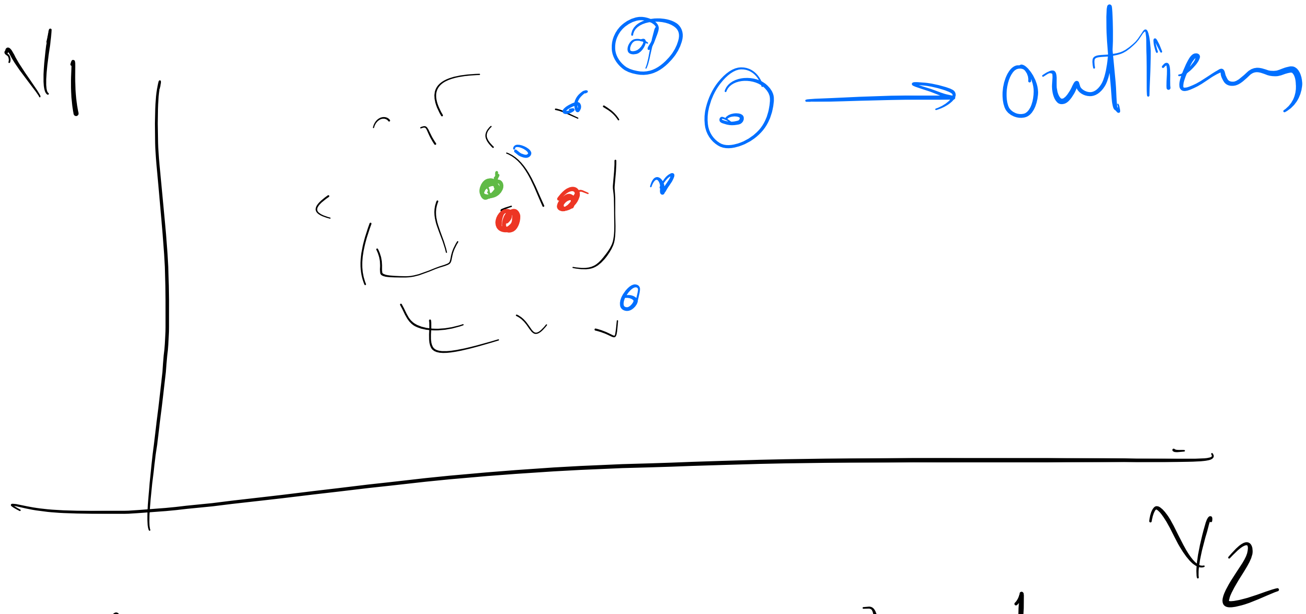
med $= 3$

1 2 3 4 10

$\mu = 4$

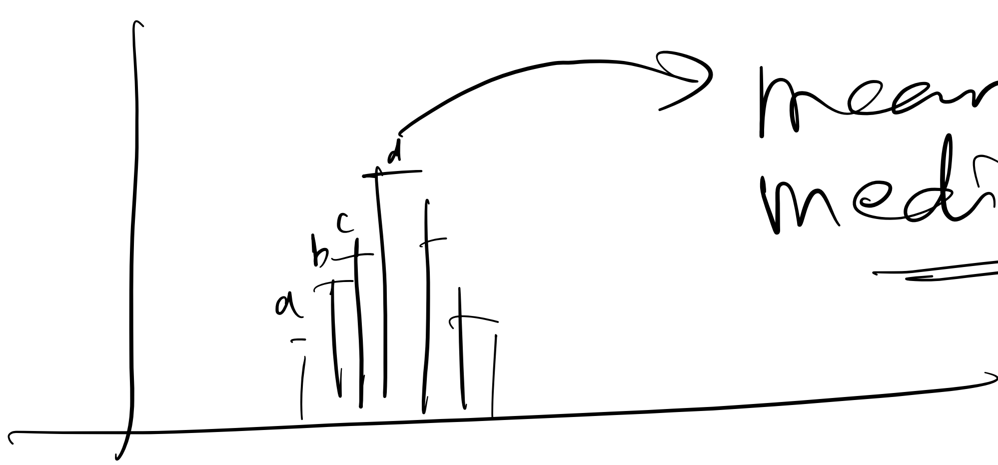
med $= 3$

median
- sort the sample values
- pick the center value

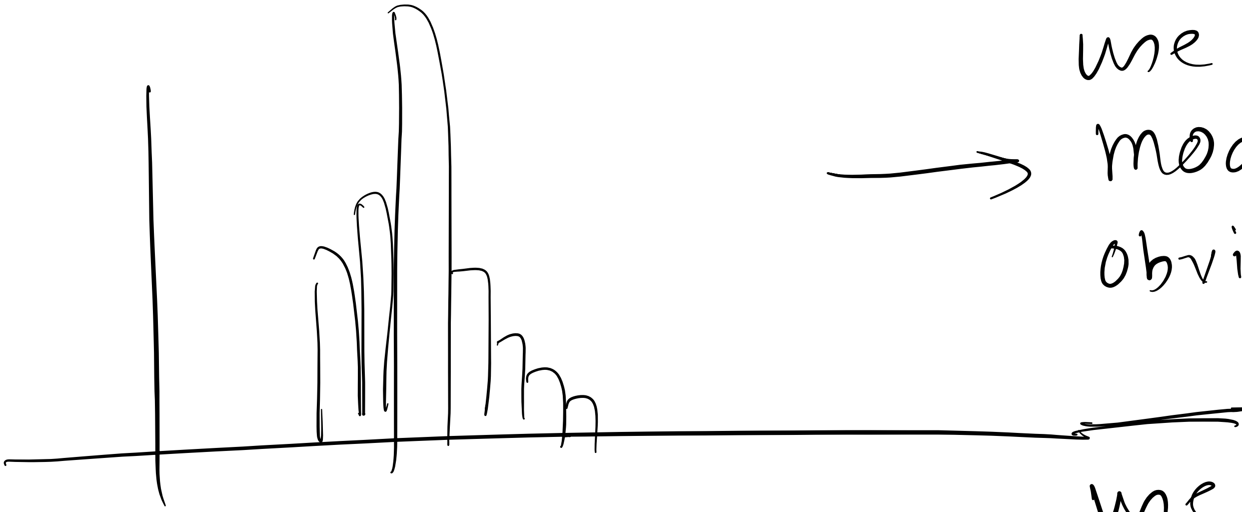


median is more robust to (minority) outliers

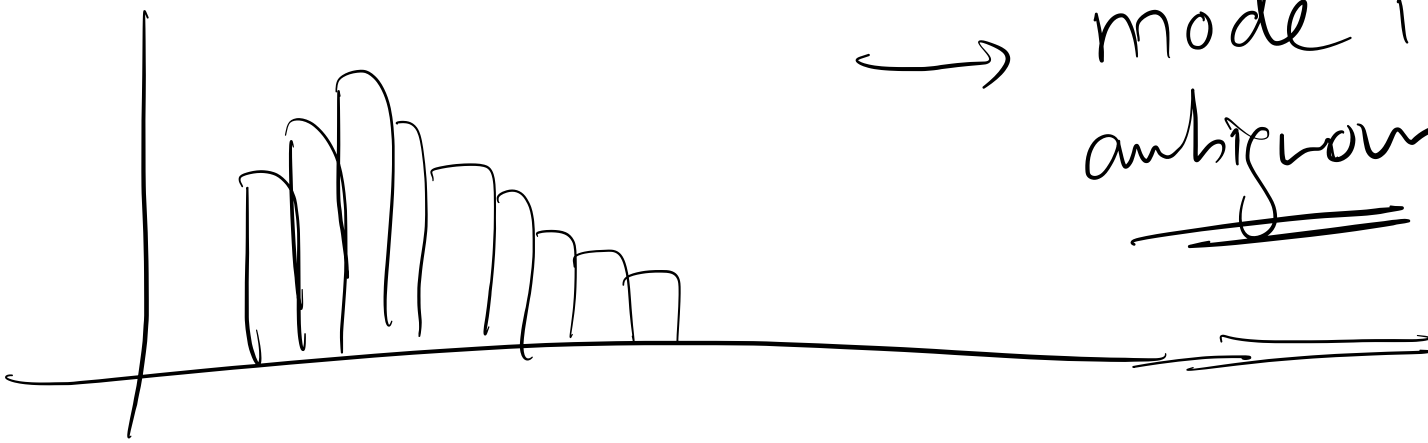
V mode: Most frequent value



mean, mode
median



use of mode is obvious



use of mode is ambiguous

percentiles

p^{th} percentile: Is that of p for which there are Np values lower than that value of p

where N: total no. of samples

medium → 0.5 percentile (50th)

quantile → 0.25 percentile (25th)

0.75 " (75th)

Interpretation of probability

I Frequency interpretation

Estimates of probability or descriptive statistics is completely dependent of sampling

II Classical interpretation

Estimates the probabilities based on concepts of likelihood, based on experience/knowledge, w/o doing an experiment

↳ An overarching mathematical framework

- transformation of variables (e.g. deriving attributes)
- Extracting meaningful quantities
- Defining computational methods (e.g. statistical tests, ML algorithms etc)

Probability theory

- Sample space: A set consisting of all possible <u>outcomes</u> for an experiment

eg. Roll of dice: $S = \{1, 2, \ldots 6\}$ → integer-valued scalar

weather data: $S = \begin{bmatrix} (a_1, b_1) \\ (a_2, b_2) \\ \vdots \\ (a_n, b_n) \end{bmatrix}$ → real-valued vectors

- Event: A subset of sample space.

Events can be represented as sets $A, B, C \ldots$

⇒ This enables us to define operations on the sets. (or events)

eg. $C = A \cup B$: Event C which is made up of outcomes of either A or B or both

$C = A \cap B$: Event C ___ outcomes from both A and B

$C = A^c$: Set of outcomes in the sample space S which are not in A

$A \subset B$: Set of outcomes in A which are also in B

The need to assign some numbers to the concepts of Sample space, outcomes, events.

denote 'probability' which is a scaling of the 'frequency' in the frequency interpretation:

$$\frac{\text{frequency of the outcome in an event}}{\text{total no. of outcomes}} \in [0,1]$$

Axioms of 'probability'

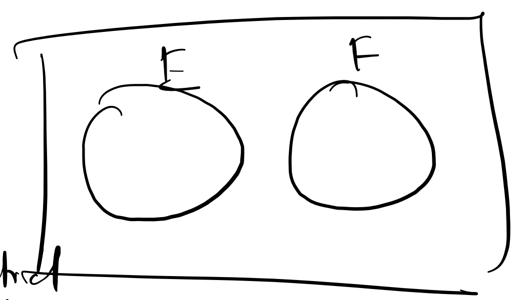
a) $0 \leq P(E) \leq 1$

b) $P(S) = 1$

c) If two events E and F are "mutually exclusive"

then

$$P(E \cup F) = P(E) + P(F)$$



assigning a probability to a relation of events | assigning a mathematical identity of '+' to compute the probability $P(E \cup F)$

More properties

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

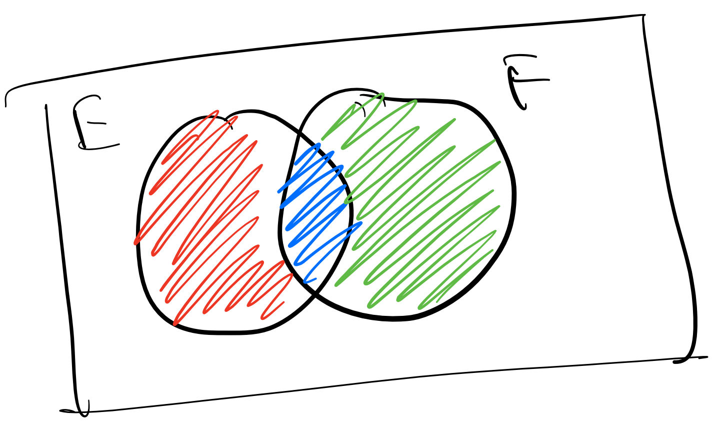
$S = E \cup E^c$

$P(S) = P(E) + P(E^c)$

$1 = P(E) + P(E^c)$

$\Rightarrow P(E^c) = 1 - P(E)$

2) $P(E \cup F) = P(E) + P(F) - P(E \cap F)$



$\Downarrow$

$P(E \cap F)$ also written as $P(EF)$

$$E \cup F = \underbrace{(E - EF)} \cup (EF) \cup \underline{(F - EF)}$$

$$P(E \cup F) = \underbrace{P(E - EF)} + P(EF) + \underline{P(F - EF)}$$

$$E = (E - EF) \cup (EF)$$

$$P(E) = P(E - EF) + P(EF)$$

$$\Rightarrow \underbrace{P(E - EF)} = P(E) - P(EF)$$

$$P(E \cup F) = P(E) - P(EF) + P(EF) + P(F) - P(EF)$$

$$\underline{P(E \cup F) = P(E) + P(F) - P(EF)}$$

$\underline{P(\emptyset) = 0}$ (Prove) HW

Conditional probability

$P(E/F)$:- probability of an event E given that an event F has happened

event is given

⇓

— probability estimates of E are affected based on additional knowledge about a related event F

$$\boxed{P(E/F) = \frac{P(EF)}{P(F)}}$$

→ Removing the effect of P(F) from P(EF) because of the fact that F is given