# Types of Analysis

1. Univariate /
2. Bivariate

③ Multivariate

1 Univariate

one variable

- data derive
- define

- summarise

Bar chart

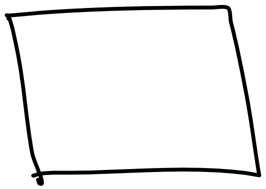
Histogram

Pie chart

CDF

PDF

# Probability Density Function (PDF)

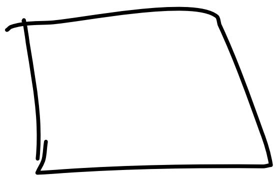


Dice

$\rightarrow$ 1, 2, 3, 4, 5, 6

$$\frac{1}{6} \qquad \frac{2}{6}$$

# Cumulative Density function (CDf)

 $\longrightarrow$ 1 2 3 6

$\frac{1}{6}, \frac{2}{6}, \frac{3}{6}, \ldots\ldots \frac{6}{6}$

$\downarrow$

1

## 2 Bivariate Analysis

Bi → 2

1 Numerical-Numerical

- Scatter plot
- Linear correlation

2) Categorical - Categorical

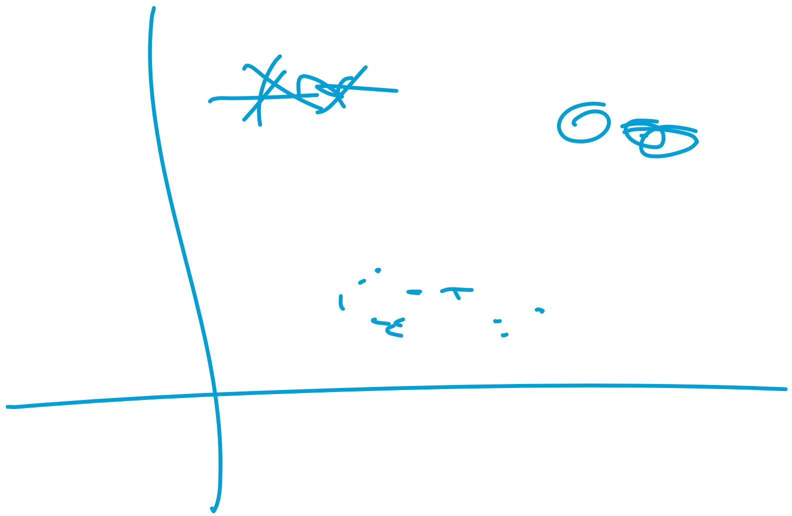
- Chi-Square

3) Numer - Cat

- Z-test, t-test
- ANOVA

# 3 Multivariate

- Cluster Analysis
- PCA
- Correspo



# Naive Bayes Classifier

## Bayes Theorem / Conditional Probability

1. Marginal probability → $P(A)$
2. Joint " → $P(A, B)$
3. Conditional " → $P(A \mid B)$

# Conditional

$$\begin{cases} P(A,B) = P(B,A) \\ P(A|B) \neq P(B|A) \end{cases}$$

$$P(A,B) = P(A|B) \times P(B) \text{ — } \textcircled{1}$$

$$= P(B|A) \times P(A) \text{ — } \textcircled{2}$$

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

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

Bayes Theorem /

Reverend Thomas Bayes

# Bayes Theorem

$P(A) = 0.7$

$P(\bar{A}) = 1 - 0.7$
$= 0.3$

$1 - P(A)$

$$\boxed{P(B) = P(B|A) \cdot P(A) + P(B|\bar{A})\, P(\bar{A})}$$

Sunday, May 28, 2023 10:16 PM

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B|A)\,P(A) + P(B|\bar{A})\,P(\bar{A})}$$

$P(A) \longrightarrow$ Prior probability

$P(A|B) \longrightarrow$ Posterior ,,

$P(B) \longrightarrow$ Evidence

$P(B|A) \longrightarrow$ Likelihood

Posterior = like × Prior

Evidence

* Population → Cancer

  Test → Posv / Negv

**Problem**

→ positive (Test)

→ Probability (Cancer)

Sensitivity → True Positive rate

↓

85%

Base rate fallacy

$$\boxed{P(T=+ \mid Can=true) = 0.85}$$

85% মানুষের → patient cancer

wrong

Base rate → 0.0002 P

5000 patient → Low

P (cancer = True) = 0.02%

$$P(C=t \mid T=+) = \frac{P(T=+ \mid C=t)\,P(C=t)}{P(T=+)}$$

$$= \frac{0.85 \times 0.0002}{P(T=+)}$$

$P(T=+) \rightarrow 0.05016$

P.T.O

$$= 0.0033$$

$\therefore$ $0.33\% \rightarrow$ patient (cancer)

$$P(T=+) = P(+|t)\,P(t) + P(+|f)\,P(f)$$

$$P(f) = 1 - 0.0002 = 0.9998$$

$$P(+|f) = 1 - 0.95 = 0.05$$

$$= 0.85 \times 0.0002 + 0.05 \times 0.9998$$

$$= 0.00017 + 0.04999$$

$$P(T=+) = 0.05016$$

Specificity $\longrightarrow$ True Neg rate

$\downarrow$

95% 0.95

```
def bayes_theorem(   )
{

        }
```

- Modeling Hypotheses

- Classification

# Naïve Bayes Classifier

$$P(C \mid F1, F2, F3 \ldots\ldots F_n)$$

$$= \frac{P(C) \times (P(F1 \mid C)\ P(F2 \mid C) \ldots P(F_n \mid C))}{P(F)}$$

# * Bayes optimal Classifier