Basics of Probability for Machine Learning

Debaditya Roy

Why Probability in ML?

ML = decision making under uncertainty

Probabilistic models describe uncertainty over

- Observations
- Parameters
- Predictions

Core to Bayesian methods, graphical models, generative models

Random Variables

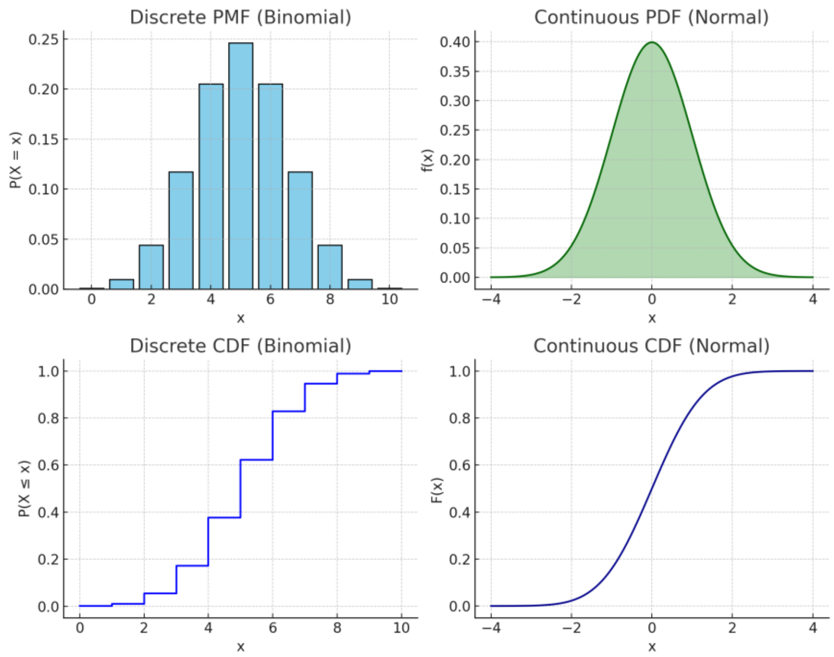
- Discrete: e.g., $X \in \{0, 1\}$
- Continuous: e.g., $X \in \mathbb{R}$
- Notation:

  - Capital letter = Random Variable : X

  - Lowercase = value: x

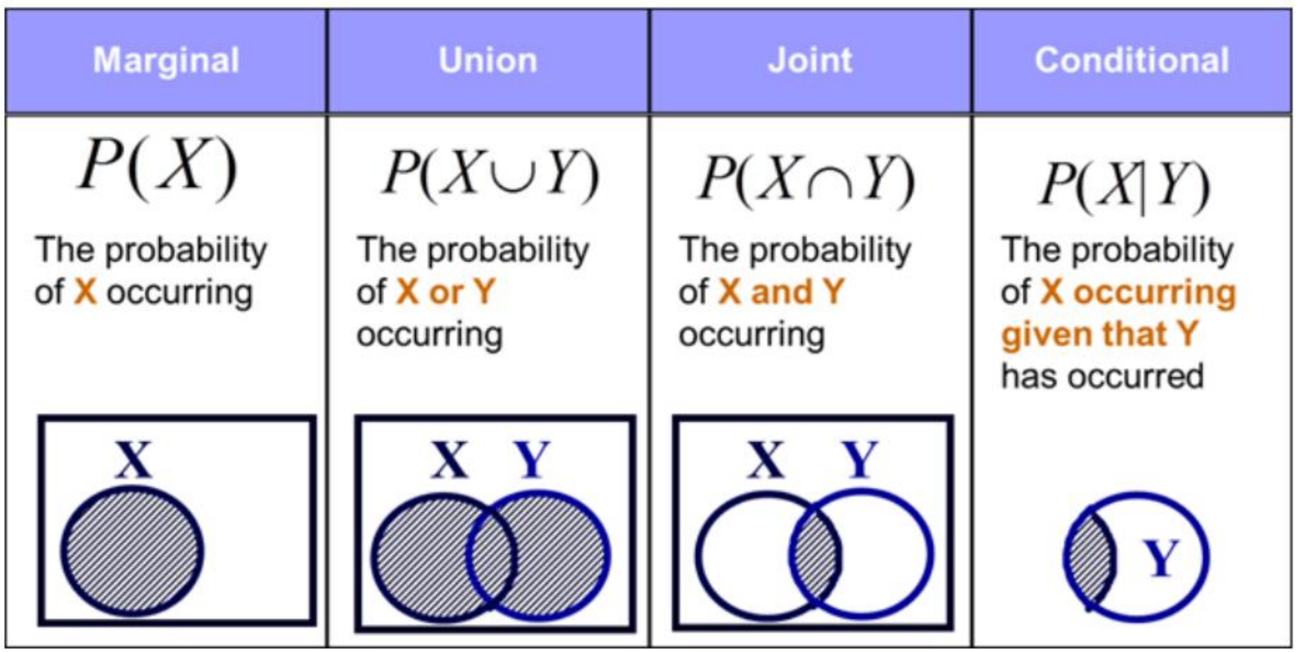
Distribution Definitions

- Probability Mass Function (PMF): $P(X = x)$
- Probability Density Function (PDF): $p(x)$
- Cumulative Density Function (CDF): $F(x) = P(X \leq x)$



Joint, Marginal, Conditional

- Joint: $P(X, Y)$
- Marginal: $P(X) = \Sigma_Y P(X, Y)$
- Conditional: $P(X|Y) = \frac{P(X,Y)}{P(Y)}$



https://slideplayer.com/slide/1660805/

Bayes' Rule

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

Used for:

- Inference
- Updating beliefs

Example:

1% of the population has a disease:P(Disease)=0.01, P(No Disease)=0.99

The test is:

99% sensitive (true positive rate):$P(Positive \mid Disease) = 0.99$

5% false positive rate:$P(Positive \mid No\ Disease) = 0.05$

Bayes' Rule

If someone **tests positive**, what's the chance they actually have the disease?

$$P(\text{Disease}|\text{Positive}) = \frac{P(\text{Positive}|\text{Disease})P(\text{Disease})}{P(\text{Positive})}$$

Using law of total probability

$$P(+\text{v}e) = P(+\text{ve}|\text{D})P(\text{D}) + P(+\text{ve}|{\sim}D)P({\sim}D)$$
$$= 0.99 \cdot 0.01 + 0.05 \cdot 0.99 = 0.0099 + 0.0495 = 0.0594$$

$$P(\text{D}|\text{Positive}) = \frac{0.99 \cdot 0.01}{0.0594} \approx 0.1667$$

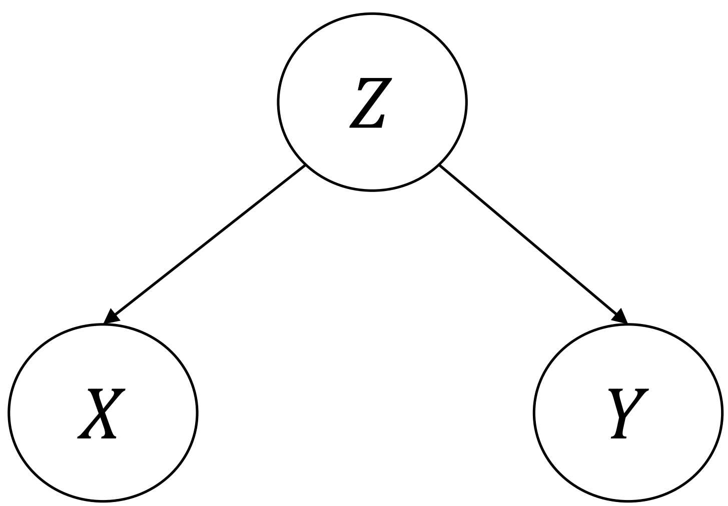
Independence

- Marginal independence

$$P(X, Y) = P(X)P(Y)$$

- Conditional independence

$$P(X \mid Y, Z) = P(X \mid Z)$$



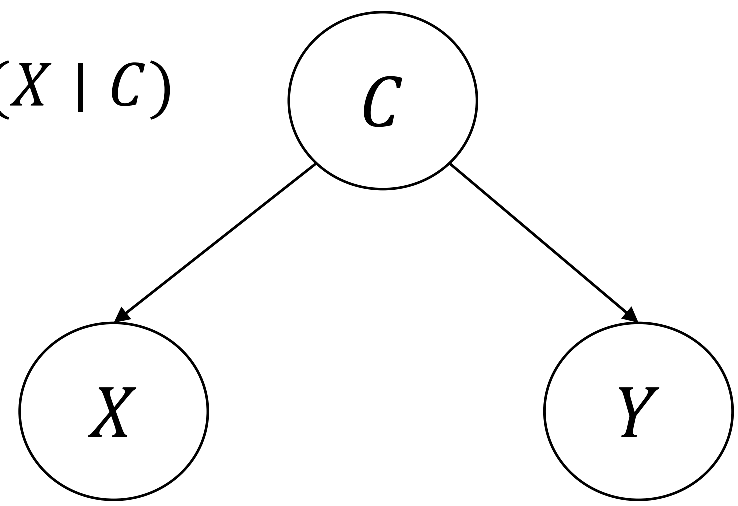
Independence

Example:

- Imagine two people (X and Y) both sneezing a lot.
- You notice this and think their sneezing is somehow connected.
- But then you find out they both had exposure to a pollen-heavy area (C)
- Once you know that, their sneezing is no longer surprising or connected—they're both reacting to the same cause.

$$P(X \mid Y, C) = P(X \mid C)$$

$$X \perp Y \mid C$$



Expectation & Variance

Expectation

$$\mathbb{E}[X] = \Sigma_x \, xP(x)$$

Variance

$$Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

Linearity

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$$

Common Distributions

Distribution	Support	Parameters	Mean	Notes
Bernoulli	$x \in \{0,1\}$	$p \in [0,1]$	p	Binary outcome (success/failure)
Binomial	$x \in \{0,1,\dots,n\}$	$n \in \mathbb{N}, p \in [0,1]$	np	Sum of n independent Bernoulli trials
Multinomial	$x_i \in \{0,\dots,n\}$ $\sum x_i = n$	$n, \boldsymbol{p} = (p_1,\dots,p_k)$	np_i per class i	Extension of binomial to multiple categories
Gaussian (Normal)	$x \in \mathbb{R}$	$\mu \in \mathbb{R}, \sigma^2 > 0$	μ	Bell-shaped continuous distribution
Exponential	$x \in [0,\infty)$	$\lambda > 0$	$\frac{1}{\lambda}$	Models time until next event (Poisson process)

Probability in ML

- Naive Bayes: uses conditional independence
- Latent Variable Models: marginals and posteriors
- EM algorithm: estimates unobserved variables
- Bayesian Inference: $P(\theta \mid D)$

Summary

Key ideas:

- Joint, marginal, conditional
- Bayes' rule
- Independence