Statistics Cheat Sheet- 4

Probability Basics

Experiment: Any action or process with uncertain outcomes (e.g., flipping a coin).

Sample Space (S): Set of all possible outcomes (e.g., S={Heads, Tails}).

Event (E): A subset of the sample space (e.g., getting a Heads).

Trial: Any particular performance of a random experiment.

Probability (P): Measure of the likelihood of an event ($0 \le P(E) \le 1$).

$$P(A) = \frac{Number\ of\ Favourable\ Outcome}{Total\ Number\ of\ Favourable\ Outcomes}$$

(e.g., P(Head) = P(Tail) = 1/2

Equally Likely Events

Events that have the same probability of occurring.

(e.g., **In a random toss of an unbiased or uniform coin, head and tail are equally likely events.)**

Independent Events

Two events (A and B) are independent if the occurrence of one does not affect the probability of the other.

(e.g., **In tossing an unbiased coin, the event of getting a head in the first toss is independent of getting a head in the second, third and subsequent throws.**

Random Variables

A random variable (X) assigns numerical values to the outcomes of an experiment.

Discrete Random Variable: Takes countable values (e.g., number of heads).

Continuous Random Variable: Takes infinite values (e.g., height, weight).

Probability Mass Function (PMF)

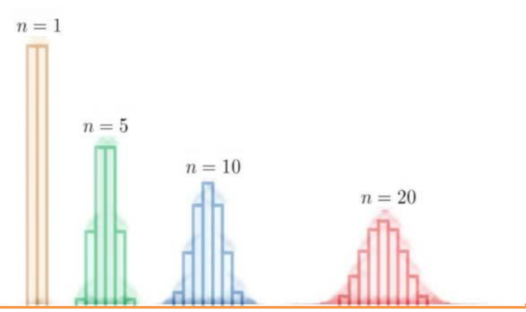
- Used for discrete random variables.
- Formula:
  P(X=x)=Probability of X taking value x.
- Example: In a dice roll, P(X=2)= 1/6 .

Probability Density Function (PDF)

- Used for continuous random variables.
- Describes the relative likelihood of the random variable taking a specific value.
- The probability is the area under the curve between two points.
- Example: Heights of people follow a normal distribution.

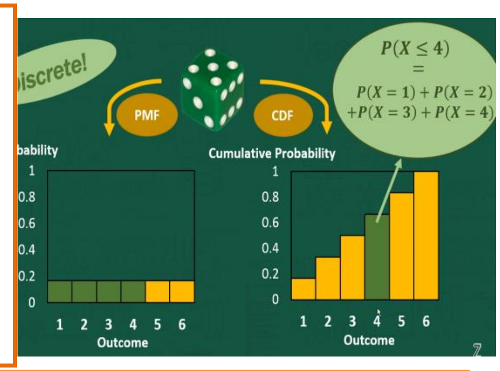
The Central Limit Theorem
- (Law of Large Numbers)

As the sample size approaches infinity, the probability distribution becomes a perfectly symmetrical(normal distribution) where the center of the curve is the mean of the population, regardless of the original distribution of the population..



Cumulative Distribution Function

The cumulative distribution function gives the probability that the random variable X is less than or equal to x.



Conditional Probability P(A|B)

Conditional probability refers to the probability of an event occurring given that another event has already occurred.

Example:

Event A: The person has the disease.

Event B: The test result is positive.

Goal: Calculate P(A|B), the probability that a person has the disease given a positive test result.

Baye's Theorem

Bayes theorem is used to determine the conditional probability of event A when event B has already occurred.

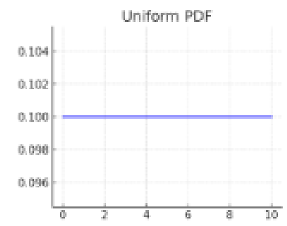
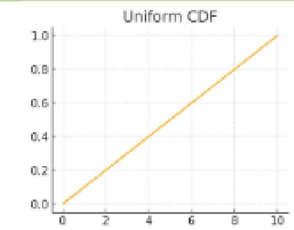
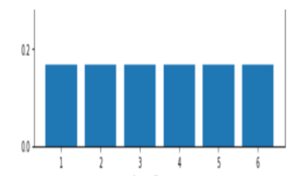
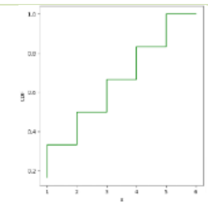
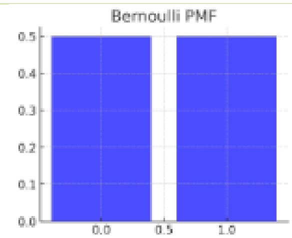
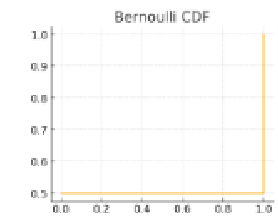
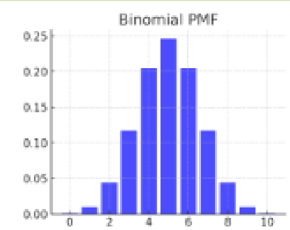
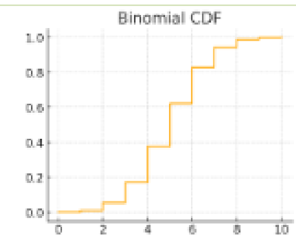
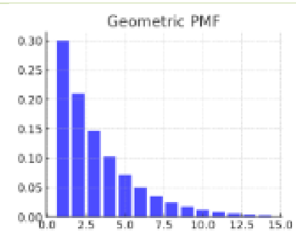
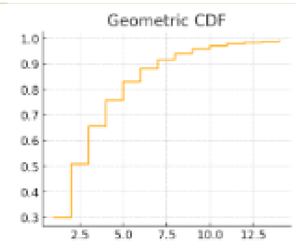
Bayesian methods are widely used in machine learning to incorporate prior knowledge and update probabilities as new information becomes available.

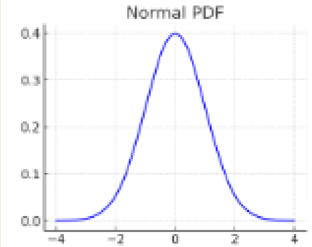
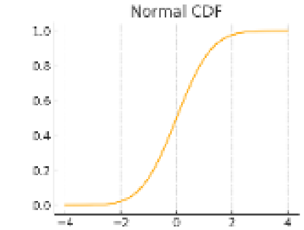
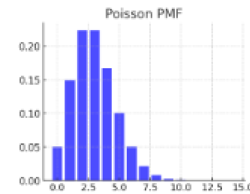
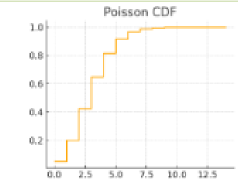
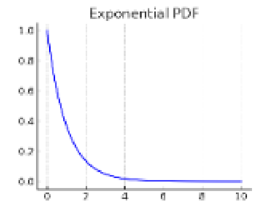
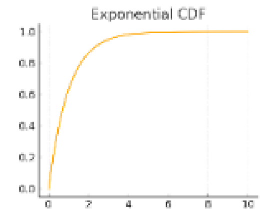
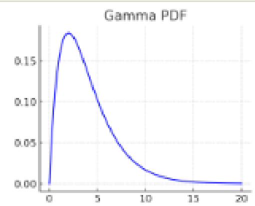
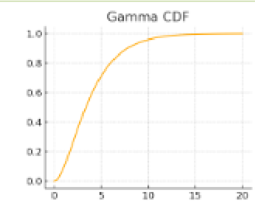
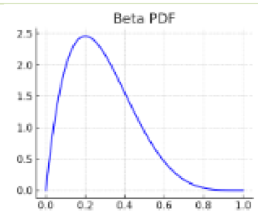
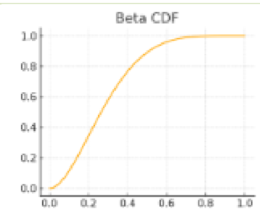
$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Bayes' Theorem Formula

where,

- **P(A)** and **P(B)** are the probabilities of events A and B also P(B) is never equal to zero.
- **P(A|B)** is the probability of event A when event B happens
- **P(B|A)** is the probability of event B when A happens

Distribution	Type	Properties	Example	Probability Function (PDF/PMF)	Parameters (mean & variance)	PDF/PMF	CDF	Applications in ML
Uniform	Continuous	when all the possible events are equally likely (Equal probability across range)	-Rolling a dice(discrete) -Picking a random point on a line segment (continuous)	$f(x)=\begin{cases}\frac{1}{b-a}, & a\le x\le b,\\ 0, & otherwise.\end{cases}$ $P(X=x)=\frac{1}{n}$ for $x=1,2,..$	Mean=$\frac{a+b}{2}$ Var=$\frac{(b-a)^2}{12}$			-Simulation -random sampling
Uniform	Discrete	when all the possible events are equally likely (Equal probability across range)	-Rolling a dice(discrete) -Picking a random point on a line segment (continuous)	$f(x)=\begin{cases}\frac{1}{b-a}, & a\le x\le b,\\ 0, & otherwise.\end{cases}$ $P(X=x)=\frac{1}{n}$ for $x=1,2,..$	Mean=$\frac{a+b}{2}$ Var=$\frac{(b-a)^2}{12}$			
Bernoulli	Discrete	Binary outcomes (has only two possible outcomes)	-Coin flip, -Success/ Failure -True/False	$f(x_i)=\begin{cases}p, & x_j=1, j=1,2,3\ldots n,\\ 1-p=q, & x_j=0, j=1,2,3..\\ 0, & otherwise.\end{cases}$	Mean = p Var = p(1-p)			-Binary classification, -feature selection -Logistic Regression
Binomial	Discrete	A Sequence of Bernoulli Events. (only 2 possible outcomes repeated n number of times-independent)	-Number of heads in coin flips -the entire quiz of 10 T/F questions	$f(x_i)=\binom{n}{x}p^x(1-p)^{n-x}$	Mean = np Var = np(1-p)			-Anomaly detection, -queue modeling, -NLP (word counts)
Geometric	Discrete	Exponential decay. Trials until first success	- Number of coin flips to get heads -the first occurrence of a failure or the first success	$f(x_i)=(1-p)^{x-1}p,$ $x=0,1,2,3,\ldots$	Mean =1/p Var=$\frac{1-p}{p^2}$			Predicting the number of attempts needed for success in recommendation systems

Normal (Gaussian)	Continuous	-Symmetric, -bell-shaped -most values clustering around a central region. -area under the curve equals 1	-Heights, -weights -exam scores	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$	$Mean = \mu$ $Var = \sigma^2$			Feature scaling (z-score), kernel methods, Gaussian Mixture Models, Linear Regression
Standard Normal	Continuous	-special case of the Normal distribution	-Z-scores, test statistics	$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	$Mean = 0$ $Var = 1$			
Poisson	Discrete	-deals with the frequency with which an event occurs within a specific interval	- Number of arrivals at a service center -The number of thefts reported in an area in a day	$f(x_i) = \frac{e^{-\lambda}\lambda^x}{x!},$ $x = 0,1,2,3, \ldots$	Mean=λ Var=λ			Anomaly detection, queue modeling, NLP (word counts)
Exponential	Continuous	-models the interval of time between the calls	-Time between arrivals at service centers -The life of an air conditioner	$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$	Mean=$1/\lambda$ Var = $\frac{1}{\lambda^2}$			Survival analysis, queueing models, reliability engineering
Gamma	Continuous	-Generalized exponential - model the time until an event occurs multiple times	- Lifetimes of multiple events - the total rainfall accumulated over a season	$f(x) = \begin{cases} \frac{\lambda}{\Gamma(r)}(\lambda x)^{r-1}e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0 \end{cases}$	Mean=k/λ Var = $\frac{k}{\lambda^2}$			Bayesian inference, survival analysis, reliability modeling
Beta	Continuous	applied in scenarios where proportions, or rates need to be estimated based on prior knowledge or observed data	-Probabilities in Bayesian models	$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$	$Mean = \frac{\alpha}{\alpha+\beta}$			A/B testing, prior distributions in Bayesian analysis