Data Science & Statistics Tutorial – 1

M.M.S CHANDRA NAGU

2320030206

1. State any 2 examples for addition, multiplication, Bayes theorem rule of probability along with its implementation?

ANS)

* Addition Rule of Probability:  
  Example 1: Probability of drawing a red card or a king from a deck of cards.  
  Example 2: Probability of rain or snow tomorrow.  
  Implementation: P(A ∪ B) = P(A) + P(B) - P(A ∩ B)
* Multiplication Rule of Probability:  
  Example 1: Probability of drawing two aces consecutively from a deck without replacement.  
  Example 2: Probability of rolling a 3 and then a 5 on two dice.  
  Implementation: P(A ∩ B) = P(A) \* P(B|A) for dependent events or P(A) \* P(B) for independent events.
* Bayes Theorem Rule of Probability:  
  Example 1: Probability of having a disease given a positive test result.  
  Example 2: Probability a randomly selected student is female given they passed an exam.  
  Implementation: P(A|B) = [P(B|A) \* P(A)] / P(B)

1. Provide a real-life example for mathematical expectation of a discrete random variable.

ANS)

Mathematical expectation (also called expected value) represents the *long-run average outcome* of a random variable over many repetitions of the same experiment.

Example:  
Imagine a small lottery game:

* You win $100 with probability 0.1 (10% chance)
* You win $50 with probability 0.2 (20% chance)
* You win nothing ($0) with probability 0.7 (70% chance)

To find the expected winnings, we multiply each possible outcome by its probability and sum them:

E(X)=(100×0.1)+(50×0.2)+(0×0.7)E(X)=10+10+0=20*E*(*X*)=(100×0.1)+(50×0.2)+(0×0.7)*E*(*X*)=10+10+0=20

So, the *expected* winning per game is $20.  
This doesn't mean you’ll win $20 every single time — rather, if you play this game a large number of times, your average winning per play will approach $20.

Real-life applications: This method is used in insurance (to estimate premium rates), in business (to decide whether a project is worth investing in), and in gambling games to calculate fair payouts.

1. Mention an example for mathematical expectation of a continuous random variable.

ANS)

For continuous random variables, the mathematical expectation is calculated using integration instead of summation.

Example: Lifetime of a light bulb  
Suppose the lifetime X*X* of a certain type of light bulb follows an exponential distribution with an average (mean) lifetime of 5 years.

The probability density function (PDF) of an exponential distribution is:

f(x)=λe−λxforx≥0*f*(*x*)=*λe*−*λx*for*x*≥0

Here, λ=1mean=15=0.2*λ*=mean1=51=0.2.

The mathematical expectation for a continuous random variable is:

E(X)=∫0∞x f(x) dx*E*(*X*)=∫0∞*xf*(*x*)*dx*

Substituting f(x)*f*(*x*):

E(X)=∫0∞x⋅0.2e−0.2x dx*E*(*X*)=∫0∞*x*⋅0.2*e*−0.2*xdx*

By solving the integration (using integration by parts), we get:

E(X)=1λ=5 years*E*(*X*)=*λ*1=5 years

Interpretation: On average, each light bulb will last 5 years before it burns out. While some might fail earlier and some might last longer, in the long run the *average lifetime* remains at 5 years.

1. List a situation where binomial distribution, Poisson distribution, geometric distribution is used?

ANS)

These three probability distributions model different types of real-world random processes. Understanding when to apply each can help in analyzing and predicting outcomes effectively.

* Binomial Distribution:

The binomial distribution applies when you have a fixed number of independent trials, each with two possible outcomes (success or failure), and the probability of success is constant across trials.

Example:  
Suppose you toss a fair coin 10 times and want to find the probability of getting exactly 6 heads. Here:

* Each toss is independent
* Each toss has two possible outcomes (head or tail)
* Probability of success (head) remains 0.5 for each toss

The binomial distribution calculates the probability of exactly *k* successes in *n* trials using:

P(X=k)=(nk)pk(1−p)n−k*P*(*X*=*k*)=(*kn*)*pk*(1−*p*)*n*−*k*

Real-life uses:

* Quality control in manufacturing (e.g., number of defective items in a batch)
* Medical trials (e.g., number of patients responding to a treatment in a fixed group)
* Sports (e.g., number of goals scored by a player in a fixed number of attempts)
* Poisson Distribution:

The Poisson distribution models the number of times an event occurs within a fixed interval of time or space, assuming these events happen independently and at a constant average rate.

Example:  
Imagine a call center receives on average 5 calls per hour. What is the probability of receiving exactly 8 calls in the next hour?

Poisson distribution formula:

P(X=k)=λke−λk!*P*(*X*=*k*)=*k*!*λke*−*λ*

where λ*λ* is the average number of events in the interval.

Real-life uses:

* Number of emails received per hour
* Number of accidents occurring at a traffic junction in a day
* Number of decay events from a radioactive source in a minute
* Geometric Distribution:

The geometric distribution models the number of independent trials until the first success occurs in a sequence of Bernoulli trials (each with success probability p*p*).

Example:  
You roll a fair six-sided die repeatedly until you get the first “6.” The number of rolls needed follows a geometric distribution with success probability p=16*p*=61.

Geometric distribution formula for the probability that the first success occurs on the k*k*-th trial:

P(X=k)=(1−p)k−1p*P*(*X*=*k*)=(1−*p*)*k*−1*p*

Real-life uses:

* Number of coin flips until the first head appears
* Number of sales calls until the first sale is made
* Number of shots taken until scoring the first goal in a basketball game.