ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

Probability & Statistics: Distributions, expectation, variance, covariance, Bayes theorem

Session – 16 & 17

AIM OF THE SESSION

To familiarize students with the basic concepts of Probability & Statistics

INSTRUCTIONAL OBJECTIVES

This Session is designed to:

1. Explain Expectation, Variance, and Co-Variance.
2. Describe different types of Discrete and Continuous distributions.

LEARNING OUTCOMES

At the end of this session, you should be able to:

1. Define the concepts of Expectation, Variance, and covariance.
2. Understand different Distributions & Bayes theorem.

SESSION INTRODUCTION

A probability distribution describes how the values of a random variable are distributed.

Two main types:

- Discrete Distributions
- Continuous Distributions

Discrete Distributions

Defined for countable outcomes

Types:

1. General Discrete Distribution
2. Bernoulli Distribution
3. Binomial Distribution
4. Hypergeometric Distribution
5. Geometric Distribution
6. Poisson Distribution

1. General Discrete Distribution

Definition: Each outcome has equal probability

PMF: p(X = x) = 1/n

Mean = $E(x) = \sum xp(x)$

Variance = $V(x) = E(x^2) - [E(x)]^2 = \sum x^2 p(x) - [\sum xp(x)]^2$

Properties:

- $E(aX_1 + b) = aE(X_1) + b$
- $V(aX_1 + b) = a^2 V(X_1)$
- $E(aX_1 + bX_2) = aE(X_1) + bE(X_2)$
- $V(aX_1 + bX_2) = a^2 V(X_1) + b^2 V(X_2) + 2ab\, cov(X_1, X_2)$
- $cov(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2)$

Note:

If X_1 and X_2 are independent, $E(X_1 X_2) = E(X_1)E(X_2)$

$cov(X_1, X_2) = 0$

1. General Discrete Distribution

Example: Rolling a fair die (1 to 6)

P(X=1) = P(X=2) = P(X=3) = P(X=4) = P(X=5) = P(X=6) = 1/6

p(X = x) = 1/n= 1/6

Mean = $E(x) = \sum xp(x) = 1 * \frac{1}{6} + 2 * \frac{1}{6} + 3 * \frac{1}{6} + 4 * \frac{1}{6} + 5 * \frac{1}{6} + 6 * \frac{1}{6} = 3.5$

Variance = $V(x) = E(x^2) - [E(x)]^2 = \sum x^2 p(x) - [\sum xp(x)]^2 = 2.917$

2. Bernoulli Distribution

Definition: One trial with two outcomes (success/failure)

PMF: $P(X=1)= p, P(X=0)= 1-p$

Example: Tossing a coin once (Head = 1, Tail = 0)

$$P(X=Head) = p = 1/2$$
$$P(X= Tail) = 1-p = 1-1/2 = 1/2$$

3. Binomial Distribution

Definition: The Binomial Distribution occurs when an experiment is performed satisfies the three assumptions of Bernoulli trials, which are

1. Only 2 outcomes are possible: success and failure
2. Probability of success (p) and failure (1-p) remain the same from trial to trial.
3. The trials are statistically independent

PMF: $p(X=r) = n_{c_r} p^r (1-p)^{n-r}$

Mean = np

Variance = np(1-p)

Example: 10 dice are thrown. What is the probability of getting exactly 2 sixes

$p(X=2) = 10_{c_2} (1/6)^2 (1-1/6)^8$ = 0.2907

4. Hypergeometric Distribution

Definition: If the probability changes from trail to trail, one of the assumptions of binomial distribution gets violated and hence binomial distribution cannot be used. In such cases hypergeometric distribution is used. This is particularly used in cases of sampling without replacement from a finite population.

PMF: $p(X=x) = \dfrac{{}^{r}C_{x} * {}^{N-r}C_{n-x}}{{}^{N}C_{n}}$

$p(X=1) = \dfrac{{}^{6}C_{1} * {}^{4}C_{2}}{{}^{10}C_{3}} = 0.3$

N = 10

r= 6 D   N-r = 4 ND

n=3

x=1D   n-x = 2 ND

Example:

There are 10 markers on a table, of which 6 are defective and 4 are not defective. If 3 are randomly taken from above lot, what is the probability that exactly 1 of markers is defective.

5. Geometric Distribution

Definition: Consider repeated trials of a Bernoulli experiment with probability **p** of success and **q = 1-p** of failure. Let **x** denote the number of times Bernoulli experiment must be repeated until finally obtaining a success. The distribution of random variable **x** is given as follows.

The experiment will be repeated **k** times only in the case that there is a sequence of **k-1** failures followed by a success.

$$P(k) = q^{k-1}p$$

Example: Suppose the probability that team A wins each game in a tournament is 60 percent. A plays until it loses.

a). Find the expected number of games that A plays = $E(x) = \frac{1}{p} = \frac{1}{0.4} = 2.5$

b). Find the probability P that A plays in at least 4 games = $P(x > r = 3) = q^r = (0.6)^3$

6. Poission Distribution

Definition: A random variable X, taking on one of the values 0, 1, 2, …. Is said to be a Possion random variable with parameter λ if for some $\lambda > 0$,

$$P(X = x) = \frac{e^{-\lambda} \lambda^{x}}{x!}$$

For Possion distribution:

Mean = $E(x) = \lambda = \alpha \Delta t$ = Variance = $V(x)$

Example: A certain airport receives on an average of 4 aircrafts per hour. What is the probability that no aircraft lands in a particular 2 hr period.

Ans: $\alpha =$ rate of occurrence of event per unit time $= \frac{4}{hr}$

$\lambda = \alpha \Delta t$ = avg. no of occurrence of event in specified observation period = 4*2 = 8

Now we wish that no aircraft should land for 2 hrs. i.e. x = 0

$$P(x = 0) = \frac{e^{-\lambda} \lambda^{x}}{x!} = \frac{e^{-8} 8^{0}}{0!} = e^{-8}$$

Continuous Distributions

Types:

1. General Continuous Distribution
2. Uniform Distribution
3. Exponential Distribution
4. Normal Distribution
5. Standard Normal Distribution

1. General Continuous Distributions

Let X be a continuous variable. A continuous distribution of X can be defined by a probability density function f(x) which is such a function such that

- Probability density function = $f(x) = \dfrac{dF}{dx}$
- Cumulative Probability = $F(x) = p(X \le x) = \int_{-\infty}^{x} f(x)dx$
- Mean = Expected value = $E(x) = \mu_x = \int_{-\infty}^{\infty} xf(x)dx$
- Variance= $V(x) = E(x^2) - [E(x)]^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - [\int_{-\infty}^{\infty} xf(x)dx]^2$
- $p(-\infty \le X \le \infty) = \int_{-\infty}^{\infty} f(x)dx = 1$

I. General Continuous Distributions (Cont…)

Example: A continuous random variable X has a probability density function $f(x) = e^{-x},\ 0 < x < \infty$. Then $P[X > 1]$ is

Ans:

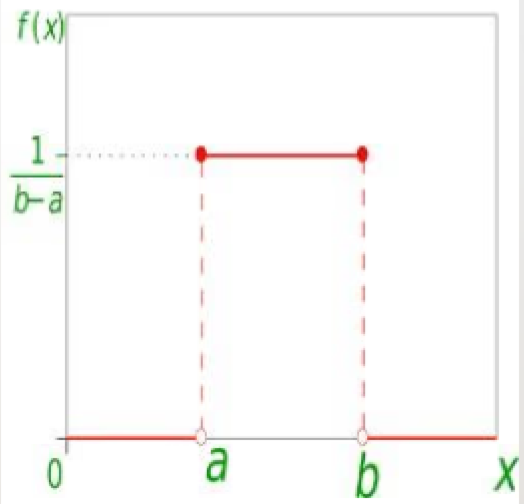
$$p(X > 1) = 1 - p(X \le 1) = 1 - \int_{-\infty}^{1} f(x)dx = 1 - \int_{-\infty}^{1} e^{-x}dx$$

$$p(X > 1) = \int_{1}^{\infty} e^{-x}dx = e^{-1} = 0.368$$

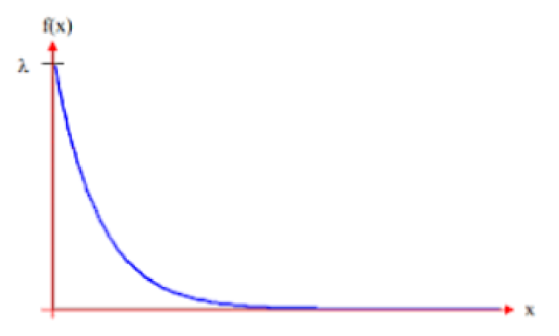
2. Uniform Distributions

Example: If X is uniformly distributed over (0,10). Calculate the probability that

- Probability density function= f(x) = $\frac{1}{b-a} = \frac{1}{10-0} = 0.1$
- Cumulative Distribution function = $F(x) = \frac{x-a}{b-a} = \frac{x-0}{10-0} = 0.1x$
- Mean = $E(X) = \frac{b+a}{2} = \frac{10+0}{2} = 5$
- Variance = $V(X) = \frac{(b-a)^2}{12} = \frac{(10-0)^2}{12} = 100 / 12$
- X<3 = $\int_0^3 f(x)\,dx = \int_0^3 0.1\,dx =$ 0.3
- X>6 = $\int_6^{10} f(x)\,dx = \int_6^{10} 0.1\,dx =$ 0.4



3. Exponential Distributions



- Probability Density function = $f(x) = \begin{cases} \lambda e^{-\lambda x}, & if\ x \geq 0 \\ 0, & if\ x < 0 \end{cases}$

- Cumulative Distribution function = $F(a) = P(X \leq a) = \int_0^a \lambda e^{-\lambda x} dx = \left(-e^{-\lambda x}\right)_0^a = 1 - e^{-\lambda a}$, a ≥ 0

- Mean = $E[X] = \frac{1}{\lambda}$
- Variance = $V(X) = \frac{1}{\lambda^2}$

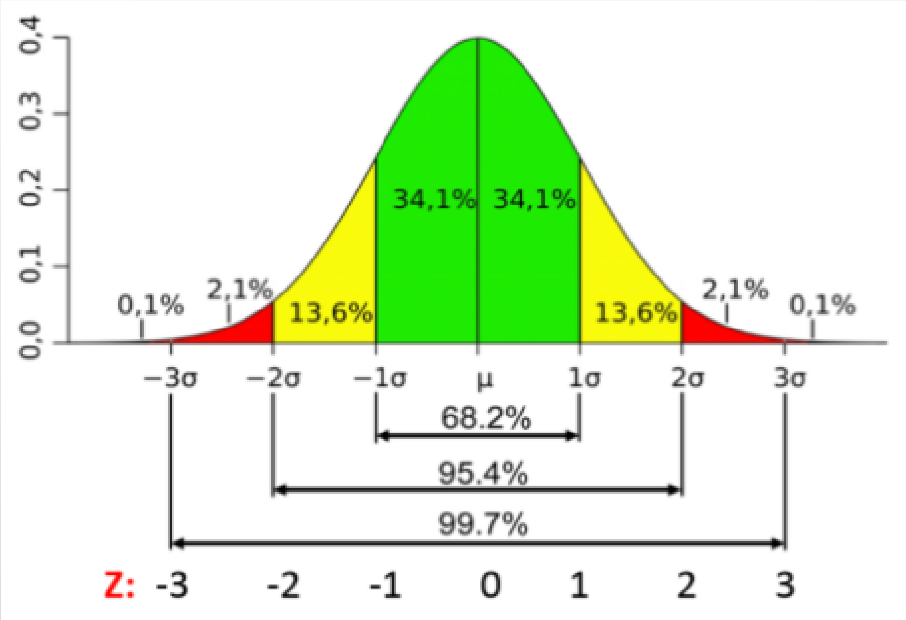
Example: Suppose that the length of a phone call in minutes is an exponential random variable with parameter $\lambda = 0.1$. If someone arrives immediately ahead of you at a public telephone booth, find the probability that you will have to wait,

(a). More than 10 minutes = $P(X > 10) = 1 - P(X \leq 10) = 1 - F(10) = 1 - (1 - e^{-\lambda 10})$=0.368

(b). Between 10 and 20 minute = $P(10 < X < 20) = F(20) - F(10) = 0.233$

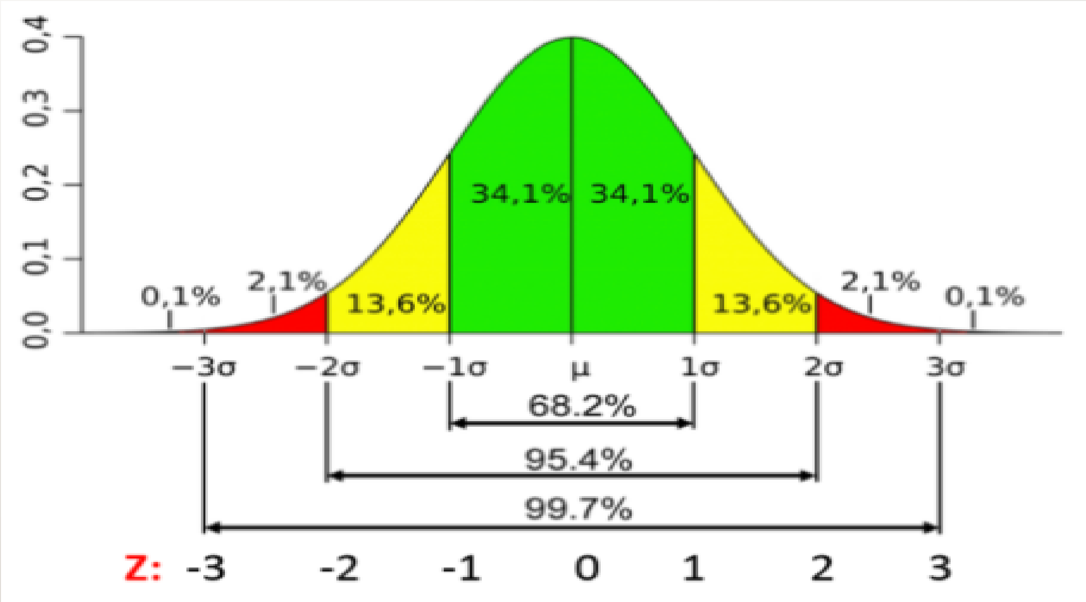
4. Normal Distributions $[N(\mu,\sigma^2)]$

- Probability Density function = $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \; -\infty < x < \infty$
- Cumulative Distribution function = $F(a) = 1 - Q(\frac{x-\mu}{\sigma})$
- Mean = $E[X] = \mu$
- Variance = $V(X) = \sigma^2$



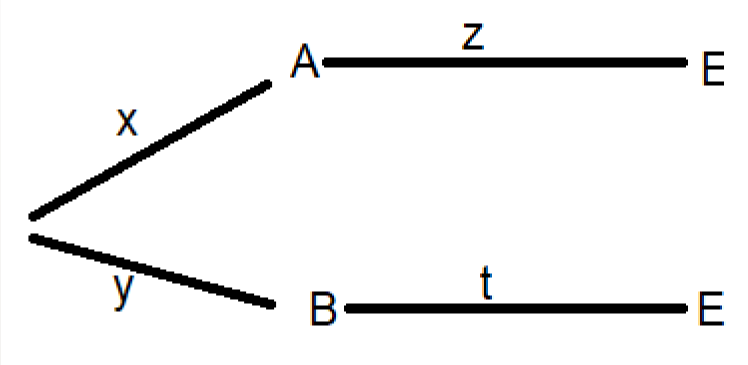
5. Standard Normal Distributions $[N(0,1)]$

- Probability Density function = $f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}, \; -\infty < x < \infty$
- Cumulative Distribution function = $F(a) = 1 - Q(x)$
- Mean = $E[X] = \mu = 0$
- Variance = $V(X) = \sigma^2 = 1$



Bayes Theorem

- $P\left(\frac{A}{E}\right) = \frac{P(A \cap E)}{P(E)} = \frac{P(A \cap E)}{P(A \cap E) + P(B \cap E)} = \frac{P(A) * P(E/A)}{P(A) * P(E/A) + P(B) * P(E/B)}$



- Then the above Bayes formula gives
- $P\left(\frac{A}{E}\right) = \frac{xz}{xz + yt}$
- $P\left(\frac{B}{E}\right) = \frac{yt}{xz + yt}$

Bayes Theorem (Cont…)

Example: Suppose we have 2 bags. Bag 1 contains 2 red & 5 green marbles. Bag 2 contains 2 red and 6 green marbles. A person tosses a coin & if it is heads goes to bag 1 and draws a marble. If it is tails, he goes to bag 2 and draws a marble. In this situation,

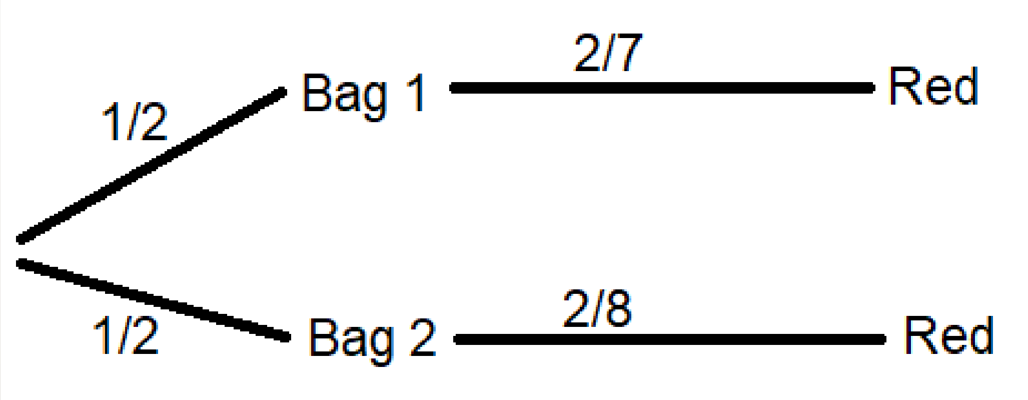
(a). What is the probability that the marble drawn this is red.

(b). Given that the marble draw is red, what is probability that it came from bag1.

(c). Given that the marble draw is red, what is probability that it came from bag2.

Ans:

- $P(Red) = \frac{1}{2} * \frac{2}{7} + \frac{1}{2} * \frac{2}{8}$
- $P\left(\frac{Bag1}{Red}\right) = \frac{xz}{xz+yt} = \frac{\frac{1}{2} * \frac{2}{7}}{\frac{1}{2} * \frac{2}{7} + \frac{1}{2} * \frac{2}{8}} =$
- $P\left(\frac{Bag2}{Red}\right) = \frac{yt}{xz+yt} = \frac{\frac{1}{2} * \frac{2}{8}}{\frac{1}{2} * \frac{2}{7} + \frac{1}{2} * \frac{2}{8}}$



Reference Books:

1. Russel and Norvig, 'Artificial Intelligence', third edition, Pearson Education, PHI, (2015)

2. Elaine Rich & Kevin Knight, 'Artificial Intelligence', 3nd Edition, Tata Mc Graw Hill Edition, Reprint(2008)

Sites and Web links:

1. https://www.virtusa.com/digital-themes/heuristic-search-techniques
2. https://towardsdatascience.com/a-star-a-search-algorithm-eb495fb156bb
3. https://www.tutorialandexample.com/local-search-algorithms-and-optimization-problem/
4. http://cgi.di.uoa.gr/~ys02/siteAI2008/local-search-2spp.pdf
5. http://aima.eecs.berkeley.edu/slides-pdf/chapter04b.pdf