8/11/2022

# MATH 103 L01 (Probability)

## 2.1 Events and Sample Space

Probability is defined on :

- The outcomes of an experiment (elements of a set)
- Sample space (universal set)
- Events, collection of outcomes (subsets of sample space)

Sample Space $\Omega$ (omega)

A particular outcome $w \in \Omega$ is a sample point

An experiment taking place corresponds to one of the $w \in \Omega$ occurring

$A = [-5, 5]$ $B = \mathbb{Z}$ $C = [0, \infty)$ all subsets of $\mathbb{R}$

$(A \cap B)^c$ complement = Entire set $[(A \cup B) \cup (A^c \cap B^c)] \setminus (A \cap B)$ In set form

$= \{x \in \mathbb{R} : x \neq -5, -4 \dots 4, 5\}$ Number form

## 3.1 The axioms of probability -

Let $\Omega$ be a sample space. The probability $P$ is a real-valued function defined on subsets of $\Omega$ following three properties -

Axiom 1 (positivity) $P(A) \geq 0$ for all $A \subset \Omega$

Axiom 2 (normalisation) $P(\Omega) = 1$

Axiom 3 (countable additivity) if $A_1$ and $A_2$ are disjoint events, $P(A_1 \cup A_2) = P(A_1) + P(A_2)$

Generally if $A_1, A_2, A_3 \dots \subset \Omega$ are events that are pairwise disjoint. ($A_i \cap A_j = \emptyset$ when $i \neq j$)

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

20/11/2022

# MATH 103 L02

## Chapter 3: The Axiomatic approach

By the first part of axiom 3 when $A_1 \dots A_n$ are pairwise disjoint then:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) \text{ for any } n \in \mathbb{N}$$

This is the finite additivity of P.

### 3.2 Consequences of axioms -

Theorem 3.4 (Monotonicity) -
If $A \subseteq B$ then $P(A) \leq P(B)$

Theorem 3.5 (Complementary events)
$P(A^c) = 1 - P(A)$

Theorem 3.8 (Partition Law)
$P(A) = P(A \cap B) + P(A \cap B^c)$

5/1/2022

MATH 103 L03 + L04

$P(A|B) = P(A \cap B) / P(B)$

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

$P(B|A)P(A) = P(A \cap B) = P(A|B)P(B)$

2/2/2022

# MATH 103 L01 → L05 Recap

2-1 Probability is defined by :
* Outcomes of an experiment (elements of a set)
* Sample space (universal set)
* Events, collection of outcomes (subsets of sample space)

The set of all possible outcomes is $\Omega$ (omega) and is known as the sample space. A particular outcome $\omega \in \Omega$ is a sample point. The set of outcomes in the sample space are exhaustive and exclusive.

An event A is a subset of the possible outcomes contained in the sample space $\Omega$. $A \subseteq \Omega$ occurs if $\omega \in \Omega$ satisfies $\omega \in A$.

Operations :
Union (OR), Intersection (AND)
A and B are mutually exclusive / disjoint if they have no outcomes in common ($A \cap B = \emptyset$)
$A^C$ = Complement of A.

A partition of the sample space splits the sample space into disjoint subsets. Generally the K sets $A_1, A_2, \dots A_K$ form a partition of the set B if the sets $A_1, A_2, \dots A_K$ are mutually exclusive and exhaustive such :

$$A_i \cap A_j = \emptyset \quad \text{for all } i \neq j$$

and

$$B = A_1 \cup A_2 \cup \dots A_K$$

2.2 Discrete Uniform Law -
This applies to situations where $\Omega$ is finite and each of the $n$ sample points are equally likely.

$$P(A) = \frac{|A|}{|\Omega|}$$

$|A|$ = number of sample points in $A$ and $|\Omega|$ = number of sample points in $\Omega$

$P(\{\omega_i\}) = 1/|\Omega|$

(Combinatorics) Counting Principle:
If an experiment involves two stages with $N_1$ and $N_2$ possible outcomes and the second outcome can't influenced by the first (independent) then total outcomes $= N_1 N_2$

Counting Principle (Principle 2) The number of ways of arranging $n$ objects in a line (permuting $n$ objects) $= n!$

Binomial coefficient:
* Number of permutations of $n$ objects composed of two types with $r$ of one type and $n-r$ of the other.

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Counting Principle (Principle 3) The number of ways of forming a committee of $k$ people out of a population of $n$ people is $\binom{n}{k}$

# Chapter 3: Axiomatic Approach

## 3.1 Axioms of probability

The probability $P$ is a real-valued function defined on subsets of $\Omega$ that satisfies the following three properties.

* Positivity $P(A) \geq 0$ for all $A \subseteq \Omega$
* Normalisation $P(\Omega) = 1$
* Countable Additivity If $A_1$ and $A_2$ are disjoint
  $P(A_1 \cup A_2) = P(A_1) + P(A_2)$

Finite additivity: whenever $A_1, \dots A_n$ are pairwise disjoint $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$

Theorem 3.4 (Monotonicity)
If $A \subseteq B$ then $P(A) \leq P(B)$

Theorem 3.5 (Complementary Events)
$P(A^c) = 1 - P(A)$

Theorem 3.8 (Partition Law)
$P(A) = P(A \cap B) + P(A \cap B^c)$

Theorem 3.9 (Total probability Theorem)
Consider pairwise disjoint events $B_1, B_2 \dots B_n$ and $\Omega = \bigcup_{i=1}^{n} B_i$ then for any event $A \subseteq \Omega$

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i)$$

Generally if $\Omega = \bigcup_{i=1}^{\infty} B_i$ $\quad P(A) = \sum_{i=1}^{\infty} P(A \cap B_i)$

Theorem 3.11 (Addition Law)

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

3.3 Conditional Probability -

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

and by consequence

$P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$

Theorem 3.18 (Law of Total Probability) -

$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$

3.4 Bayes Theorem -

If A and B are events in the sample space with $P(A), P(B) > 0$ then,

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

Proof:

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

$$P(B|A)P(A) = P(A \cap B) = B(A|B)P(B). \quad \square$$

3.5 Independent events -

In the special case when $P(A|B) = P(A)$ we say A is independent of B. A and B are independent events if and only if

$$P(A \cap B) = P(A)P(B)$$

FS1
Chp 1

# Chapter 4: Discrete Random Variables -

A random variable $R$ is a function $R: \Omega \to \mathbb{R}$
For each outcome in the sample space $\omega \in \Omega$
it associates a unique real number $R(\omega)$

Every time an experiment is conducted exactly one value of the random variable is observed this is called a realisation of the random variable.

$R((i,j)) = i+j$
$S((i,j)) = \max\{i,j\}$ the bigger of the two values.

The range of values of $\{R(\omega): \omega \in \Omega\}$ is known as the induced sample space of $R$, written $S$

Discrete random variables are when $S$ is finite.

## 4.2 Probability mass functions -

The probability mass function (pmf) of a discrete random variable, $R$ is defined by

$$p_R(r) = P(R=r)$$

for all $r \in S$

note, $p_R(r) \geq 0$

$$1 = P(\Omega) = P((R=r_1)) \cup P(R=r_2) \cup P(R=r_3)$$
$$= p_R(r_1) + p_R(r_2) + p_R(r_3) + \dots$$

$p_R(r) \geq 0$

$$\sum_{r \in S} p_R(r) = 1$$

e.g $\Omega = \{a, b, c, d\}$ $R(a) = 2$ $R(b) = 4$ $R(c) = 3$ $R(d) = 2$
$S = \{2, 3, 4\}$ so,
$p_R(2) = P(\{a, d\}) = 1/2$
$p_R(3) = P(\{c\}) = 1/4$
$p_R(4) = P(\{b\}) = 1/4$

4.3 Probability of an event.

Lemma 4.10 -
Let $E \subseteq S$ be an event in the induced sample space
probability of E,
$$P(R \in E) = \sum_{r \in E} p_R(r)$$

Cumulative distribution function (cdf) of a random variable R is a function $F_R : \mathbb{R} \to \mathbb{R}$
$$F_R(m) = P(R \leq m)$$
The cdf is given by,
$$F_R(m) = P(R \leq m) = \sum_{r=0}^{m} p_R(r)$$

a cdf has properties.
$F_R(r_1) \leq F_R(r_2)$ etc.
$\lim_{r \to -\infty} F_R(r) = 0$
$\lim_{r \to \infty} F_R(r) = 1$

4.4 Expectation.
outcomes $r_1, r_2 \dots$ The mean observed value of R
is $\dfrac{r_1 + r_2 + \dots + r_n}{n}$
$$= \sum_{r \in S} r n_r$$
$$= \sum_{r \in S} r p_R(r)$$

Expectation
$$E(R) = \sum_{r \in S} r P(R)(r) = \sum_{w \in \Omega} R(w) P(\{w\})$$

3/2/2022

# MATH 103 L06 (5 - 5.6)

## Chapter 5: Models for discrete random variables -

Following identities useful in further developments -

Arithmetic progression: $\sum_{i=1}^{n} i = n(n+1)/2$

Sum of Squares: $\sum_{i=1}^{n} i^2 = n(n+1)(2n+1)/6$

Exponential Series: $\exp(x) = 1 + x + \frac{x^2}{2!} + \dots = \sum_{i=0}^{\infty} \frac{x^i}{i!}$

$e = \exp(1)$

Partial Geometric sum:

$$S_m = 1 + x + x^2 \dots = \sum_{i=0}^{m} x^i = \frac{1 - x^{m+1}}{1 - x}$$

Geometric sum:

$$S_\infty = 1 + x + x^2 \dots = \sum_{i=0}^{\infty} x^i = \frac{1}{1-x} \text{ for } |x| < 1$$

Weighted Geometric sum:

$$(1-x)^{-1} = \sum_{i=0}^{\infty} x^i \qquad (1-x)^{-2} = \sum_{i=1}^{\infty} i x^{i-1}$$

$$2(1-x)^{-3} = \sum_{i=2}^{\infty} i(i-1) x^{i-2} \text{ (By differentiating)}$$

Binomial Expansion:

$$(p+q)^n = \binom{n}{0} p^n q^0 + \binom{n}{1} p^{n-1} q^1 + \dots$$

Variance = MSMSM $E(X^2) - E(X)^2$

e.g find expectation and variance of $P_R(r) = \frac{1}{m+1}$

$$E(X) = \sum_{r=0}^{m} r \frac{1}{m+1} = \frac{1}{m+1} \times \tfrac{1}{2} m(m+1) = m/2$$

$$E(X^2) = \sum_{r=0}^{m} r^2 \frac{1}{m+1} = \frac{1}{m+1} \times \tfrac{1}{6} m(m+1)(2m+1) = \frac{m(2m+1)}{6}$$

$$\mathrm{Var}(X) = \frac{m(2m+1)}{6} - \frac{m^2}{4} = \frac{m(m+2)}{12}$$

5.3 Bernoulli Random variables -

An experiment of sample space $\{0, 1\}$ and probability of one is $\theta$ $(0 \leq \theta \leq 1)$. This is termed a Bernoulli random variable.

$$PR(0) = 1-\theta \quad PR(1) = \theta \quad R \sim Bernoulli(\theta)$$

$E(x) = 0(1-\theta) + 1\theta = \theta$

$E(x^2) = 0^2(1-\theta) + 1^2\theta = \theta$

$Var(x) = \theta - \theta^2$

5.4 Binomial Random variables -

An experiment of $n$ independent Bernoulli trials with each having probability $\theta$. Sample space $= \{0, 1, \dots n\}$

$R \sim Bin(n, \theta)$

e.g $PR(0) = FFF = (1-\theta)^3$ $PR(1) = FFT = 3\theta(1-\theta)^2$

etc This is a Binomial model.

$PR(r) = \binom{n}{r} \theta^r (1-\theta)^{n-r}$ $E(R) = n\theta$ $Var(R) = n\theta(1-\theta)$

5.5 Geometric Random variables -

An experiment on independent Bernoulli trials. This continues until there is a success and counts the number of trials upto and not including the success. $R(FFFFT) = 4$.

$PR(r) = (1-\theta)^r\theta$ $E(R) = \dfrac{1-\theta}{\theta}$ $Var(R) = \dfrac{1-\theta}{\theta^2}$

5.6 Poisson Random variables -

$PR(r) = \dfrac{\lambda^r \exp(-\lambda)}{r!}$ $R \sim Pois(\lambda)$

5.9 Summary -

8/2/2022

# MATH 103 L07

## Chapter 6: Continous Random Variables

In continous events we focus on probabilities of events instead of probabilities of single outcomes. $P(X \leq x)$.
For discrete random variables the cumulative distribution function $F_R(r) = P(R \leq r)$ for discrete values r.

### 6.2 - Cumulative Distribution Function
properties of $F_X(x)$:

$0 \leq F_X(x) \leq 1$ with $\lim_{x \to -\infty} F_X(x) = 0$ $\lim_{x \to \infty} F_X(x) = 1$

Probability of intervals

$P(a < X \leq b) = P(X \leq b) - P(X \leq a)$

### 6.3 Probability density function -
$F(x) := \int_{-\infty}^{x} f(s)\, ds$

e.g suppose $f(x) = \begin{cases} 0 & \text{when } x < 1 \\ 1/x^2 & \text{otherwise} \end{cases}$

Then for $x < 1$ $F(x) = 0$

when $x \geq 1$

$\int_{-\infty}^{x} f(s)\,ds = \int_{1}^{x} f(s)\, ds = \left[-\frac{1}{s}\right]_1^x = 1 - \frac{1}{x}$

For $x \geq 1$ $\int f(x)\,dx = -\frac{1}{x} + C$

Definition:

$f_X(x) = \frac{d}{dx} F_X(x)$

$F_X(x) = \int_{-\infty}^{x} f(x)(s)\, ds$

e.g

$P(X \leq 10) = F_X(10) = \int_{-\infty}^{10} f_X(s)\, ds$

Key Idea: $P(x < X \leq x + \delta) = \int_{x}^{x+\delta} f_X(s)\,ds = f_X(x)\delta$
for some very small interval $\delta$.

6.4 Expectation and Variance -

$E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx \quad E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx$

Quantiles : $x_p$ is the $100p\%$ quantile defined by $F_X(x_p) = p$.

Median : $F(x_{0.5}) = 0.5$

MATH 103 L08

Chapter 7 : Models for continuous random variables -

7.1 Uniform Distribution :

A continuous random variable for which all outcomes in a given range have equal chance of occurring is said to be uniformly distributed. X has a uniform distribution over (a, b) if the pdf =

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

and write $X \sim U(a,b)$

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_a^b \frac{x}{b-a}\,dx = \frac{b+a}{2}$$

$$E(X^2) \ldots = \frac{b^2+ab+a^2}{3}$$

$$Var(X) = \frac{(b-a)^2}{12}$$

7.2 Exponential Distribution :

A random variable X has an exponential distribution with rate $\beta$ if its pdf is given by

$$f_X(x) = \begin{cases} C\exp(-\beta x) & x \geq 0 \\ 0 & \text{otherwise} \end{cases} \qquad X \sim Exp(\beta)$$

where $\beta > 0$ and C = normalising constant.

Lack of Memory :

A Key property of the exponential distribution is it's lack of memory property.

A random variable satisfies this property if :

$$P(X > s+t \mid X > t) = P(X > s)$$

Gamma function : $\Gamma(\alpha)$

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1}\exp(-t)\,dt$$

Expectation and variance of Exp(B) rvs:

The $r^{th}$ moment is defined as $E(X^r)$

$$E(X^r) = \beta \int_0^\infty x^r \exp(-\beta x)\, dx$$

Lemma: $\int_0^\infty x^{\alpha-1} \exp(-\beta x)\, dx = \frac{\Gamma(\alpha)}{\beta^\alpha}$

$E(X) = \frac{1}{\beta}$ $\quad$ $Var(X) = \frac{1}{\beta^2}$ for exponential random variables.

7.3 Gamma Distribution:

A random variable X has a gamma distribution with shape parameter $\alpha$ and rate parameter $\beta$ if it's pdf is given by

$$f_X(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x) & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\alpha > 0$ and $\beta > 0$. We write $X \sim Gamma(\alpha, \beta)$

Lemma 7.9 -

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

$r^{th}$ moment of a gamma random variable.

$$E(X^r) = \frac{\Gamma(r+\alpha)}{\beta^r \Gamma(\alpha)}$$

$E(X) = \frac{\alpha}{\beta}$ $\quad$ $E(X^2) = \frac{(\alpha+1)\alpha}{\beta^2}$ $\quad$ $Var(X) = \frac{\alpha}{\beta^2}$

7.4 Normal Distribution -

A random variable X has a normal distribution (Gaussian distribution)

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

MATH 103 L09 + L10

Chapter 8 : (More than one random variable)

8.1 Joint mass probability functions -

Let X and Y be discrete random variables defined on the sample space $\Omega$. The joint probability mass function is

$$p_{X,Y}(x,y) = P(X = x, Y = y)$$

Properties :

* For all $x$ and $y$ $0 \leq p_{X,Y}(x,y) \leq 1$
* $\sum_{\text{all } x,y} p_{X,Y}(x,y) = 1$
* $P((X,Y) \in A) = \sum_{(x,y) \in A} p_{X,Y}(x,y)$

Marginal probability mass functions:

$$p_X(x) = \sum_{y=0}^{\infty} p_{X,Y}(x,y)$$

$$p_Y(y) = \sum_{x=0}^{\infty} p_{X,Y}(x,y)$$

8.2 Independence -

Two random variables X and Y are independent if the events $\{X \in A\}$ and $\{Y \in B\}$ are independent for all sets A and B

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

for all sets A, B.

Theorem 8.4 -

Two discrete random variables X and Y are independent iff $p_{X,Y}(x,y) = p_X(x)p_Y(y)$ for all $x$ and $y$.

If X and Y are discrete random variables the conditional pmfs

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} \qquad p_{Y|X}(y|x) = \frac{p_{XY}(x,y)}{p_X(x)}$$

Theorem 8.8 - Two continuous random variables X and Y are independent if

$$f_{XY}(x,y) = f_X(x) f_Y(y).$$

This result is used in likelihood-based estimates in statistics.

8.3 Weak law of large numbers -

Let $X_1, X_2 \dots X_n$ be jointly distributed random variables with finite expectation and variance.

* $E(X_1 + X_2 + \dots X_n) = E(X_1) + E(X_2) + \dots E(X_n)$
* If $X_1, X_2, \dots X_n$ are independent
  $Var(X_1 + X_2 \dots) \quad " = Var(X_1) + Var(X_2) + \dots$

Let $X_i$ be the measured value on the $i^{th}$ experiment.

$\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots X_n)$

$E(\bar{X}) = \mu$

$Var(\bar{X}) = \frac{\sigma^2}{n}$

$$P(|\bar{X} - \mu| > k\sigma) \leq \frac{1}{k^2 n}$$

As n gets large, probability that sample average $\bar{X}$ is more than $k\sigma$ away from expected value decreases to 0. Since k is arbitrary we can say $\bar{X}$ converges to $\mu$. This is called the weak law of large numbers.