Unit I Introduction to Data Modelling (07 Hours)

Basic probability:
Discrete and continuous random variables, independence, covariance, central limit theorem, Chebyshev inequality, diverse continuous and discrete distributions.
Statistics, Parameter Estimation, and Fitting a Distribution: Descriptive statistics, graphical statistics, method of moments, maximum likelihood estimation
Data Modeling Concepts • Understand and model subtypes and supertypes • Understand and model hierarchical data • Understand and model recursive relationships • Understand and model historical data

Basic Probability Concepts :-
ref
https://web.stanford.edu/class/hrp259/2007/discrete/discrete_259_2007.ppt

- **Probability** – the chance that an uncertain event will occur (always between 0 and 1)

- **Impossible Event** – an event that has no chance of occurring (probability = 0)

- **Certain Event** – an event that is sure to occur (probability = 1)

Assessing Probability

There are three approaches to assessing the probability of an uncertain event:

1. *a priori -- based on prior knowledge of the process*

2. empirical Probability

3. Subjective Probability

Assuming all outcomes are equally likely

probability of occurrence

probability of occurrence

based on a combination of an individual's past experience, personal opinion, and analysis of a

2. Example of empirical probability

Find the probability of selecting a male taking statistics from the population described in the following table:

	Taking Stats	Not Taking Stats	Total
Male	84	145	229
Female	76	134	210
Total	160	279	439

Subjective probability

- Subjective probability may differ from person to person
  - A media development team assigns a 60% probability of success to its new ad campaign.
  - The chief media officer of the company is less optimistic and assigns a 40% of success to the same campaign
- The assignment of a subjective probability is based on a person's experiences, opinions, and analysis of a particular situation
- Subjective probability is useful in situations when an empirical or a priori probability cannot be computed

Events

Each possible outcome of a variable is an event.

- Simple event
  - An event described by a single characteristic
  - e.g., A day in January from all days in 2015
- Joint event
  - An event described by two or more characteristics
  - e.g. A day in January that is also a Wednesday from all days in 2015
- Complement of an event A (denoted A')
  - All events that are not part of event A
  - e.g., All days from 2015 that are not in January

Random Variable

A random variable *X* takes on a defined set of values with different probabilities.

For example, if you roll a die, the outcome is random (not fixed) and there are 6 possible outcomes, each of which occur with probability one-sixth.

For example, if you poll people about their voting preferences, the percentage of the sample that responds "Yes on Proposition 100" is a also a random variable (the percentage will be slightly different every time you poll).

Roughly, probability is how frequently we expect different outcomes to occur if we repeat the experiment over and over ("frequentist" view)

Random variables can be discrete or continuous

- **Discrete** random variables have a countable number of outcomes
  - Examples: Dead/alive, treatment/placebo, dice, counts, etc.
- **Continuous** random variables have an infinite continuum of possible values.
  - Examples: blood pressure, weight, the speed of a car, the real numbers from 1 to 6.

Probability functions

- A probability function maps the possible values of *x* against their respective probabilities of occurrence, *p(x)*
- *p(x)* is a number from 0 to 1.0.
- The area under a probability function is always 1.