MM 225 – AI and Data Science

Day 3: Random Variable

Instructors: Hina Gokhale, MP Gururajan, N. Vishwanathan

1 AUGUST 2024

Problem from the last class

Suppose that an insurance company classifies people into one of three classes — good risks, average risks, and bad risks. Their records indicate that the probabilities that good, average, and bad risk persons will be involved in an accident over a 1-year span are, respectively, .05, .15, and .30. If 20% of the population are "good risks," 50 %"average risks," and 30% are "bad risks," what proportion of people have accidents in a fixed year? If policy holder A had no accidents in 1987, what is the probability that he or she is a good risk?

Random Variable -- rationale

- Numerical outcome of a random experiment is desirable.
- Not all random experiment result in numeric outcome. Sometimes outcomes are describable. Many times such outcomes also serve the purpose.
- However, it is useful if such outcomes can be "mapped" to a numeric value.
- Such a Mapping of outcome of a Random Experiment is called a Random Variable.

Random Variable - Definition

- Let S denote sample space.
- Define a function X : S → **R,** where R is the set of Real Numbers
- Such a function X is called random variable.
- If the range of RV X is finite or countable then X is called discrete RV
- Otherwise, X is called continuous RV

Random Variable - Example

- X = Run made by Virat Kohli in a match of T20
  - Is it continuous or discrete RV?
- Y = average of Virat Kohli
  - Is Y discrete or Continuous?
- Z = number of players who averaged more than 50
  - Is Z discrete of Continuous?
- $W = \begin{cases} S & if\ India\ won\ the\ World\ Cup \\ F & if\ India\ lost\ the\ Worl\ Cup \end{cases}$
  - Is W Discrete or Continuous?

Random Variable - Notation

- Random Variables are denoted by CAPITAL LETTERS: X, Y, …
- The values that random variable take are denoted by small letters: x, y, ….
-
- Example:
- Consider heights of students in this class: RV X
- Height of a particular student = x

Cumulative Distribution Function - CDF

Let X be any random variable, cumulative distribution function (CDF) of X is defined as :

$$F(t) = P(X \leq t)$$

Therefore, for any random variable X we have:

1. $0 \leq F(t) \leq 1$
2. If $x \leq y$ then $F(x) \leq F(y)$

Discrete RV – Probability Mass Function

- Let X be a discrete random variable taking on values $\{x_1, x_2, x_3, \ldots, x_n\}$, then **Probability Mass Function (pmf)- $f(x_i)$** is a function such that

1. $f(x_i) = P(X = x_i)$
2. $f(x_i) \geq 0$
3. $\sum_{i=1}^{n} f(x_i) = 1$

CDF for Discrete RV

- Let X be a discrete random variable taking on values $\{x_1, x_2, x_3, \ldots, x_n\}$, with Probability Mass Function (pmf)- $f(x_i)$. Then Cumulative Distribution Function of X is defined as

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

Mean and Variance of DRV

- Let X be a discrete random variable taking on values $\{x_1, x_2, x_3, \ldots, x_n\}$, with Probability Mass Function (pmf)- $f(x_i)$.
- Mean of X is $\mu = E(X) = \sum_{i=1}^{n} x_i\, f(x_i) = \sum_x x f(x)$
- Variance of X = $\sigma^2 = Var(X) = E(X-\mu)^2 = \sum_x (x-\mu)^2 f(x)$

$$= \sum_x x^2 f(x) - \mu^2$$

- Standard deviation of X $= \sigma = \sqrt{\sigma^2}$

Example of DRV

- A performance test is carried out on a newly developed device.
- Probability that the device would pass the test is 0.85.
- Two devices are tested independently.
- Let X = number of devices pass the test.
- Values RV X can take is :
  - pp → $X = 2$ and $P(X = 2) = 0.85 * 0.85 = 0.7225$
  - pf → $X = 1$ and $P(X = 1) = 0.85 * 0.15 = 0.1275$
  - fp → $X = 1$ and $P(X = 1) = 0.15 * 0.85 = 0.1275$
  - ff → $X = 0$ and $P(X = 0) = 0.15 * 0.15 = 0.0225$

Example…continue….

- Here X takes on values 0, 1 and 2.
- Hence pmf is
  - f(X = 0) = 0.0225
  - f(X = 1) = 0.2550
  - f(X = 2) = 0.7225

- CDF of X is

- $$F(x) = \begin{cases} 0 \;\; for\; x < 0 \\ 0.0225\; for\; 0 \;\leq x < 1 \\ 0.2775\; for\; 1 \;\leq x < 2 \\ 1\; for\; x \;\geq 2 \end{cases}$$

In class problem solving…..

- Here X takes on values 0, 1 and 2.
- Hence pmf is
  - $f(X = 0) = 0.0225$
  - $f(X = 1) = 0.2550$
  - $f(X = 2) = 0.7225$

- $E(X) =$

- $Var(X) =$

Thank you