Discrete Distribution

DR.ALAA MOHAMMED

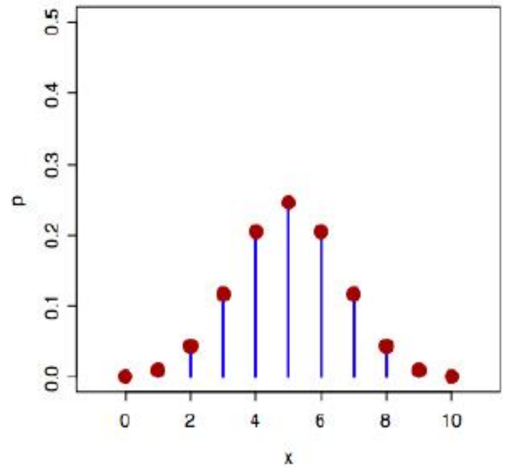
3RD LEVEL, 2ND SEMESTER, BIOMEDICAL INSTRUMENTATIONS AND BIOMECHANICS BRANCHES, BIOMEDICAL ENG. DEPARTMENT

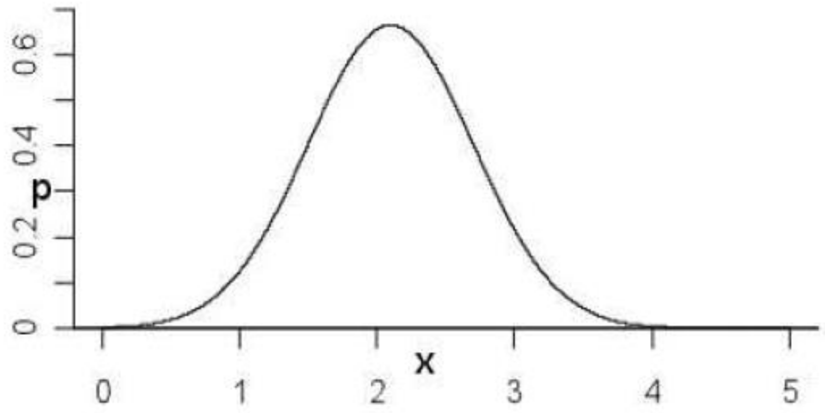
Random Variables

Informally, a random variable X denotes possible outcomes of an event can be discrete (i.e., finite many possible outcomes) or continuous

A random variable is a random number determined by chance, or more formally, drawn according to a probability distribution

A random variable is a variable that assumes *numerical values* associated with random outcomes of an experiment





Random Variables (cont.)

- the probability distribution can be given by the physics of an experiment (e.g., throwing dice)
- the probability distribution can be synthetic
- discrete & continuous random variables
- Typical random variables in Machine Learning Problems
- the input data
- the output data
- noise

Random Variables (cont.)

❖ Important concept in learning:

The data generating model e.g., what is the data generating model for:

I. throwing dice,
II. regression,
III. classification,
IV. visual perception?

Random Variables (cont.)

Some examples of discrete r.v.

- A random variable $x \in \{0,1\}$ denoting outcomes of a coin-toss
- A random variable $x \in \{1,2,\ldots,6\}$ denoting outcome of a dice roll

Some examples of continuous r.v.

- A random variable $x \in \{0,1\}$ denoting the bias of a coin
- A random variable x denoting heights of students in this class
- A random variable x denoting time to get to your hall from the department

Continuous Random Variables

Random variables which consist of measurements are usually not discrete. For example, height of students in a tutorial class.

They are called *continuous* random variables and can take on any real value in a given interval, rather than being restricted to integers.

Further examples include weight of babies, the amount of sugar in a blood sample, etc.

Discrete Random Variables

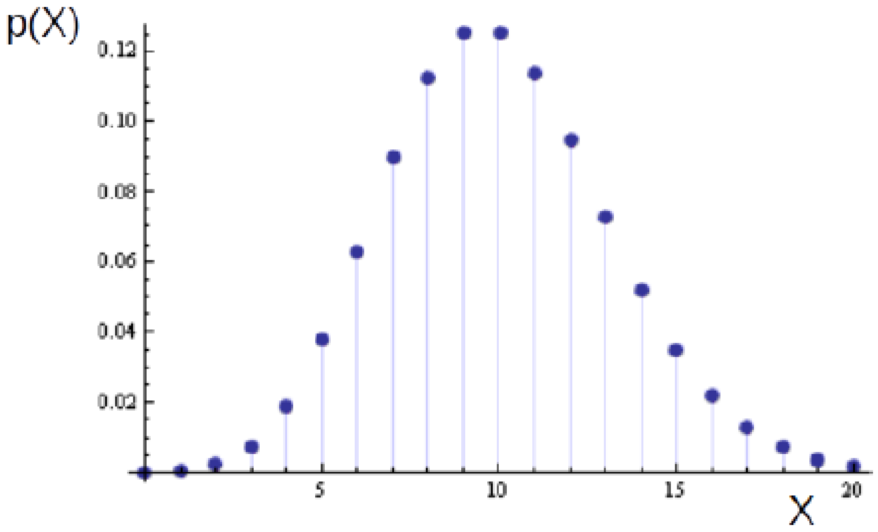
For a discrete $r.v. X, p(x)$ denotes the probability that $p(X \; = \; x)$

$p(x)$ is called the probability mass function (PMF)

p(x) $\geq$ 0

p(x) $\leq$ 1

$\sum_x p(x) = 1$



Examples:

The following are discrete random variables:

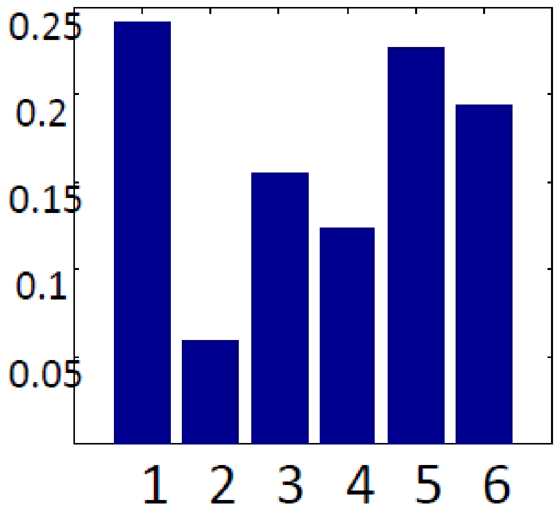
- number of children per family;
- attendance of MATH1015 lectures;
- number of patients admitted to an ER each day, etc..

Discrete Probability Distributions

- The random variables only take on **discrete** values

  e.g., throwing dice: possible values

$$v_i \in \{1,2,3,4,5,6\}$$



- The probabilities sum to 1

$$\sum_i P(v_i) = 1$$

- Discrete distributions are particularly important in classification
- Probability Mass Function or Frequency Function (normalized histogram)

Classic Discrete Distributions (I)

Bernoulli Distribution

If a random variable X can take on only two values (labelled $0 =$ "failure" and $1 =$ "success"), then the distribution of X is, by definition, Bernoulli(p), where p is the probability of success.

$P(0) = p$ and $P(1) = 1 - p$, or in compact notation:

$$P(x) = \begin{cases} p^x(1-p)^{1-x}, \text{if } x = 0 \text{ or } x = 1 \\ 0, \text{ otherwise} \end{cases}$$

Bernoulli distributions are naturally modeled by sigmoidal activation functions in neural networks with binary inputs.

Example

Say a patient (unknowingly) has strep throat. The doctor uses a throat swab test, which comes back negative with probability 1- p or positive with probability p.

Defining correct diagnosis as success, the outcome of the test has a Bernoulli(p) distribution.

In this context, the probability p is known as the *sensitivity* of the test.

Classic Discrete Distributions (II)

Binomial Distribution

Like Bernoulli distribution, binary input variables: 0 or 1, and probability $P(0) = p$ and $P(1) = 1\text{-}p$

What is the probability of *k* successes, *P*(*x*), in a series of *n* independent trials? $(n \geq x)$

$P(x)$ is a binomial random variable:

$$p(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n.$$

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

Classic Discrete Distributions (II)

Binomial variables are important for density estimation networks, e.g. “what is the probability that *k* data points fall into region R?”

Bernoulli distribution is a subset of binomial distribution (i.e., n=1)

Example

If *n* patients with strep throat come to the ER, then the number who receive a positive (correct) diagnosis is distributed as Binomial(*p*) (assuming that test results are independent, e.g. swabs are not processed in batches).

The probability that at least *k* of these *n* people receive a positive diagnosis is

$$P(X \geq k) = \sum_{x=k}^{n} p(x)$$

Classic Discrete Distributions (III)

Geometric Distribution

If we again observe *n* independent trials, each of which results in “success" with probability *p* and “Failure" with probability $1 - p$, then the number of *trials* required to observe the first success, X has, by definition, a Geometric(*p*) distribution.

Specifically, the pmf of *X* is:

$$p(x) = P(X = x) = (1 - p)^{x-1}p, \qquad x = 1, 2, \ldots .$$