Random Variables and Probability Distributions

by Dr. Mohd. Javed

November 3, 2020

Random Variable

- A Random variable is used to map the outcome of a random process to numbers.
- 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, each of which occur with probability one-sixth.

Event	Probability
x	$p(x) = P(X = x)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Figure: Rolling of a die

Random Variables

It can be discrete or continuous

- **Discrete** : random variables have a countable number of outcomes.
    - For example yes/no, dice, counts etc.
- **Continuous** : random variables have an infinite continuum of possible values.
    - For example blood pressure, weight, the speed of a car, the real numbers from 1 to 6 etc.

Probability Functions

- A probability functions 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.

Probability Density Functions

- A probability density function is most commonly associated with absolutely continuous univariate distributions.
- A random variable X has density f_X, where f_X is a non-negative Lebesgue-integrable function, if:

  $$\Pr[a \leq X \leq b] = \int_a^b f_X(x)\, dx.$$
- If F_X is the cumulative distribution function of X, then:

  $F_X(x) = \int_{-\infty}^{x} f_X(u)\, du$, and (if f_X is continuous at x)

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

Discrete example: roll of a die

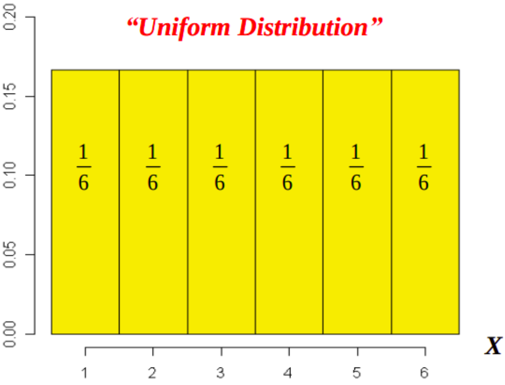


Figure: Rolling of a die

Discrete example: roll of a die

Event	Probability
x	$p(x) = P(X = x)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Figure: Rolling of a die

Cumulative distribution function (CDF)

- The cumulative distribution function of a real-valued random variable X is the function given by: $F_X(x) = \mathrm{P}(X \leq x) F_X(x) = \mathrm{P}(X \leq x)$ where the right-hand side represents the probability that the random variable X takes on a value less than or equal to x.

- The probability that X lies in the semi-closed interval $(a, b]$, where $a < b$, is therefore $\mathrm{P}(a < X \leq b) = F_X(b) - F_X(a) \mathrm{P}(a < X \leq b) = F_X(b) - F_X(a)$

Cumulative distribution function (CDF)

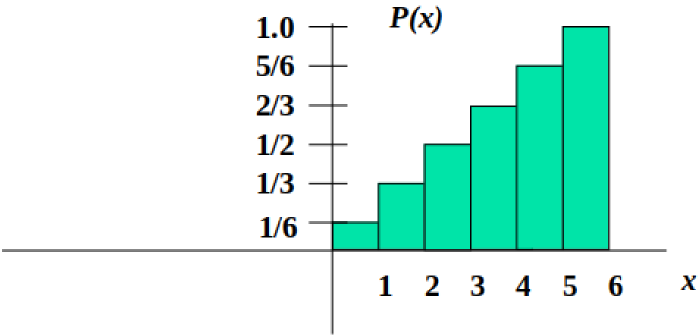


Figure: CDF example

Cumulative distribution function (CDF)

x	$P(x \leq A)$
1	$P(x \leq 1) = 1/6$
2	$P(x \leq 2) = 2/6$
3	$P(x \leq 3) = 3/6$
4	$P(x \leq 4) = 4/6$
5	$P(x \leq 5) = 5/6$
6	$P(x \leq 6) = 6/6$

Table: CDF table

Examples

1. What's the probability that you roll a 3 or less?

2. What's the probability that you roll a 5 or higher?

Examples

1. What's the probability that you roll a 3 or less?
   Ans: $P(x \leq 3) = 1/2$

2. What's the probability that you roll a 5 or higher? Ans : $P(x \geq 5) = 1 - P(x \leq 4) = 1 - 2/3 = 1/3$

Practice Problem

- Which of the following are probability functions?

  a. f(x) = .25 for x = 9,10,11,12

  b. f(x) = (3-x)/2 for x= 1,2,3,4

  c. f(x) = $(x^2 + x + 1)/25$ for x = 0,1,2,3

Practice Problem

- Which of the following are probability functions?

a. f(x) = .25 for x = 9,10,11,12
Ans : Yes

b. f(x) = (3-x)/2 for x= 1,2,3,4
Ans : No

c. f(x) = $(x^2 + x + 1)/25$ for x = 0,1,2,3
Ans : No

Types of Discrete Distribution

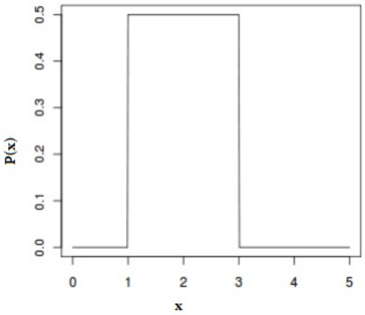
There are a variety of discrete probability distributions that you can use to model different types of data.

- **Bernoulli distribution** to model binary data, such as coin tosses.
- **Binomial distribution** is SUCCESS or FAILURE outcome in an experiment or survey that is repeated multiple times.
- **Poisson distribution** to model count data, such as the count of library book checkouts per hour.
- **Uniform distribution** to model multiple events with the same probability, such as rolling a die

Uniform distribution

- Uniform distributions are probability distributions with equally likely outcomes.
- This distribution is defined by two parameters, a and b:
  - a is the minimum.
  - b is the maximum.

  The distribution is written as U(a,b).
- The following graph shows the distribution with $a = 1$ and $b = 3$:



- **Expected Value**
  The expected value (i.e. the mean) of a uniform random variable X is:
  E(X) = (1/2) (a + b)
  This is also written equivalently as:
  E(X) = (b + a) / 2.
  "a" in the formula is the minimum value in the distribution, and "b" is the maximum value.
- **Variance**
  The variance of a uniform random variable is:
  Var(x) = (1/12)(b-a)2

Poisson distribution

- The Poisson distribution is used to model the number of events occurring within a given time interval. The formula for the Poisson probability mass function is

$$p(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!} for\, x = 0, 1, 2, \tag{1}$$

λ is the shape parameter which indicates the average number of events in the given time interval.

Shape of Poisson Distribution

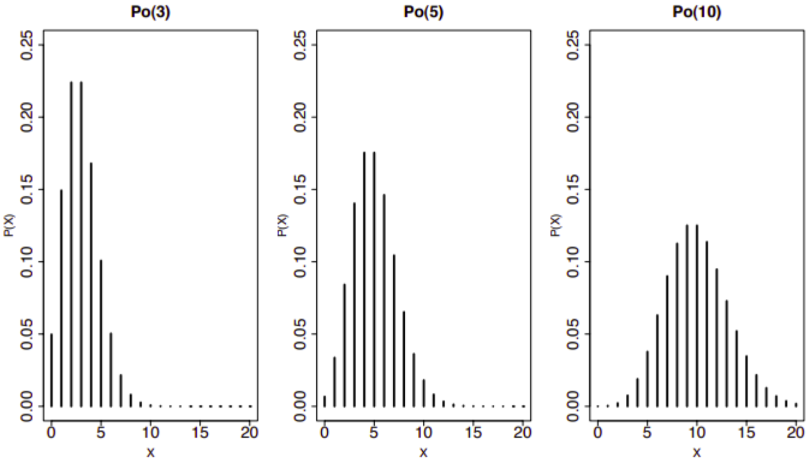


Figure: Poisson Distribution

Shape of Poisson Distribution

We observe that the Poisson distributions

1. are unimodal;
2. exhibit positive skew (that decreases as λ increases);
3. are centred roughly on λ;
4. have variance (spread) that increases as λ increases.

- **Expected Value**
  The expected value (i.e. the mean) of a Poisson random variable X is:
  $E(X) = \lambda$

- **Variance**
  The variance of a Poisson random variable is:
  $Var(x) = \sqrt{\lambda}$

Bernoulli Distribution

- The Bernoulli distribution is a discrete distribution having two possible outcomes labelled by n = 0 and n = 1 in which n = 1 ("success") occurs with probability p and n = 0 ("failure") occurs with probability q = 1-p, where 0 < p < 1.
- It therefore has probability density function

$$p(n) = \begin{cases} 1-p, & \text{for } n = 0 \\ p, & \text{for } n = 1 \end{cases} \tag{2}$$

which can also be written

$$p(n) = p^n(1-p)^{1-n}. \tag{3}$$

- The performance of a fixed number of trials with fixed probability of success on each trial is known as Bernoulli trial.

- **Expected Value**
  The expected value (i.e. the mean) of a Bernoulli random variable X is:
  $E(X) = p$

- **Variance**
  The variance of a Bernoulli random variable is:
  $Var(x) = p(1-p)$

Binomial Distribution

- The binomial distribution gives the discrete probability distribution $P_p(n/N)$ of obtaining exactly n successes out of N Bernoulli trials (where the result of each Bernoulli trial is true with probability p and false with probability q=1-p).
- The binomial distribution is therefore given by

$$\mathrm{P}_p(\mathrm{n}/\mathrm{N}) = \binom{N}{n} p^n q^{N-n} = \frac{N!}{n!(N-n)!} p^n q^{N-n}$$

where $\binom{N}{n}$ is a binomial coefficient.

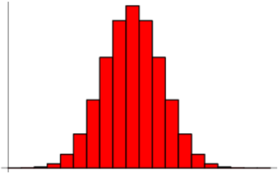


Figure: The above plot shows the distribution of n successes out of N=20 trials with p=q=1/2.

- **Expected Value**
  The expected value (i.e. the mean) of a Binomial random variable X is:
  $E(X) = np$

- **Variance**
  The variance of a Binomial random variable is:
  $Var(x) = np(1\text{-}p)$

- Continuous random variables are described with probability density function (pdfs) curves
- Normal pdfs are recognized by their typical bell-shape.

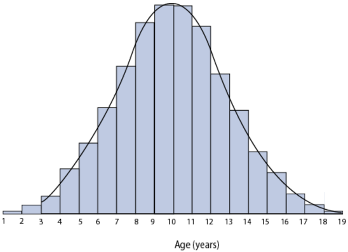


Figure: Age distribution of a pediatric population with overlying Normal pdf.

Gaussian Distribution

- The general form of its probability density function is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{4}$$

- The parameter μ is the mean or expectation of the distribution (and also its median and mode), while the parameter σ is its standard deviation.
- The variance of the distribution is σ^2
- A random variable with a Gaussian distribution is said to be normally distributed, and is called a normal deviate.

- The mean, mode and median are all equal.
- The curve is symmetric at the center (i.e. around the mean, μ).
- Exactly half of the values are to the left of center and exactly half the values are to the right.
- The total area under the curve is 1.

- **Expected Value**
  The expected value (i.e. the mean) of a Standard Gaussian random variable X is:
  $E(X) = 0$

- **Variance**
  The variance of a Standard Gaussian random variable is:
  $Var(x) = 1$

The End