COMP1046
Mathematics for Computer Science

Continuous Probability

Topic Outline

- Continuous Probability Distributions
- Uniform Distribution
- Normal (Gaussian) Distribution

Recommended textbook:

- Ross, Sheldon M., Introduction to Probability and Statistics for Engineers and Scientists (San Diego: Elsevier Science & Technology 2009)

Sections 4.1, 4.2, 5.4 & 5.5.

Available online from the library.

Continuous Random Variable

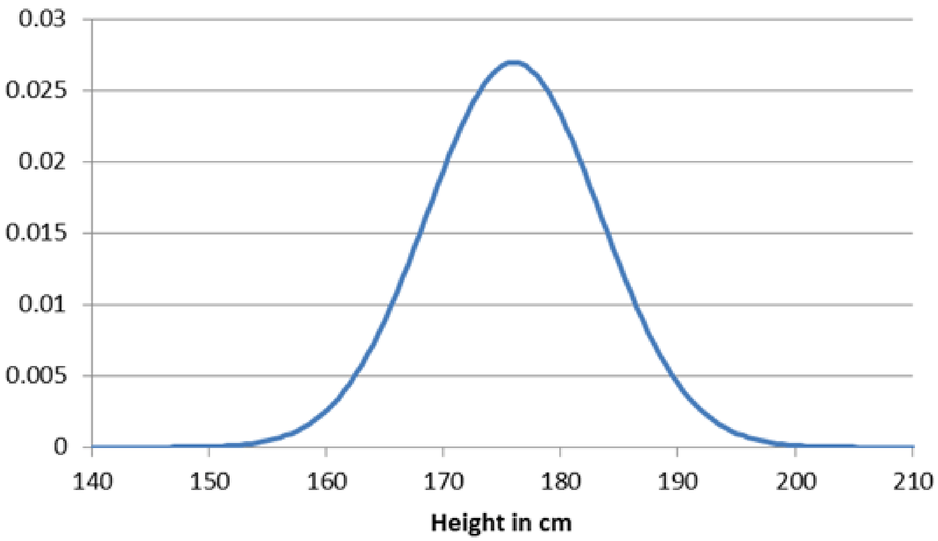
- A discrete random variable is across a finite number of possible values.
- For example, a die gives one of only 6 possible outcomes: 1, 2, 3, 4, 5, 6.

- A continuous random variable is a real number.
- Examples: Height, weight, distance.

* Textbook ref: Sections 4.1 & 4.2 (pages 89-95).

Continuous Random Variable: Example

- Suppose X is the "Height of a person, in meters, in the COMP1046 class".
- We can draw a distribution, something like:



- What is $P(X > 190)$? How would you measure this?
- What is $P(X = 150)$? Can we measure this?

Exercise

Which of these measured quantities do you think are discrete and which continuous random variables, or neither?

1. The rainfall in Ningbo in any one week.
2. The number of slides in a lecture slide set.
3. The distance between two stars in the sky.
4. The distance between two clouds in the sky.
5. Money.
6. The speed of a car.
7. The number of ways three people could form a queue.

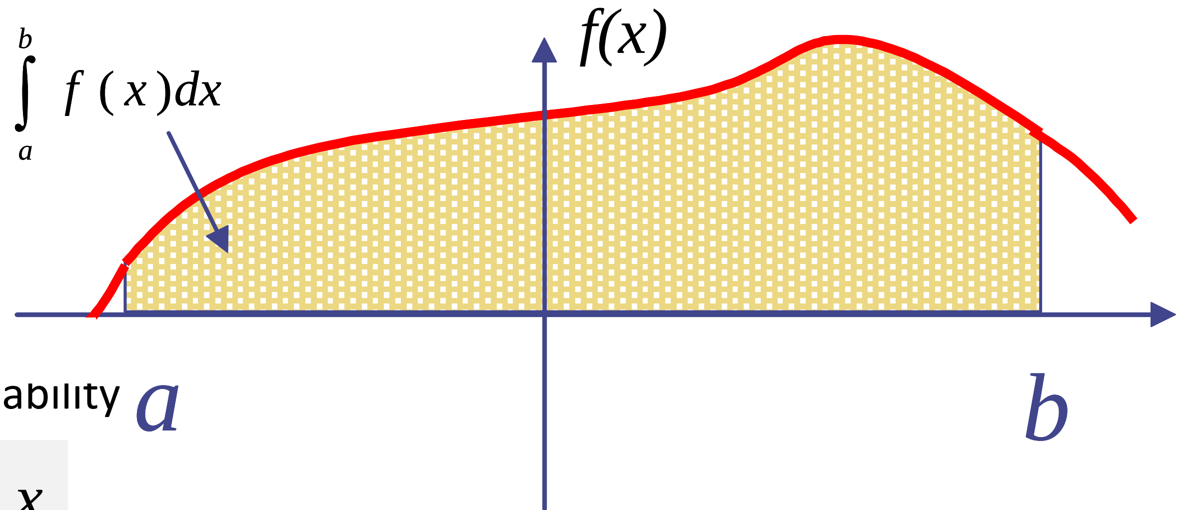
Continuous Probability Distributions

- Differences from Discrete Random Variables
  - Probability of specific value outcomes make no sense.
  - Probability of values within an interval is more helpful.
  - Cannot list all possible outcomes – instead we need to use a function.

- **Probability Density Function** (PDF)
  - Write as density function of random variable X: $f(x)$.
  - Rule: $f(x) \geq 0$ for all x.
  - Probability that X lies between values *a* and *b* is equal to the area under the curve between *a* and *b.*
  - Area under the curve sums to 1.

Continuous vs. Discrete Distributions

- Population parameters
  - Mean = μ
    - Expected Value of X or E(X)
  - Standard Deviation = σ
- Discrete
  - Multiply each value by its probability



$$\mu = E(X) = \sum_{i=1}^{n} p_i x_i$$

$$\sigma^2 = \sum_{i=1}^{n} p_i (x_i - \mu)^2$$

- Continuous
  - Requires integration to calculate μ and σ^2

$$\mu = \int_a^b t \times f(t)dt$$

$$\sigma^2 = \int_a^b (t-\mu)^2 \times f(t)dt$$

Integration across density function

- The area under the density function can be taken across whole range of the random variable, so mean and variance across the whole range of values are

$$\mu = \int_{-\infty}^{\infty} t\, f(t)\, dt, \qquad \sigma^2 = \int_{-\infty}^{\infty} (t-\mu)^2\, f(t)\, dt$$

- Notice that (by definition),

$$\int_{-\infty}^{\infty} f(t)\, dt = 1$$

- Relationship between density and probability:

$$F(x) = P(X < x) = \int_{-\infty}^{x} f(t)\, dt$$

The function F is called the cumulative probability distribution.

Note: You will not be required to perform any integration as part of assessment for this module.

Continuous Probability Distributions

- **Cumulative Distribution Function** (CDF)
  - $F(t) = P(X \leq t)$ or the probability that random variable X does not exceed t.
  - $F(t) = \int_{-\infty}^{t} f(x)\mathrm{d}x$
  - $0.0 \leq F(t) \leq 1.0$
  - $F(b) \geq F(a)$ if $b > a$ (increasing)

- Simple Rules
  - $P(X \leq t) = F(t)$
  - $P(X > t) = 1 - F(t)$
  - $P(c \leq X \leq d) = F(d) - F(c)$
  - $P(X = t) = 0$

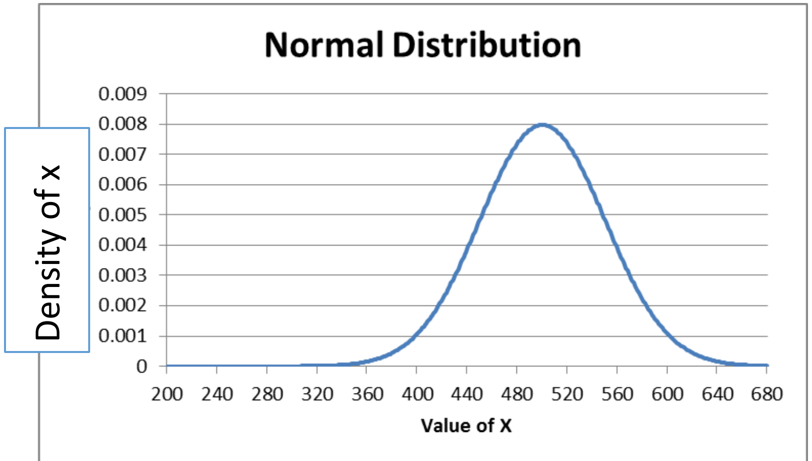
Continuous Probability Distributions

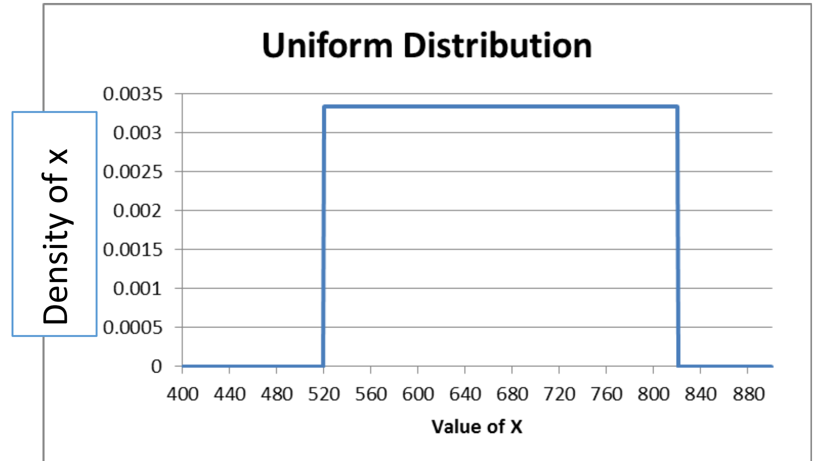
- Uniform distribution
- Normal distribution
- Many more…

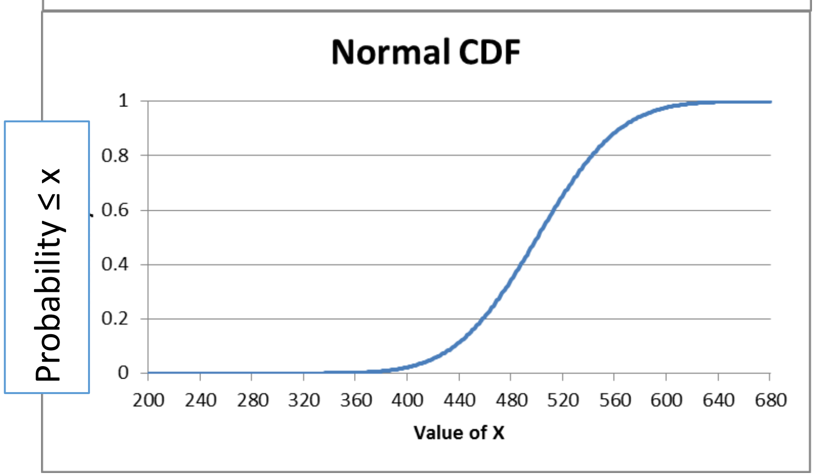
Parametric distributions:

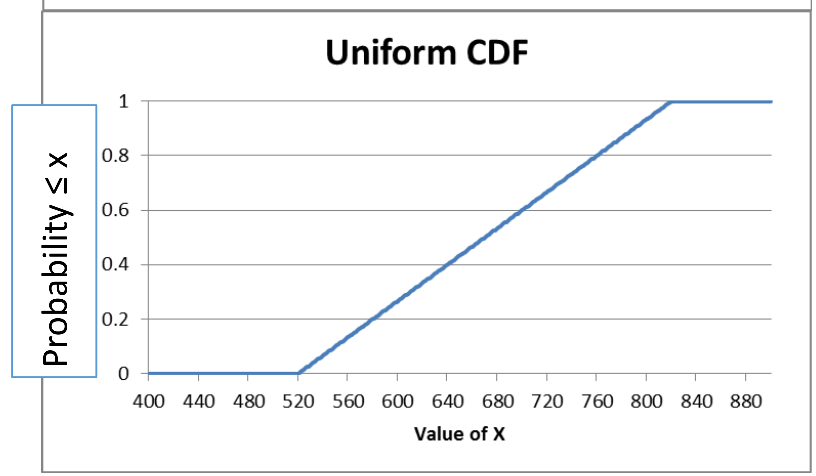
- The distribution is characterized by a fixed number of parameters.
- For example, with two parameters a, b we write $f(x \mid a, b)$ for the PDF.

Continuous Probability Functions









Uniform Distribution

- X is uniformly distributed over the range a to b, where $b > a$, or $X \sim U(a, b)$:

$$f(t \mid a,b) = \begin{cases} \dfrac{1}{b-a} & if \;\; a \le t \le b \\ 0 & otherwise \end{cases} \qquad F(t \mid a,b) = \begin{cases} 0 & if \;\; t < a \\ \dfrac{t-a}{b-a} & if \;\; a \le t \le b \\ 1 & if \;\; t > b \end{cases}$$

- Characteristics
  - Rectangular distribution with constant probability and implies the fact that each range of values that has the same length on the distributions support has equal probability of occurrence.
  - The density function integrates to unity.
  - Each of the inputs that go in to form the function have equal weighting.

Exercise 1

You want to generate a uniformly distributed random number between 1 and 10.

1. What the probability that it is between 3 and 5?

2. Supposing that $F(t \mid a, b) = 0.5$. What is the value of t in this case?

Exercise 1 Solutions

1. You want to generate a random number (Uniformly distributed) between 1 and 10. What the probability that it is between 3 and 5?

   In this case a=1 and b=10, then

$$P(3 \leq X \leq 5) = F(5) - F(3) = \frac{5-1}{10-1} - \frac{3-1}{10-1} = \frac{2}{9}$$

2. Supposing that $F(t \mid a, b) = 0.5$. What is the value of t in this case?

$$F(t \mid a, b) = \frac{t-a}{b-a} = 0.5 \quad \Rightarrow \quad t = \frac{a+b}{2} = \frac{11}{2}$$

Exercise 2

Consider the function for some parameter $a > 0$,

$$F(t \mid a) = \begin{cases} 0 & \text{if } t < 0 \\ \gamma t^2 & \text{if } 0 \le t \le a \\ 1 & \text{if } t > a \end{cases}$$

1. Show that F is a CDF when $\gamma = 1/a^2$.

2. Compute $F(a/2 \mid a)$.

3. Draw the shape of this CDF for $a = 5$.

4. What is $P(1 \le X \le 2)$? Write your answer as an expression.

Exercise 2 Solution

1. Show that F is a CDF when $\gamma = 1/a^2$.

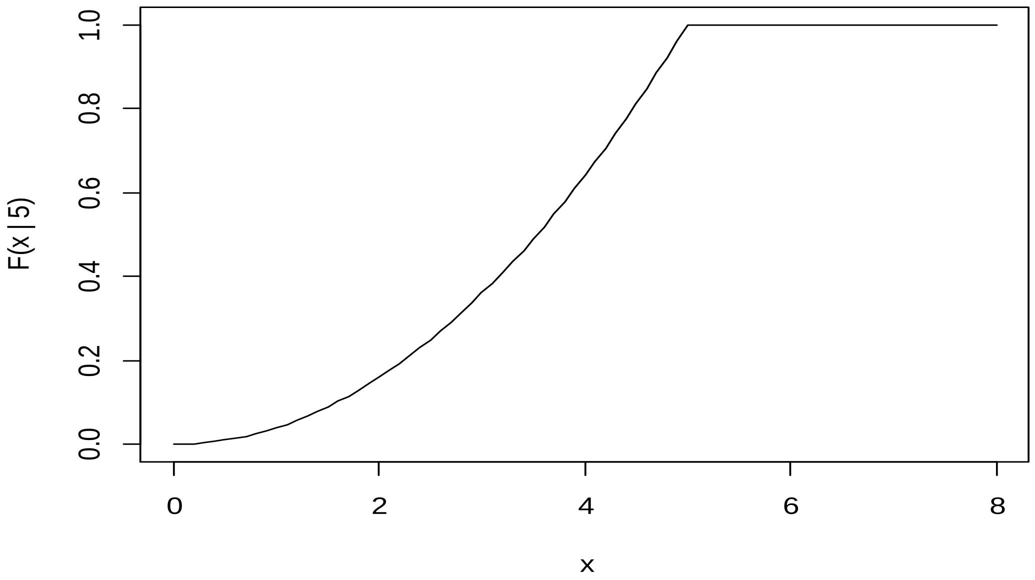
- Firstly, prove $0 \leq F(x \mid a) \leq 1$ for all x.
  - This is true for case $t < 0$ and $t > a$.
  - For case $0 \leq t \leq a, \gamma = 1/a^2$ means $0 \leq \gamma t^2 \leq 1$.
- Clearly, $F(-\infty \mid a)$=0 and $F(+\infty \mid a) = 1$
- For $s < t, F(s \mid a) \leq F(t \mid a)$: consider these cases:-

$s < t < 0$	$F(s \mid a) = F(t \mid a) = 0$
$s < 0, 0 \leq t \leq a$	$F(s \mid a) = 0 \leq \gamma t^2$
$0 \leq s < t \leq a$	$\gamma s^2 < \gamma t^2$ means $F(s \mid a) < F(t \mid a)$
$t > a$	$F(t \mid a) = 1$, therefore $F(s \mid a) \leq F(t \mid a)$, since $0 \leq F(s \mid a) \leq 1$ is already proved above.

Exercise 2 Solution

2. Compute $F(a/2 \mid a)$.
   $F(a/2 \mid a) = \gamma(a/2)^2 = 1/4$

3. Draw the shape of this CDF for $a = 5$.



Exercise 2 Solution

4. What is $P(1 \leq X \leq 2)$? Write your answer as an expression.

$$P(1 \leq X \leq 2) = F(2\,|a) - F(1\,|a)$$

and

$$F(2 \mid a) = \begin{cases} 4/a^2 & \text{if } 2 \leq a \\ 1 & \text{if } 2 > a \end{cases}$$

$$F(1 \mid a) = \begin{cases} 1/a^2 & \text{if } 1 \leq a \\ 1 & \text{if } 1 > a \end{cases}$$

So

$$P(1 \leq X \leq 2) = \begin{cases} 0 & \text{if } a < 1 \\ 1 - 1/a^2 & \text{if } 1 \leq a \leq 2 \\ 3/a^2 & \text{if } a > 2 \end{cases}$$

The Normal (Gaussian) Distribution

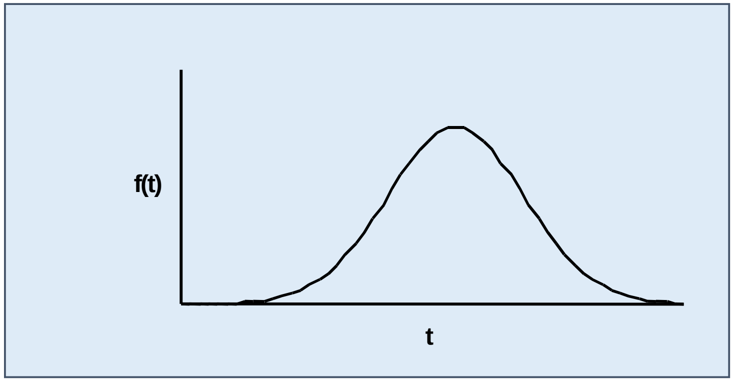
- X is normally distributed with mean μ and standard deviation σ, or X~N(μ, σ):

$$f(x \mid \mu, \sigma) = \frac{1}{(2\pi)^{1/2}\sigma} e^{\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]}$$

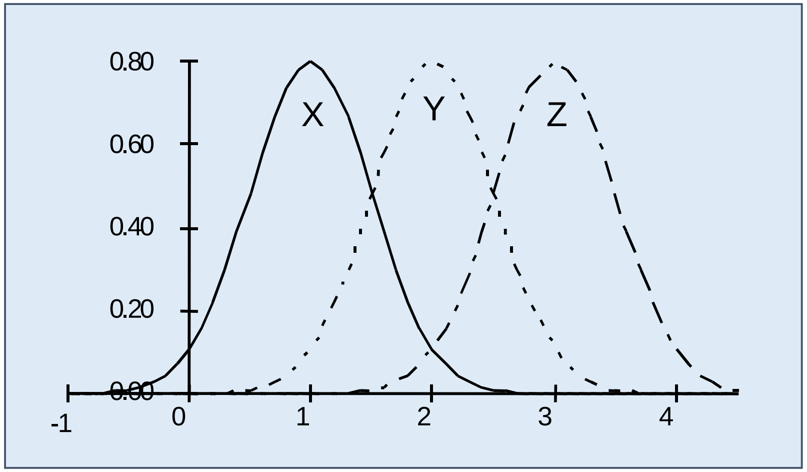
- Characteristics
  - Most commonly used distribution – many analysis assume ~ N
  - High point in “bell curve” occurs at mean.
  - Symmetric about the mean.
  - The mean “shifts” the distribution – but not the “shape”.
  - The standard deviation changes the “shape” but does not “shift” it.

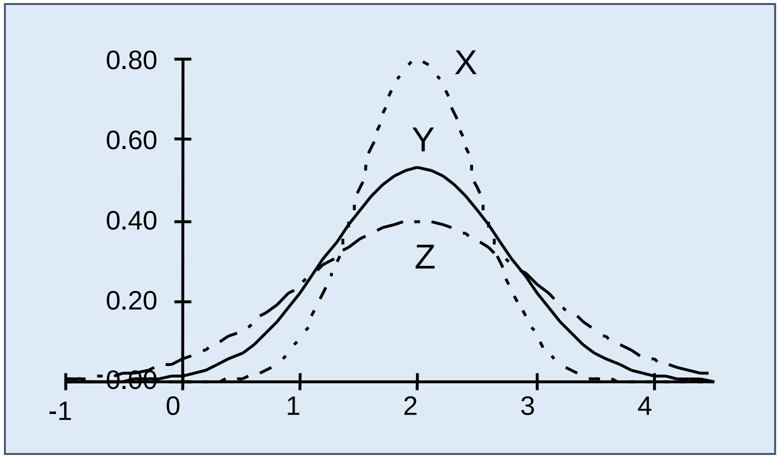
The Normal Distribution

- Density function is the familiar “bell-shaped” curve



- Completely described by mean μ and standard deviation s: N(μ,s)





Z scores and Standard Normal Distribution

- Z score given by $Z = (X - \mu)/\sigma$.

Then:

- $Z \sim N(0,1)$
- Allows for use of standard tables
- Area under the curve is 1
- Able to assess the probability of an event
- A z score can be positive or negative

Example

- Griffin Inc. ships products to an area where the distance traveled $\sim N(650, 100)$.

- You will need a calculator or standard tables to complete this:

  - What is the z-score for X = 575 miles?

  - What is the probability that distance >600?

  - What distance can I expect 50% of the shipments to be shorter than? What about 90%? 95%?

Example (Solutions)

- Griffin Inc. ships products to an area where the distance traveled $\sim N(650, 100)$.
  - What is the z-score for X = 575 miles?

$$Z = \frac{X - \mu}{\sigma} = \frac{575 - 650}{100} = -0.75$$

  - What is the probability that distance >600?

$$Z = \frac{600 - 650}{100} = -0.5$$

Look up on a scientific calculator,

$$P(Z > -0.5) = 1 - P(Z < -0.5) = 0.691$$

Example (Solutions)

- Griffin Inc. ships products to an area where the distance traveled $\sim N(650, 100)$.
  - What distance can I expect 50% of the shipments to be shorter than?

For standard normal,

$$P(Z < 0) = 0.5$$

So

$$\frac{X-\mu}{\sigma} = \frac{X-650}{100} < 0$$

Hence $X < 650$