Introduction to Statistical Modelling

STAT2507D

Chapter 6-1

The Normal Probability Distributions

DISTRIBUTIONS FOR CONTINUOUS RANDOM VARIABLES

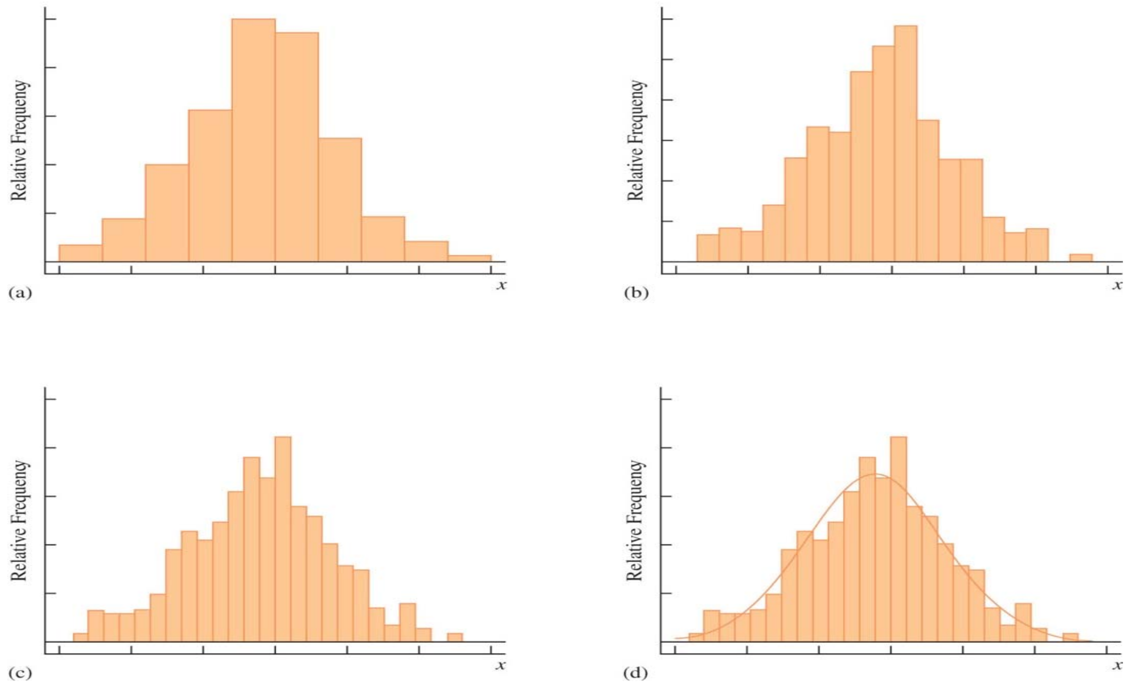
- Continuous random variables can assume the infinitely many values corresponding to points on a line interval
- Examples:
  - Heights, weights: 170.6cm, 50.7kg
  - Length of life of a light bulb : 105.6hrs

DISTRIBUTIONS FOR CONTINUOUS RANDOM VARIABLES

- For small number of measurements, a small number classes is sufficient.
- As the number of measurements becomes very large, number of classes need to be increased (Q 1.6) and the class widths become very narrow (more classes), the relative frequency histogram appears like smooth curve

Relative Frequency Histograms for Large Samples

- This smooth curve describes the probability distribution of the continuous random variable.

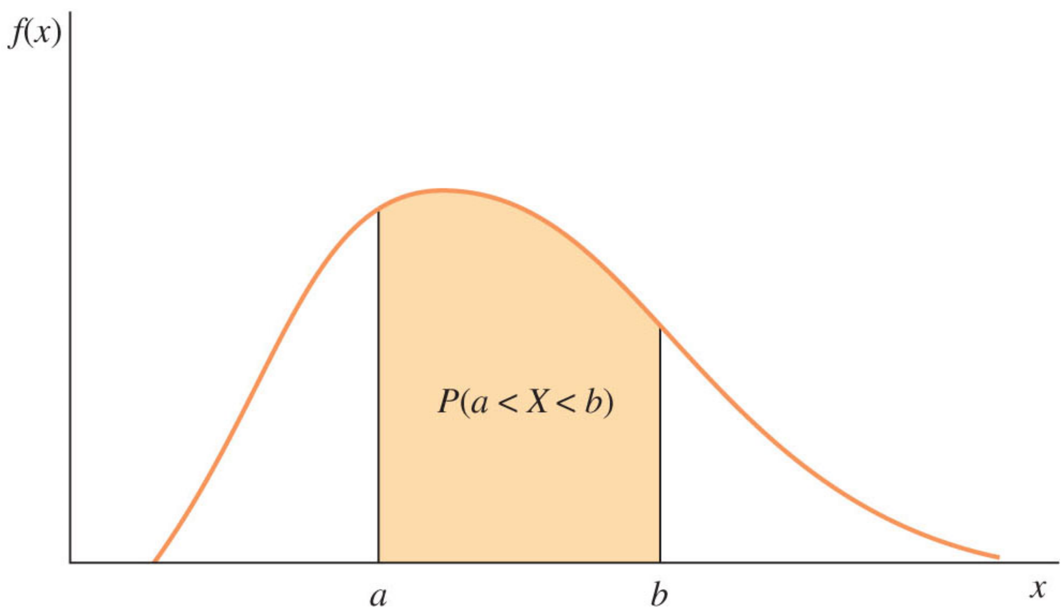


PROBABILITY DISTRIBUTIONS

- The depth or density of the probability, which varies with x, may be described by a mathematical formula $f(x)$, called the **probability distribution** or **probability density function** for the random variable x

Characteristics of Continuous Probability Distribution

- The area under a continuous probability distribution is equal to 1
- P(a < x < b) = area under the curve between a and b



PROPERTIES OF CONTINUOUS PROBABILITY DISTRIBUTION

- Since a probability for any continuous random variable is an area under the probability density function, there is no probability attached to any single value of x; i.e.

$P(x = a) = 0$ is always valid.

PROPERTIES OF CONTINUOUS PROBABILITY DISTRIBUTION

- This implies that
  - $P(x \leq a) = P (x < a)$
  - $P (x \geq a) = P(x > a)$;
  - $P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$.
  - This is not true in general for discrete random variables.

PROPERTIES OF CONTINUOUS PROBABILITY DISTRIBUTION

- $f(x) \geq 0$, for any value of x
- $P(a \leq X \leq b)$ is given by the area under the probability curve between a and b
- The total area under the curve of f(x) is equal to 1

PROBABILITY DISTRIBUTIONS

- There are many different types of continuous random variables
- We try to pick a model that
  - Fits the data well
  - Allows us to make the best possible inferences using the data

UNIFORM DISTRIBUTIONS

- The values of a uniform random variable are evenly distributed over a given interval
- The formula that generates the uniform distribution

$$f(x) = \frac{1}{(b-a)}, a \le x \ \le b$$

Uniform Distributions

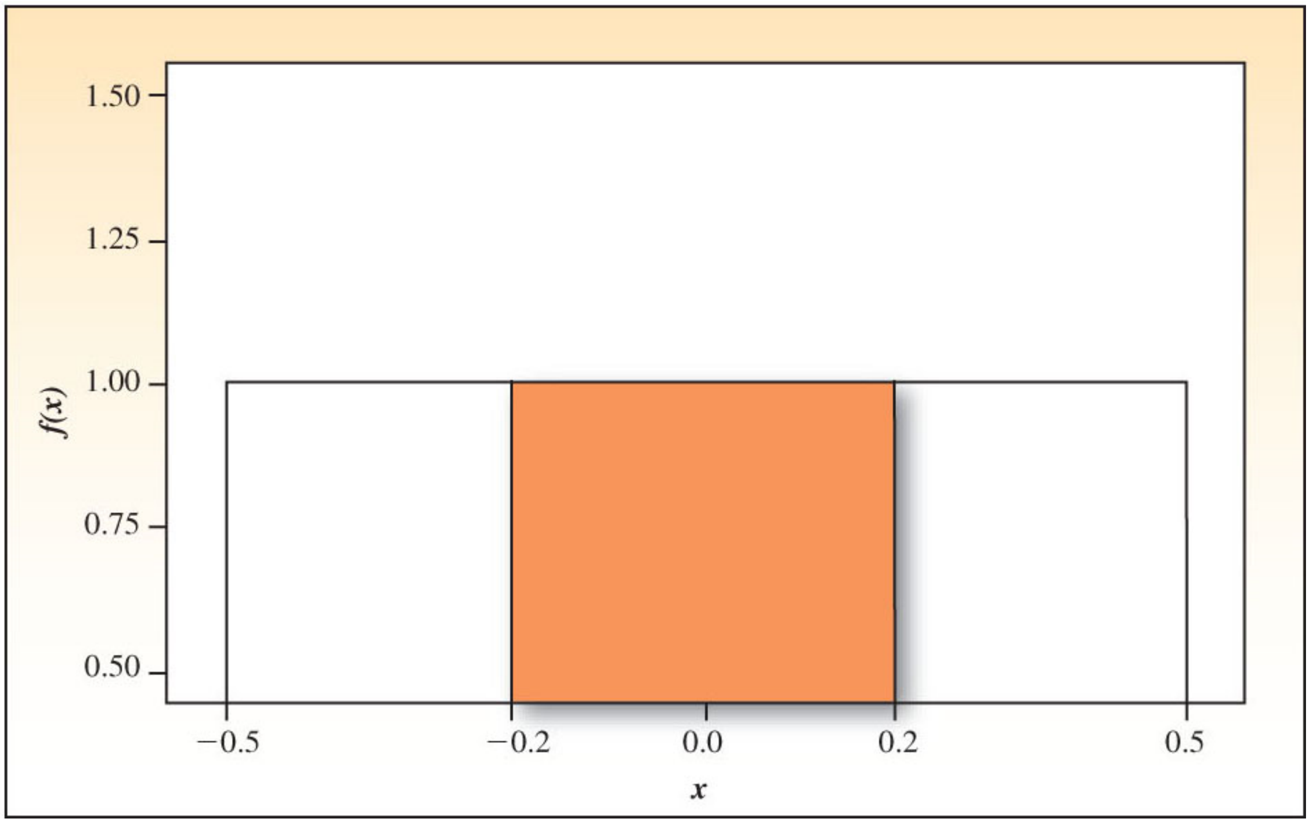
- If X is a uniform random variable , then it is denoted by $X \sim U(a,b)$
  - $E(X) = \mu_x = \frac{a+b}{2}$
  - $Var(X) = \sigma_x^2 = \frac{(b-a)^2}{12}$

EXERCISE

6.1) The error x introduced by rounding an observation to the nearest centimetre has a uniform distribution over the interval [-0.5, 0.5]. What is the probability that the rounding error is less than 0.2?

- $P(-0.2 < x < 0.2) = [0.2 - (-0.2)] \times 1 = 0.4$

UNIFORM DISTRIBUTIONS, CONT'D



EXERCISE

6.2) The wait time for a specific bus, in minutes, is uniformly distributed between 0 and 15 minutes

a. Write the formula for the probability density function and sketch it
b. Find the probability that X= 4
c. What is the probability that the wait time is more than 10 minutes?
d. What is the probability that the wait time is between 5 and 10 minutes?

EXPONENTIAL DISTRIBUTION

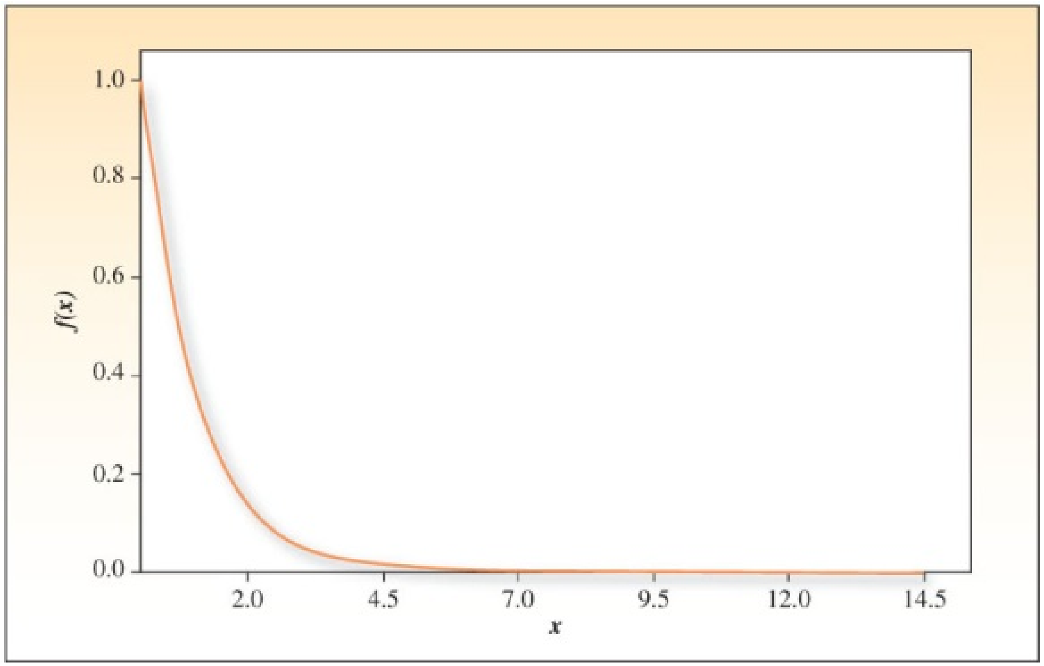
- The exponential random variable is used to model continuous random variables such as waiting times or lifetimes. Exponential probability distribution is given by

$$f(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}},\ 0 \le x \le \infty,\ \lambda > 0$$

- Parameter μ is the mean of the exponential distribution. $E(X) = \mu_x = \lambda; Var(X) = \lambda^2$

EXPONENTIAL DISTRIBUTION

- When $\lambda=1$, the area under the curve is as shown



EXPONENTIAL DISTRIBUTION

- Probabilities of a given interval can be easily estimated
  - $P(x < a) = 1 - e^{-\frac{a}{\lambda}}; P(x > a) = e^{-\frac{a}{\lambda}}$
  - $P(x < b) = 1 - e^{-\frac{b}{\lambda}}$
  - $P(a < x < b) \; = \; P(x \; < b) - P(x < a) = e^{-\frac{a}{\lambda}} - e^{-\frac{b}{\lambda}}$

EXERCISE

6.3) Suppose the magnitude of earthquakes in a region of Indonesia can be modelled by exponential with mean of 4. Find the probability that the next earthquake to hit this region will exceed 4 on a Richer scale. Assuming that earthquakes are independent, what is the probability that out of the next seven earthquakes at least one will exceed 4.0 on the Richer scale?

MEMORY LESS PROPERTY OF EXPONENTIAL DIST'N

6.4) Suppose X has an exponential probability density function with mean λ.

a. Show that $P(X > a + b \mid X > a) = P(X > b)$, $a>0$ and $b>0$.
b. If $X \sim \exp(4)$, $a = 20$ and $b = 15$, then what is $P(X > 35 \mid X > 20)$ using part (a)?