PROBABILIY AND STATISTICS I

LECTURE TWELVE

Continuous random variables

Lecturer: Dr. Emily Roche

INTRODUCTION

This lecture will focus on some special continuous probability distributions.

Intended learning outcomes

At the end of this lecture, you will be able to explain properties of uniform, exponential and normal distributions and apply uniform, exponential and normal distributions to solve real life problems.

References

These lecture notes should be supplemented with relevant topics from the book listed in the Bibliography at the end of the lecture.

Continuous probability distributions

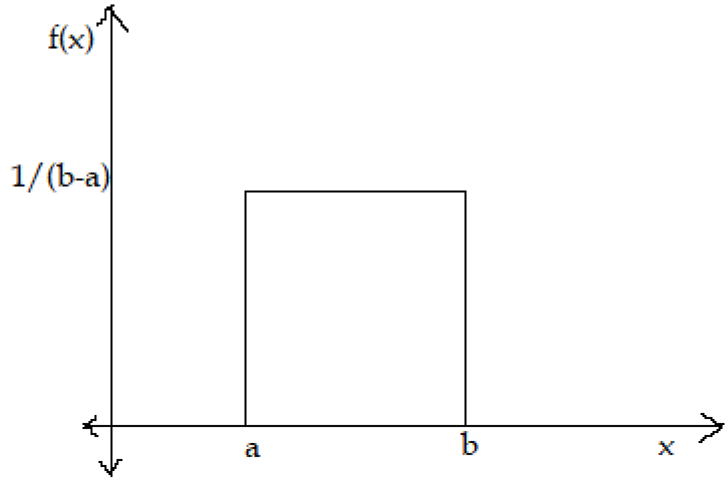
i. **Uniform Distribution.**

This is the simplest continuous probability distribution. It is rectangular in shape and defined by minimum and maximum values. The probability of the random variable is equal for all the values of the random variables.

Probability density function for the uniform distribution is:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & if \quad a \leq x \leq b \\ 0 & otherwise \end{cases}$$

It can be presented graphically as:



The area under the curve equals to one.

The Cumulative Distribution Function (cdf) is:

$$F(x) = P(X \leq x) = \frac{x-a}{b-a}$$

Expectation of the Uniform distribution is obtained as:

$$E(X) = \int_a^b xf(x)\,dx = \int_a^b x\left(\frac{1}{b-a}\right) = \left.\frac{x^2}{2(b-a)}\right|_a^b = \frac{b^2}{2(b-a)} - \frac{a^2}{2(b-a)}$$

$$= \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2} \Rightarrow \text{mid point of } [\text{a, b}]$$

While its variance is given as:

$$Var(X) = E[X - E(X)]^2 = \int_a^b \left(X - \frac{b+a}{2}\right)^2 \left(\frac{1}{b-a}\right) dx = \frac{(b-a)^2}{12}$$

Exercise

Suppose that a research department of a steel manufacturer believes that one of the company's rolling machine is producing sheets of varying thickness. The thickness is uniformly distributed with the values between $150mm$ and $200mm$. Any sheet that has less than $150mm$ must be scrapped because it is unacceptable to buyers.

a. What is the probability that a sheet selected at random will be unacceptable?
b. Calculate and interpret the mean and standard deviation of X.

Application:

It is used in situations where by you imagine that no value with a specific range is more likely than others. In most situations it is used where you know the worst and the best range but don't want to make assumption within the range.

ii. Exponential Distribution:

Example

If an engineer is responsible for the quality of copper wire for use in domestic wiring systems, he or she might be interested in knowing both the number of faults in a given length of wire and also the distance between such faults. While the number of faults may be analysed by using the Poisson distribution, the distances between faults along the wire give rise to the **Exponential Distribution**.

Examples of Exponential random variables:

1. You have observed the number of hits to your web site at a rate of maybe 2 per day. This follows a Poisson distribution.
   The time (X), in days, between hits follow an Exponential distribution.
2. You have observed the number of calls that arrive each day over a period of one year, and note that the arrivals follow a Poisson distribution at a rate of 3 per day.
   The waiting time (X) between calls follow an Exponential distribution.
3. Records show that job submissions to a busy computer center have a Poisson distribution with an average of 4 per minute.
   The time X in minutes between submissions follow an Exponential distribution.
4. Records indicate that messages arrive to a computer server following a Poisson distribution at the rate of 6 per hour.
   The time X (hours) that elapse between messages follow an Exponential distribution.

In summary the Exponential distribution describes waiting time between Poisson occurrences. In practice this distribution is used for time periods between observing an event for a Poisson occurrence, in other words, waiting time until the first event in a Poisson process with rate λ. The events are systematic or approximately regular.

For example: Flight or Train timetables, the arrival of waves on a beach, Radioactive decay, the occurrences of a rare disease in a large population, arrival of a packet of information on the internet, times between successive job arrivals at a computing center.

Therefore, a random variable X with pdf

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & otherwise \end{cases}$$

with parameter $\lambda \geq 0$, being a constant (rate) is said to be Exponentially distributed.

The cdf of X is given by:

$$\Pr[X \leq x] = F(x) = \int_0^x f(x)\, dx$$

$$= \int_0^x \lambda\, e^{-\lambda x} dx = 1 - e^{-\lambda x}$$

$$\therefore \Pr[X \leq x] = F(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Expectation of the Exponential distribution:

$$E(X) = \int_0^\infty x\, f(x) dx = \int_0^\infty x\, \lambda e^{-\lambda x} dx = \frac{1}{\lambda} \text{ (use integration by parts to verify)}$$

Variance of the Exponential distribution:

$$Var\ (X) = E[\,X - E(X)]^2 = E\left[X - \frac{1}{\lambda}\right]^2$$

$$= \int_0^\infty \left(x - \frac{1}{\lambda}\right)^2 f(x) dx = \int_0^\infty \left(x - \frac{1}{\lambda}\right)^2 \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2}$$

Illustration:

Let X be the interval between births at a county hospital for which the average time between births is seven days. Assuming the distribution of the time between births follows an exponential distribution. Multiple births may violate the assumption of independence and therefore a "birth" is defined to be a birth for one mother regardless of the number of babies born. The unit of time is day and the corresponding average rate of events is one birth every seven days;

So $\lambda = \frac{1}{7}$

The probability distribution function therefore is:

$$f(x) = \begin{cases} \frac{1}{7} e^{-\frac{1}{7}x} & x \geq 0 \\ 0 & otherwise \end{cases}$$

From the above illustration, what is the chance that there is a birth in the next 10 days? 10 hours? 10 minutes?

$$\Pr[X < x] = 1 - e^{-\lambda x} \qquad x > 0$$

$$= 1 - e^{-\frac{1}{7}x}$$

(i) $\Pr[X \leq 10] = 1 - e^{-\frac{10}{7}} = 0.760$

(ii) $10\ hours = \frac{5}{12}\ days$

$$\Pr\left[x \leq \frac{5}{12}\right] = 1 - e^{\left(-\frac{1}{7} \times \frac{5}{12}\right)} = 0.058$$

(iii) $10\ minutes = \frac{1}{144}\ days$

$$\Pr\left[X \leq \frac{1}{144}\right] = 0.00099$$

iii. Normal distribution.

This is the most popular and commonly used distribution. It is also called the Gauss distribution after its German discoverer. The normal distribution is a probability distribution which is used to determine probabilities of continuous variables

Examples of continuous variables are:

Distances	Weights	Capacity e.t.c
Times	Heights	

Usually continuous variables are those, which can be measured by using the appropriate units of measurement.

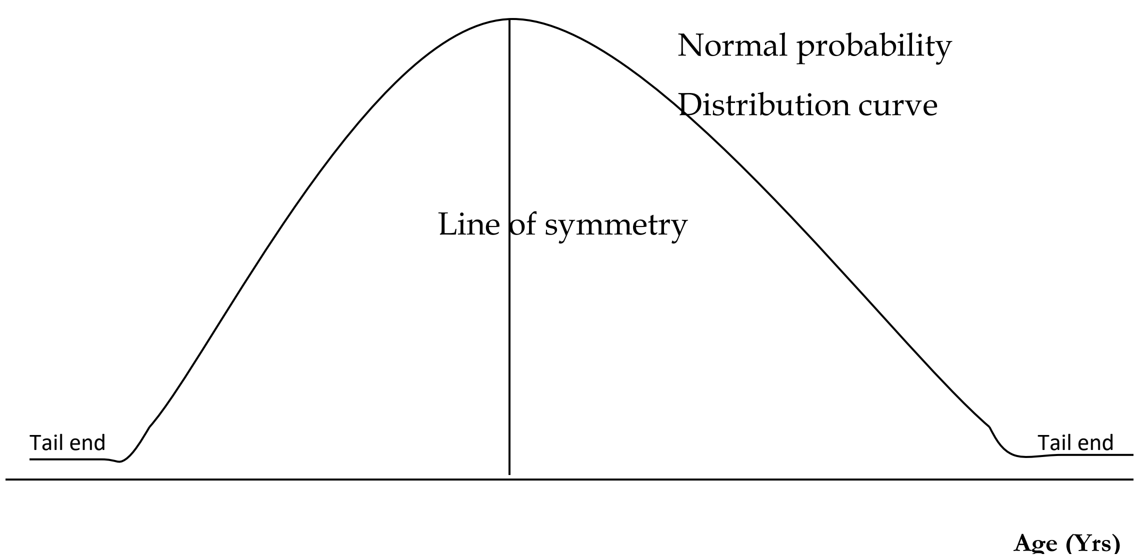
A random variable X is said to follow the normal distribution iff its pdf is:

$$f(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} & -\infty < x < \infty \\ 0 & otherwise \end{cases}$$

If X is normally distributed, then it's written as $X \sim N(\mu, \sigma^2)$ read as X is nornally distributed with mean μ and variance σ^2.

The properties of the normal distribution are:

1. The total area under the curve is = 1 which is equivalent to the maximum value of probability



2. The line of symmetry divides the curve into two equal halves

3. The two ends of the normal distribution curve continuously approach the horizontal axis but they never cross it
4. The values of the mean, mode and median are all equal

Note:

The above distribution curve is referred to as normal probability distribution curve because if a frequency distribution curve is plotted from measurements of a given sample drawn from a normal population then a graph similar to the normal curve must be obtained.

It should be noted that 68% of any population lies within one standard deviation, $\pm 1\sigma$

95% lies within two standard deviations $\pm 2\sigma$

99% lies within three standard deviations $\pm 3\sigma$

Where σ = standard deviation

5. In case $\mu = 0$ and $\sigma = 1$, then the density function for X becomes:

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} & -\infty < x < \infty \\ 0 & otherwise \end{cases}$$

In this situation the distribution of X is said to have a standard normal distribution.

6. In practice, in order to easily compute probabilities involving the normal variate, the random variable has to be standardized first that is, to be made to have the normal standard density. Notation wise,

$$X \sim N(0,1)$$

Standardization of variables

This is done using the formula; $Z = \frac{X-\mu}{\sigma}$. Where

$X =$ Value to be standardized

Z = Standardization of X

μ = population mean

σ = Standard deviation

Computation of probabilities of the normal variate when plotted takes the familiar bell shape

The standard normal probabilities have been evaluated and tabled.

Example

A sample of students had a mean age of 35 years with a standard deviation of 5 years. A student was randomly picked from a group of 200 students. Find the probability that the age of the student turned out to be as follows

Lying between 35 and 40

The standardized value for 35 years

$$Z = \frac{X - \mu}{\sigma} = \frac{35 - 35}{5} = 0$$

The standardized value for 40 years

$$Z = \frac{X - \mu}{\sigma} = \frac{40 - 35}{5} = 1$$

∴ the area between $Z = 0$ and $Z = 1$ is 0.3413 (These values are checked from the normal tables see attached)

The value from standard normal curve tables.

When $Z = 0, p = 0.50000$ and when $Z = 1, p = 0.84134$

Now the area under this curve is the area between $Z = 1$ and $Z = 0$

$$= 0.84134 - 0.50000 = 0.34134$$

∴ the probability age lying between 35 and 40 yrs is 0.34134

Lying beyond 45 yrs - P(beyond 45 years) is determined as $P(x > 45)$

$$Z = \frac{X - \mu}{\sigma} = \frac{45 - 35}{5} = 2$$

Probability corresponding to $P(Z = 2) = 0.97725$

$$\therefore P(Age > 45yrs) = 1.00000 - 0.97725 = 0.02275$$

<u>Bibliography</u>

Gupta, SP (Dr.), (2014). *Statistical methods* (43rd Ed.). Sultan Chand & Sons.

S. C. Gupta and V. K. Kapoor, (2020). *Fundamentals of mathematical Statistics* (12th Ed). Sultan Chand & Sons.