ST221 Introduction to Statistics

Dr Rafael Moral

Dept of Mathematics and Statistics
Maynooth University

rafael.deandrademoral@mu.ie

3 Continuous Random Variables

A continuous random variable takes values anywhere in an interval on the real line, e.g. $(-\infty, \infty)$, $(0, \infty)$, $(5, 25)$ etc. Examples:

- Length of a caterpillar in mm
- Distance travelled in km in a training session
- Weight of a book in grams.

3.1 The probability density function and cumulative distribution function

For a continuous random variable, the **probability density function (pdf)** $f(x)$ has the following properties:

1. $f(x) \geq 0$, but we don't require $f(x) \leq 1$.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.
3. $P(a \leq X \leq b) = \int_a^b f(x)dx$.

Notes:

- $f(x) \neq P(X = x)$
- $P(X = a) = \int_a^a f(x)dx = 0$.

For a continuous random variable, the **cumulative distribution function (cdf)** $F(x)$ is:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du.$$

Graphically:

Notes:

- $F(-\infty) = 0, F(\infty) = 1$
- $f(x) = \dfrac{dF(x)}{dx}$.

Notes on pdf and cdf for continuous random variables

- If X takes values in a continuous set, e.g. $(0, \infty), (0, 1)$ etc, then $P(X = x) = 0$ for all x, since there are infinitely many x's in a continuous set.

- We use the cdf $F(x) = P(X \leq x)$ to work out probabilities.

  E.g. $P(a < X < b) = P(a \leq X \leq b)$
  $= P(X \leq b) - P(X \leq a) = F(b) - F(a)$.

  Note: the endpoints are not very important for continuous random variables but are very important for discrete random variables.

- The pdf is not a probability but it is very useful as it gives the shape of the distribution of X.

Example

Suppose I choose a number randomly from [0,1].

Let X be the resulting number, then $X \sim$ Uniform [0,1] and

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{if } x > 1 \end{cases}$$

Example contd.

Suppose I choose a number randomly from [0,1].

Let X be the resulting number, then $X \sim$ Uniform [0,1] and

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}$$

Example

Suppose I choose a number randomly from [a,b].

Let X be the resulting number, then $X \sim$ Uniform [a,b] and

$$f(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{if } x > b \end{cases}$$

Example contd.

Suppose I choose a number randomly from [a,b].

Let X be the resulting number, then $X \sim$ Uniform [a,b] and

$$F(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \leq x \leq b \\ 1 & \text{if } x > b \end{cases}$$

Example

Suppose $F(x) = x^2$ for $0 \leq x \leq 1$, i.e.

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x^2 & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}$$

Find f from F and graph both f and F.

Definition

If X is a continuous random variable, then

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

and

$$Var(X) = E[X^2] - (E[X])^2$$

Note that

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

i.e.

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)dx$$

Example

Suppose $X \sim$ Uniform $(0,1)$. Find $\mathrm{E}[X]$ and $\mathrm{Var}(X)$.

For the remainder of this section we will consider 'special' continuous distributions. We have already seen the uniform distribution. We will also consider:

- The exponential distribution
- The normal distribution.

3.2 The exponential distribution

The exponential distribution is commonly used to model, for example,

- the lifetime of components
- time between earthquakes, arrivals at an ATM, etc.

Example

Let Y be the number of earthquakes in a year and assume that we can model Y with a Poisson distribution, i.e. $Y \sim \text{Poisson}(\lambda)$.

In this case, $\text{E}[Y]=\lambda$ (the mean number of earthquakes in a year).

We could also consider X=time between earthquakes.
Then $X \sim \exp(\lambda)$.

Properties of the exponential distribution

Assume that $X \sim \exp(\lambda)$.

Probability density function

$$f(x) = \lambda e^{-\lambda x} \qquad \text{with } x \geq 0$$

Cumulative distribution function

$$\begin{aligned} F(x) &= P(X \leq x) = \int_0^x \lambda e^{-\lambda u} du \\ &= -e^{-\lambda u}\Big|_0^x = -e^{-\lambda x} - (-e^{-\lambda 0}) \\ &= 1 - e^{-\lambda x} \end{aligned}$$

Properties of the exponential distribution contd.

$$E[X] = \int_0^\infty x\lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

$$Var(X) = \frac{1}{\lambda^2}$$

Example

Let X = time in minutes between arrivals at a supermarket checkout. On average there are 2 arrivals per minute. We can model this as $X \sim$ exponential($\lambda = 2$).

1. What is the probability that the time between arrivals is ≤ 1 minute?
2. What is the probability that the time between arrivals is in $[1, 2]$ minutes?
3. What is the probability that no customers arrive in a 3 minute period?
4. What is the mean time between arrivals?

Example

Suppose that the length of a phone call in minutes is an exponential random variable with parameter $\lambda = \frac{1}{10}$.

If someone arrives immediately ahead of you at a public telephone booth, find the probability that you will have to wait

1. more than 10 minutes,

2. between 10 and 20 minutes.

3.3 The normal distribution

This continuous distribution is one of the most commonly used distributions in statistics. It is used to model, for example,

- experimental errors in scientific measurements
- test scores in aptitude tests
- heights of people selected at random from a population.

Properties of the normal distribution

Probability density function

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

with $-\infty < x < \infty$.

- The shape of the function is:

- The parameters μ and σ are the mean and standard deviation respectively. $\mathrm{E}[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$.

Properties of the normal distribution contd.

- We write $X \sim \mathrm{N}(\mu, \sigma^2)$.
- The shape of the normal pdf is described as bell shaped. It is symmetric about μ (half of the area under the curve is above and half below μ).
- Approximately 68% of the area under the curves lies in $(\mu - \sigma,\ \mu + \sigma)$. Approximately 95% of the area under the curves lies in $(\mu - 2\sigma,\ \mu + 2\sigma)$. Approximately 99.7% of the area under the curves lies in $(\mu - 3\sigma,\ \mu + 3\sigma)$. This is known as the 68-95-99.7 rule. Graphically it is represented by:

Properties of the normal distribution contd.

- Probabilities $P(X \leq x)$ cannot be obtained as a closed form expression and the integral must be evaluated numerically.
- Normal probability tables contain tabulated probabilities for the standard normal distribution.
  I.e. for $Z \sim N(\mu = 0, \sigma^2 = 1)$.
- Theorem: If $X \sim N(\mu, \sigma^2)$, then $Z = \dfrac{x - \mu}{\sigma} \sim N(0,\ 1)$.
  As a result of this theorem, probabilities for any normal distribution can be obtained from the standard normal tables.
- The standard normal tables are in the Departmental Statistical Tables.

Example

Let $Z \sim \mathrm{N}(0, 1)$. Find:

1. $\mathrm{P}(0 < Z < 1.14)$.

2. $\mathrm{P}(Z < 1.14)$.

Example contd.

Let $Z \sim \mathrm{N}(0, 1)$. Find:

3. $\mathrm{P}(-1.14 < Z < 0)$.

4. $\mathrm{P}(|Z| < 1.14)$.

5. $\mathrm{P}(|Z| > 1.14)$.

Example

Suppose test scores for a group of students for some exam are normally distributed with mean 50 and standard deviation 10.

What is the probability of a randomly selected student having a score greater than 80?

3.4 Independent random variables

Definition: Two discrete random variables X and Y are said to be independent if their joint probability mass function factors into their individual probability mass functions.

$$\mathrm{P}(X = x, Y = y) = \mathrm{P}(X = x)\mathrm{P}(Y = y) \qquad \forall\ x, y.$$

Definition: Two jointly continuous random variables are said to be independent if their joint probability density function factors into their individual probability density functions.

$$\mathrm{f}(x, y) = \mathrm{f}(x)\mathrm{f}(y)$$

Example

Let

X = result of throwing a die.
Y = result of throwing the die again.

X and Y are independent.

Let

U = sum of face values $X + Y$.
V = difference $X - Y$.

U and V are dependent because information abut U provides information about V.

We can extend the definition of independence to n random variables.