Edexcel International A Level (IAL) Maths: Statistics 2

Your notes

Cumulative Distribution Function

Contents

Cumulative Distribution Function

What is the cumulative distribution function (c.d.f.)?

- For a **continuous random variable**, X, with **probability density function f(x)** the **cumulative distribution function** (c.d.f.) is defined as

$$\mathrm{F}(x_0) = P(X \le x_0) = \int_{-\infty}^{x_0} \mathrm{f}(t)\ \mathrm{d}t$$

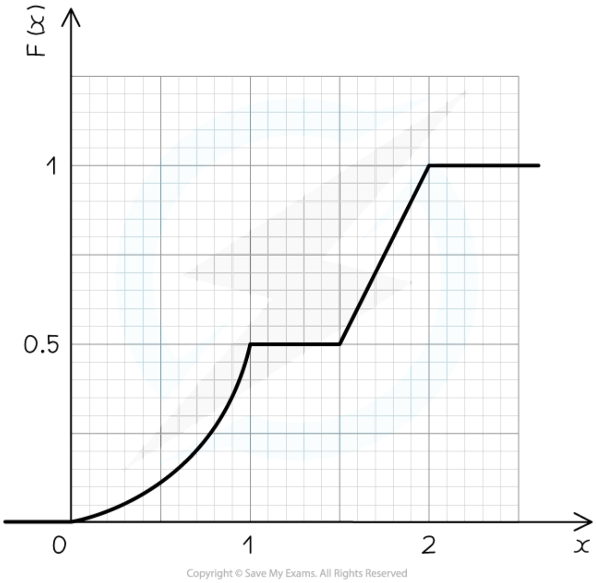
- Compare this to the cumulative distribution function for a **discrete random variable**

$$\mathrm{F}(x_0) = P(X \le x_0) = \sum_{x \le x_0} \mathrm{P}(X = x)$$

- $\mathrm{F}(x_0)$ is the probability that X is a value less than or equal to x_0
- Notice the use of **uppercase** F for the **c.d.f.** but **lowercase** f for the **p.d.f.**
- On the graph of the p.d.f. $y = \mathrm{f}(x)$ this would be the area under the graph up to the (vertical) line $x = x_0$
- F(x) should be defined for all values of $x \in \mathbb{R}$
- The graph of the c.d.f. y = F(x) will
  - start on the x-axis (i.e. start at a probability of 0)
  - end at x = 1 (i.e. finish at a probability of 1)
  - will be **continuous** function, even when defined **piecewise**

e.g.

$$\mathrm{F}(x) = \begin{cases} 0 & x < 0 \\ 0.5x^2 & 0 \le x \le 1 \\ 0.5 & 1 \le x \le 1.5 \\ x - 1 & 1.5 \le x \le 2 \\ 1 & x > 2 \end{cases}$$



Your notes

- The horizontal lines at F(x) = 0 and F(x) = 1 may not always be shown

How do I find probabilities using the cumulative frequency distribution?

- $\mathrm{P}(a \le X \le b) = \mathrm{F}(b) - \mathrm{F}(a)$
- Although $\mathrm{P}(X = k)$, for all values of k , F(k) is not necessarily zero

How do I find the cumulative frequency distribution (c.d.f.) from the probability density function (p.d.f.) and vice versa?

- To find the c.d.f.,F(x) , from the p.d.f.,f(x), **integrate**

$$\mathrm{F}(x) = \int_{-\infty}^{x} \mathrm{f}(t)\ dt$$

- Ensure you define F(x) fully for $x \in \mathbb{R}$ so include values of x for which F(x) = 0 and values of x for which F(x) = 1
- For piecewise functions as well as integrating you will need to add on the value of the c.d.f. at the end of the previous part
  - Suppose there are two sections to a p.d.f. $x \le a$ and $x > a$
  - For $x > a$:

$$\mathrm{F}(x) = \int_{-\infty}^{x} \mathrm{f}(t)\ \mathrm{d}t = \int_{-\infty}^{a} \mathrm{f}(t)\ \mathrm{d}t + \int_{a}^{x} \mathrm{f}(t)\ \mathrm{d}t = \mathrm{F}(a) + \int_{a}^{x} \mathrm{f}(t)\ \mathrm{d}t$$

- Therefore the c.d.f can be calculated for the interval $a < x < b$ by using

$$\mathrm{F}(x) = \mathrm{F}(a) + \int_a^x \mathrm{f}(t)\,\mathrm{d}t$$

- See part (b) in the Worked Example below
- To find the p.d.f from the c.d.f., **differentiate**

$$\mathrm{f}(x) = \frac{d}{dx}\mathrm{F}(x)$$

- Any part of a c.d.f that is constant corresponds to the p.d.f. for that part being zero (the derivative of a constant is zero)

How do I find the median, quartiles and percentiles using the cumulative frequency distribution (c.d.f.)?

- For piecewise functions, first identify the section the required value lies in
  - To do this find the upper limit of each section of the c.d.f.
- To find the **median**, m, solve the equation F(m) = 0.5
  - The median is sometimes referred to as the second quartile, Q_2
- To find the **lower quartile**, Q_1, solve the equation $F(Q_1) = 0.25$
- To find the **upper quartile**, Q_3 , solve the equation $F(Q_3) = 0.75$
- To find the **n^th^ percentile**, solve the equation $\mathrm{F}(p) = \frac{n}{100}$

Worked Example

a) The continuous random variable, X , has cumulative distribution function

$$\mathrm{F}(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{4}x(4-x) & 0 \le x \le 2 \\ 1 & x > 2 \end{cases}$$

Find

(i) $\mathrm{P}(X > 1.5)$

(ii) $\mathrm{P}(0.5 \le X \le 1)$

(iii) The lower quartile of X.

(b) The continuous random variable, X, has probability density function

$$f(x) = \begin{cases} 0.5x & 0 \le x \le 1 \\ 0.5 & 1 \le x \le 2.5 \\ 0 & \text{otherwise} \end{cases}$$

Find the cumulative frequency distribution, $\mathrm{F}(x)$.

Answer:

a) (i) $P(X > 1.5) = 1 - P(X \leqslant 1.5)$

$= 1 - F(1.5)$

$= 1 - \left[\frac{1}{4}(1.5)(4 - 1.5)\right]$

$\therefore P(X > 1.5) = \frac{1}{16}$

(ii) $P(0.5 \leqslant X \leqslant 1) = P(X \leqslant 1) - P(X \leqslant 0.5)$

$= F(1) - F(0.5)$

$= \frac{1}{4}(1)(4 - 1) - \frac{1}{4}(0.5)(4 - 0.5)$

$\therefore P(0.5 \leqslant X \leqslant 1) = \frac{5}{16}$

(iii) For the lower quartile solve "$F(Q_1) = 0.25$"

$\frac{1}{4} Q_1 (4 - Q_1) = 0.25$

$Q_1^2 - 4Q_1 + 1 = 0$

$Q_1 = 2 + \sqrt{3}$ or $2 - \sqrt{3}$

Use formula or calculator to solve

reject since $0 \leqslant x \leqslant 2$

$\therefore$ The lower quartile, Q_1, is $2 - \sqrt{3}$

b) P.d.f. to c.d.f. integrate but include previous parts.

for $x < 0$, $F(x) = 0$ and for $x > 2.5$, $F(x) = 1$

for $0 \leqslant x \leqslant 1$ $\quad F(x) = 0.5\int_0^x t \, dt$

$$= 0.5\left[0.5t^2\right]_0^x$$

$$= 0.5\left[0.5x^2 - 0\right] = 0.25x^2$$

for $1 \leqslant x \leqslant 2.5$ $\quad F(x) = F(1) + 0.5\int_1^x dt$

The 'area'/'probability' used so far

$$= 0.25 + 0.5\left[t\right]_1^x$$

$$= 0.25 + 0.5\left[x - 1\right] = 0.5x - 0.25$$

Now fully define F(x)

$$F(x)\begin{cases} 0 & x < 0 \\ 0.25x^2 & 0 \leqslant x \leqslant 1 \\ 0.5x - 0.25 & 1 \leqslant x \leqslant 2.5 \\ 1 & x > 2.5 \end{cases}$$

Your notes

Examiner Tips and Tricks

- Remember that P(X=k) = 0 , for any value of k, is zero
  - This can be easily missed when working with c.d.f. rather than a p.d.f.
- A quick check you can do is verify that your c.d.f. is continuous
  - The value of the c.d.f. at the upper limit of one section should equal the value of the c.d.f at the lower limit of the next section