Comparing Discrete & Continuous Random Variables

Consider the probability distribution of the number of heads obtained when tossing two coins i.e. X = number of heads tossed

x	0	1	2
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Definitions

random variable: X

probability function: $P(X = x)$

cumulative probability function: $F(t) = P(X \leq t) = \sum_{x=a}^{t} P(X = x)$

cumulative probability function is also known as the **cumulative distribution function**

e.g. $F(1) = \frac{3}{4}$

For a continuous random variable, a probability function does not exist

Consider the random variable X = height of adults

Between any two values, say 170 cm and 175 cm, there is an infinite number of possible outcomes, hence $P(X = 172.4) = 0$

$$n(172.4) = 23$$
$$n(170 \leq x \leq 175) = \infty$$
$$P(x = 172.4) = \lim_{n \to \infty} \frac{23}{n}$$
$$= 0$$

The probability of any single value in a continuous distribution is zero

It allows continuous data to be treated as discrete data

By using relative frequencies and placing data into classes, a probability function could be evaluated e.g. Let X = waiting time in minutes for a ski lift

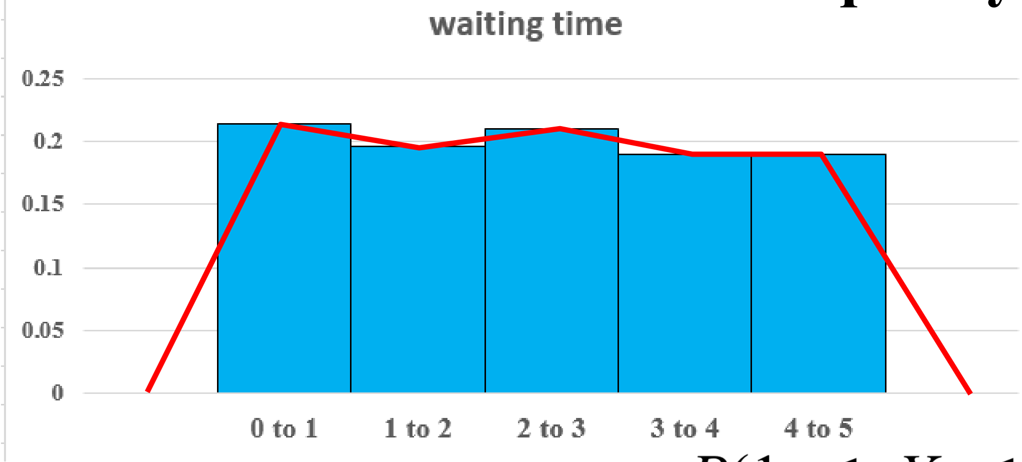
x	**0 – 1**	**1 – 2**	**2 – 3**	**3 – 4**	**4 - 5**
f_r	0.214	0.196	0.210	0.190	0.190
cf_r	0.214	0.410	0.620	0.810	1

Find $P(1 \leq X \leq 3)$

Method 1: use relative frequencies

$$P(1 \leq X \leq 3) = 0.196 + 0.210$$
$$= \underline{0.406}$$

Method 2: use the relative frequency histogram



As the widths of each column is one, the working would be the same as method 1

$$P(1 \leq X \leq 3) = 0.196 + 0.210$$
$$= \underline{0.406}$$

Method 3: use the relative frequency polygon

(approximation using the trapezoidal rule)

x	**1**	**1.5**	**2**	**2.5**	**3**
f_r	0.205	0.196	0.203	0.210	0.2

$$P(1 \le X \le 3) \approx \frac{0.5}{2}\{0.205 + 2(0.196 + 0.203 + 0.210) + 0.2\}$$

$$= 0.40575$$

Method 4: use cumulative relative frequencies

$$P(1 \le X \le 3) = F(3) - F(1)$$

$$= 0.620 - 0.214$$

$$= 0.406$$

Q: how could we find $P(1.5 \le X \le 2.5)$?

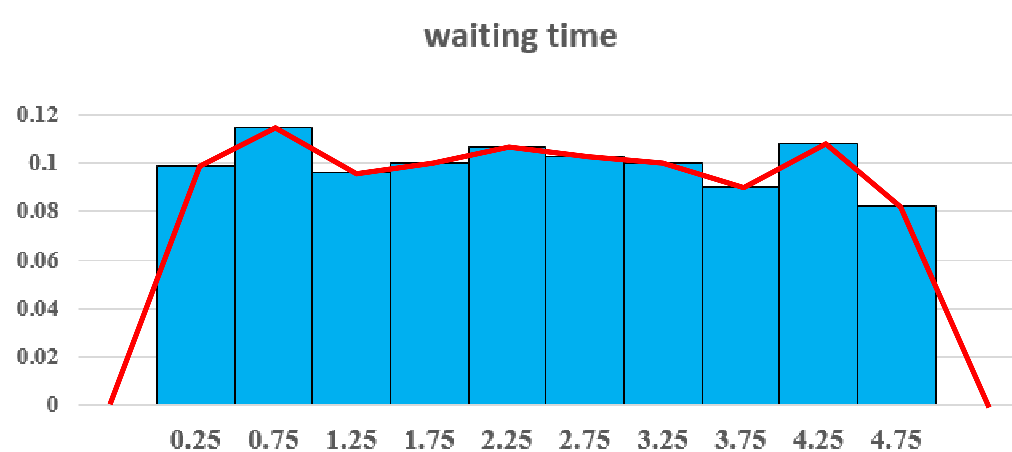
A: reduce the class widths to 0.5

x	0 – 0.5	0.5 – 1	1 – 1.5	1.5 - 2	2 – 2.5	2.5 - 3	3 – 3.5	3.5 - 4	4 – 4.5	4.5 - 5
f_r	0.099	0.115	0.096	0.100	0.107	0.103	0.100	0.090	0.108	0.082
cf_r	0.099	0.214	0.310	0.410	0.517	0.620	0.720	0.810	0.918	1

Method 1: use relative frequencies

$$P(1.5 \leq X \leq 2.5) = 0.100 + 0.107$$
$$= 0.207$$

Method 2: use the relative frequency histogram



as the area of all of the rectangles is $\frac{1}{2}$ we must multiply by 2 to ensure the probabilities sum to 1

$$P(1.5 \leq X \leq 2.5) = \left(\frac{1}{2} \times 0.100 + \frac{1}{2} \times 0.107\right) \times 2$$
$$= 0.207$$

Method 3: use the relative frequency polygon

(approximation using the trapezoidal rule)

x	**1.5**	**1.75**	**2**	**2.25**	**2.5**
f_r	0.098	0.100	0.1035	0.107	0.105

$$P(1.5 \le X \le 2.5) \approx \frac{0.5}{2}\{0.098 + 2(0.100 + 0.1035 + 0.107) + 0.105\} \times 2$$

$$= 0.20675$$

Method 4: use cumulative relative frequencies

$$P(1.5 \le X \le 2.5) = F(2.5) - F(1.5)$$

$$= 0.517 - 0.310$$

$$= 0.207$$

Q: what happens as the width of the class $\to 0$

A: the area under the polygon $\to$ the area of the histograms

$$= F(b) - F(a)$$

i.e. the difference in the cumulative distribution function

Probability Density Functions

The **probability density function** is the equation of the relative frequency polygon as the width of the classes $\to 0$, and has the following properties;

$$(1)\quad f(x) \geq 0 \quad ; \forall\, x$$

$$(2)\quad \int_{-\infty}^{\infty} f(x)dx = 1$$

$$(3)\quad \int_{a}^{b} f(x)dx = P(a \leq X \leq b)$$

e.g. A bid made at an auction for a real estate property, in millions of dollars, can be modelled by the random variable X with the probability density function

$$f(x) = \begin{cases} k(16 - x^2) & 1 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$$

a) Find the value of k

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$k\int_{1}^{4} (16 - x^2)dx = 1$$

$$k\left[16x - \frac{1}{3}x^3\right]_1^4 = 1$$

$$k\left(64 - \frac{64}{3} - 16 + \frac{1}{3}\right) = 1$$

$$27k = 1$$

$$k = \frac{1}{27}$$

b) Find the cumulative distribution function

$$F(x) = \frac{1}{27}\int_{1}^{x} (16 - x^2)dx$$

$$F(x) = \frac{1}{27}\left[16x - \frac{1}{3}x^3 \right]_1^x$$

$$= \frac{1}{27}\left(16x - \frac{1}{3}x^3 - 16 + \frac{1}{3} \right)$$

$$= \frac{1}{81}(48x - x^3 - 47)$$

$$F(x) = \begin{cases} 1 & x > 4 \\ \frac{1}{81}(48x - x^3 - 47) & 1 \leq x \leq 4 \\ 0 & x < 1 \end{cases}$$

c) Find the probability that a bid of more than 3 million dollars will be made

$$P(X \geq 3) = 1 - F(3)$$

$$= 1 - \frac{1}{81}(70)$$

$$= \frac{11}{81}$$

Finding Quartiles

Q_1	Median (M)	Q_3
$\int_{-\infty}^{Q_1} f(x)dx = \frac{1}{4}$	$\int_{-\infty}^{M} f(x)dx = \frac{1}{2}$	$\int_{-\infty}^{Q_3} f(x)dx = \frac{3}{4}$

e.g. The function $f(x)$ is a pdf for a random variable X.

$$f(x) = \begin{cases} \frac{1}{9}x^2 & 0 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$$

Find the median

$$\frac{1}{9}\int_0^M x^2 dx = \frac{1}{2}$$

$$\frac{1}{27}\left[x^3\right]_0^M = \frac{1}{2}$$

$$M^3 = \frac{27}{2}$$

$$M = \frac{3}{\sqrt[3]{2}}$$

median is 2.38 (to 2 dp)

Finding the Mode

mode is the global maximum of the probability distribution function

i.e. $f'(x) = 0$ and $f''(x) < 0$

Note: global maximum could occur at the extremities of the domain

e.g. determine the mode of the pdf defined by

$$f(x) = \begin{cases} \frac{3}{175}(-x^2 + 2x + 15) & 0 \leq x \leq 5 \\ 0 & \text{otherwise} \end{cases}$$

as $f(x)$ is a quadratic, maximum will occur on the axis of symmetry

$$x = \frac{-2}{-2}$$

$$= 1$$

mode is 1

Exercise 16B; 1a, 3, 4, 5bd, 6 (use 5b), 7, 10, 11, 12, 13, 14, 15