

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

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

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

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 $P(X = 172.4) = 0$

$$\begin{aligned}n(172.4) &= 23 \\n(170 \leq x \leq 175) &= \infty \\P(x = 172.4) &= \lim_{n \rightarrow \infty} \frac{23}{n} \\&= 0\end{aligned}$$

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

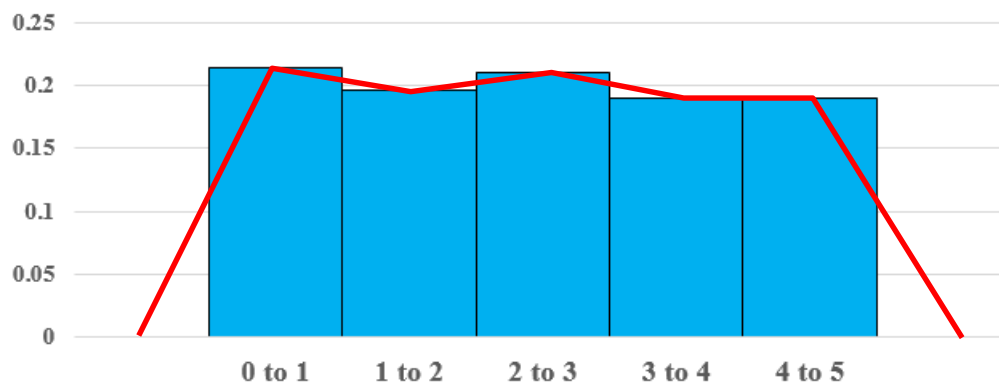
Find $P(1 \leq X \leq 3)$

Method 1: use relative frequencies

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

Method 2: use the relative frequency histogram

waiting time



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

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

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

$$\begin{aligned}P(1 \leq X \leq 3) &\approx \frac{0.5}{2} \{0.205 + 2(0.196 + 0.203 + 0.210) + 0.2\} \\ &= \underline{0.40575}\end{aligned}$$

Method 4: use cumulative relative frequencies

$$\begin{aligned}P(1 \leq X \leq 3) &= F(3) - F(1) \\ &= 0.620 - 0.214 \\ &= \underline{0.406}\end{aligned}$$

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

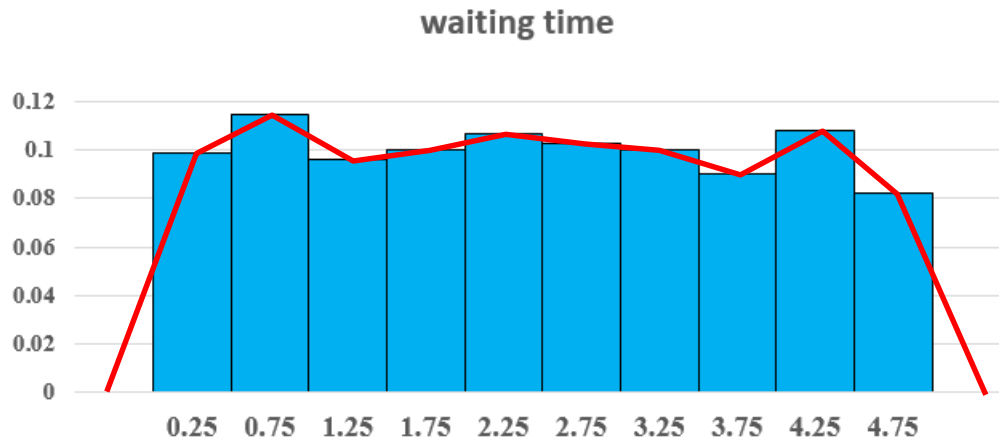
A: reduce the class widths to 0.5

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

$$\begin{aligned} P(1.5 \leq X \leq 2.5) &= 0.100 + 0.107 \\ &= \underline{0.207} \end{aligned}$$

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

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

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

$$\begin{aligned} P(1.5 \leq X \leq 2.5) &\approx \frac{0.5}{2} \{0.098 + 2(0.100 + 0.1035 + 0.107) + 0.105\} \times 2 \\ &= \underline{0.20675} \end{aligned}$$

Method 4: use cumulative relative frequencies

$$\begin{aligned} P(1.5 \leq X \leq 2.5) &= F(2.5) - F(1.5) \\ &= 0.517 - 0.310 \\ &= \underline{0.207} \end{aligned}$$

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

A: the area under the polygon \rightarrow 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 $\rightarrow 0$, and has the following properties;

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

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

$$(3) \int_a^b f(x)dx = P(a \leq X \leq b)$$

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

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

a) Find the value of k

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

$$k \int_1^4 (16 - x^2) dx = 1$$

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

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

$$27k = 1$$

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

b) Find the cumulative distribution function

$$F(x) = \frac{1}{27} \int_1^x (16 - x^2) dx$$

$$\begin{aligned} F(x) &= \frac{1}{27} \left[16x - \frac{1}{3} x^3 \right]_1^x \\ &= \frac{1}{27} \left(16x - \frac{1}{3} x^3 - 16 + \frac{1}{3} \right) \\ &= \frac{1}{81} (48x - x^3 - 47) \end{aligned}$$

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

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

$$\begin{aligned}P(X \geq 3) &= 1 - F(3) \\&= 1 - \frac{1}{81}(70) \\&= \frac{11}{81}\end{aligned}$$

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 .

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

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

$$\text{i.e. } f'(x) = 0 \text{ and } f''(x) < 0$$

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

$$\begin{aligned} x &= \frac{-2}{-2} \\ &= 1 \\ \text{mode is } &\underline{1} \end{aligned}$$

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