Exercise:
1. In the $N$-bit format, how many real numbers that can be represented exactly?
$$2^N$$

2. Use the 64-bit format to find the decimal equivalent of the binary floating-point machine number $\hat{x}$. For each, find also the interval which contains all real numbers represented by $\hat{x}$.

(a) $s = 1$, $e = 2^{10} + 2^3 + 2^2 = 1036$, $f = \frac{1}{2^5} + \frac{1}{2^5} + \frac{1}{2^5} = \frac{13}{32}$
$$\hat{x} = (-1)^1 2^{1036-1023} \left(1 + \frac{13}{32}\right) = -2^{13} \left(1 + \frac{13}{32}\right) = -11520$$
$$\hat{x}_{\text{larger}} = \hat{x} + 2^{13} \left(\frac{1}{2^{52}}\right) = -11520 + \frac{1}{2^{39}}$$
$$\hat{x}_{\text{small}} = \hat{x} - 2^{13} \left(\frac{1}{2^{52}}\right) = -11520 - \frac{1}{2^{39}}$$
the interval: $(-11520 - \frac{1}{2^{39}}, -11520 + \frac{1}{2^{39}})$

(b) $s = 0$, $e = \sum_{k=1}^{9} 2^k = 1022$, $f = \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^5} + \frac{1}{2^7} + \frac{1}{2^8} = \frac{91}{256}$
$$\hat{x} = (-1)^0 2^{1022-1023} \left(1 + \frac{91}{256}\right) = 2^{-1} \left(1 + \frac{91}{256}\right) = \frac{347}{512} = 0.677734375$$
$$\hat{x}_{\text{larger}} = \hat{x} + 2^{-1} \left(\frac{1}{2^{52}}\right) = 0.677734375 + \frac{1}{2^{35}}$$
$$\hat{x}_{\text{small}} = \hat{x} - 2^{-1} \left(\frac{1}{2^{52}}\right) = 0.677734375 - \frac{1}{2^{35}}$$
the interval: $0.677734375 - \frac{1}{2^{35}}, 0.677734375 + \frac{1}{2^{35}}$

3. Let $P_4(x) = 2x^4 + 3x^3 - x^2 - 5x + 2$. Evaluate $P_4(-2)$ using the nested method.
$$P_4(-2) = (((2(-2) + 3)(-2) - 1)(-2) - 5)(-2) + 2 = 16$$

4. Let $P(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0$ be a polynomial and $c$ be a real number.
   a. Give the numbers (in terms of $n$) of multiplications and additions needed to evaluate $P(c)$ without using the nested method.
      - $x^2$, $(x^2)x$, ..., $(x^{n-1})x$, $n - 1$ multiplications
      - $a_1 x$, $a_2(x^2)$, ..., $a_n(x^n)$, $n$ multiplications
      - $a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$, $n$ additions
      - a total of $2n - 1$ multiplications and $n$ additions.
   b. Give the numbers (in terms of $n$) of multiplications and additions needed to evaluate $P(c)$ using the nested method.
\[ P(x) = (a_n x + a_{n-1})x^{n-1} + a_{n-2}x^{n-2} + \ldots + a_1 x + a_0 \]
\[ = ((a_n x + a_{n-1})x + a_{n-2})x^{n-2} + a_{n-3}x^{n-3} + \ldots + a_1 x + a_0 \]
\[ = (((a_n x + a_{n-1})x + a_{n-2})x + a_{n-3})x^{n-3} + \ldots + a_1 x + a_0 \]
\[ = (((((a_n x + a_{n-1})x + a_{n-2})x + a_{n-3})x + \ldots + a_1) x + a_0 \]

\( n \) multiplications and \( n \) additions.

5. Let \((x_1, y_1)\) and \((x_2, y_2)\) be two points on a straight line \( L \) where \( y_1 \neq y_2 \). By the slope-point form, the equation of the line passing through these two points is

\[ y - y_1 = \frac{y_2 - y_1}{x_2 - x_1} (x - x_1). \]

To find the \( x \)-intercept, we first set \( y = 0 \) and then solve \( x \) in terms of \( x_1 \), \( x_2 \), \( y_1 \) and \( y_2 \). Two formulas are derived for the \( x \)-intercept:

(i) \[ x = \frac{x_1 y_2 - x_2 y_1}{y_2 - y_1} \] and (ii) \[ x = x_1 - \frac{(x_2 - x_1)y_1}{y_2 - y_1} \]

\[ \text{a. Show algebraically these formulas are the same.} \]

(ii) \[ x = x_1 - \frac{(x_2 - x_1)y_1}{y_2 - y_1} = \frac{x_1 y_2 - x_2 y_1 + x_1 y_1}{y_2 - y_1} = \frac{x_1 y_2 - x_2 y_1}{y_2 - y_1} \quad (i) \]

\[ \text{b. Let } (x_1, y_1) = (1.31, 3.24) \text{ and } (x_2, y_2) = (1.93, 4.76). \text{ Using the 3-digit rounding arithmetic, compute } x \text{ using both formula. Which one is better? Why?} \]

\[
\begin{array}{|c|c|c|c|c|}
\hline
& x_1 y_2 & x_2 y_1 & y_2 - y_1 & \frac{x_1 y_2 - x_2 y_1}{y_2 - y_1} \\
\hline
\text{i. } \hat{x} & 6.24 & 6.25 & 1.52 & -0.01 \quad \text{and } \frac{x_1 y_2 - x_2 y_1}{y_2 - y_1} = -0.00658 \quad \hat{x} = -0.00658 \\
\hline
\text{ii. } \hat{x} & 0.62 & 2.01 & 1.52 & 1.32 \quad \frac{x_1 y_2 - x_2 y_1}{y_2 - y_1} = 1.32 \quad \hat{x} = -0.01 \\
\hline
\end{array}
\]

the true value \( x = \frac{x_1 y_2 - x_2 y_1}{y_2 - y_1} = \frac{1.31(4.76) - 1.93(3.24)}{4.76 - 3.24} = -1.15789474 \times 10^{-2} \)

Computing the relative errors of two approximations:

i. \[ \frac{|\hat{x} - x|}{|x|} = \frac{|-0.00658 - (-1.15789474 \times 10^{-2})|}{-1.15789474 \times 10^{-2}} = 0.431727274 \]

ii. \[ \frac{|\hat{x} - x|}{|x|} = \frac{|-0.01 - (-1.15789474 \times 10^{-2})|}{-1.15789474 \times 10^{-2}} = 0.136363639 \]

the method in ii. gives a better approximation.