Q.1] If $\beta = 2$, fraction is of 3 bits and exponent is of 6 bits, then what will be the non-negative lowest and highest number that can be generated using the Normalized form of floating point representation? N.B: Exponent range starts from -1.

Soln: exponent is of 6 bits, so for exponent $2^6 = 64$ possible values exists.

non-negative lowest = 0.1000×2^{-1}

highest = 0.1111×2^{62}

Q2] Derive the formula for G_H for the Normalized form of floating point representation.

Soln: Same as lecture note.

Q3] For a system in $\beta = 2, m = 4$ and $e \in \{-3, 3\}$ then how many non-negative numbers can be represented in a system following a Normalized form?

Soln: Normalized form: $0.1 d_1 d_2 d_3 d_4 \times 2^e$

Total non-negative numbers = $2^4 \times 7 = 112$.

Q4] $fl(x*y) = ?$ where $x = 3/8$, $y = 5/8$, $m = 4$, $e \in \{-3, 3\}$ and system follows Normalized form. Also find the rounding error.

Soln: $x \times y = \frac{15}{64} = 0.001111 = 0.1111 \times 2^{-2}$

So no need to round.

$\therefore fl(x \times y) = 0.1111 \times 2^{-2} = 15/64$

error = 0.