Info Processing & the Brain CW 1 (All logs are base 2 unless oterwise specified)

Q1. Marginal of Y, $P(Y)$

$$P(Y=1) = \sum_{x \in X} P(Y=1, X=x)$$

$$= P(Y=1, X=a) + P(Y=1, X=b)$$

$$= 1/16 + \frac{1}{2}$$

$$= \frac{9}{16}$$

$$P(Y=2) = \sum_{x} P(Y=2, X=x)$$

$$= 0 + 1/4$$

$$= 1/4$$

$$P(Y=3) = \sum_{x} P(Y=3, X=x)$$

$$= \frac{1}{16} + \frac{1}{8}$$

$$= \frac{3}{16}$$

Marginal of X, $P(X)$

$$P(X=a) = \sum_{y \in Y} P(Y=y, X=a)$$

$$= P(Y=1, X=a) + P(Y=2, X=a) + P(Y=3, X=a)$$

$$= \frac{1}{16} + 0 + \frac{1}{16}$$

$$= \frac{1}{8}$$

$$P(X=b) = \sum_{y} P(Y=y, X=b)$$

$$= \frac{1}{2} + \frac{1}{4} + \frac{1}{8}$$

$$= \frac{7}{8}$$

Conditional, $P(Y \mid X=a)$

$$P(Y=1 \mid X=a) = \frac{P(Y=1, X=a)}{P(X=a)}$$

$$= \frac{\frac{1}{16}}{\frac{1}{8}}$$

$$= \frac{1}{2}$$

$$P(Y=2 \mid X=a) = P(Y=2, X=a)/P(X=a)$$

$$= 0/{}^1/_8$$

$$= 0$$

$$P(Y=3 \mid X=a) = P(Y=3, X=a)/P(X=a)$$

$$= {}^1/_{16} / {}^1/_8$$

$$= {}^1/_2$$

Q2 i) $H(X) = -\sum_x p(x) \log p(x)$

$$= \frac{1}{8}\log 8 + \frac{7}{8}(\log 8 - \log 7)$$

$$= \frac{3}{8} + \frac{21}{8} - \frac{7}{8}\log 7$$

$$= 3 - \frac{7}{8} \cdot \log 7 = 0.5436$$

$$H(Y) = -\sum_y p(y) \log p(y)$$

$$= \frac{9}{16}(\log 16 - \log 9) + \frac{1}{4}\log 4 + \frac{3}{16}(\log 16 - \log 3)$$

$$= \frac{36}{16} + \frac{2}{4} + \frac{12}{16} - \frac{9}{16}\log 9 - \frac{3}{16}\log 3$$

$$= 3.5 - \frac{9}{16}\log 9 - \frac{3}{16}\log 3$$

$$= \cancel{0.9147}\ \underline{1.4197}$$

~~H(X|Y) =~~

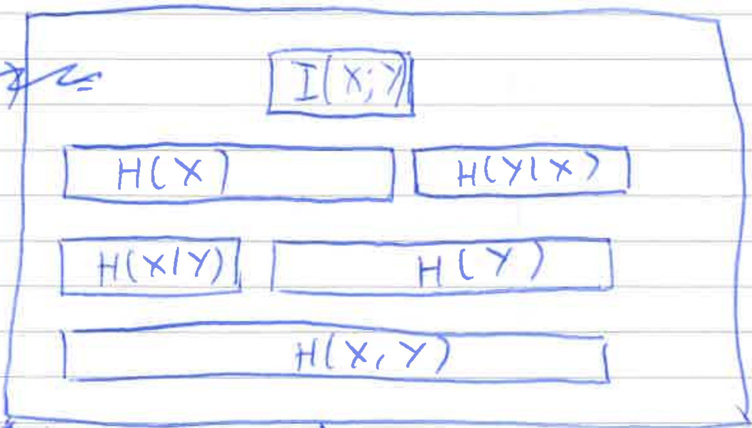
$$H(X,Y) = \sum_x \sum_y p(x,y) \log p(x,y)$$

$$= \frac{1}{16}\log 16 + \frac{1}{16}\log 16 + \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{8}\log 8$$

$$= \frac{1}{4} + \frac{1}{4} + \frac{1}{2} + \frac{1}{2} + \frac{3}{8}$$

$$= 1 + \frac{7}{8} = \underline{1.875}$$

~~H(X|Y) =~~



*The relationship of information measures

$$H(X|Y) = H(X,Y) - H(Y)$$

$$= 1 + \frac{7}{8} - 3 + \frac{7}{8}\log 7$$

$$= -2 + \frac{7}{8}(1 + \log 7)$$

$$= \frac{7}{8}(1 + \log 7) - 2$$

$$\cancel{H(Y|X) = H(X,Y) - H(X)}$$

$$= 1 + \frac{7}{8} -$$

$$H(X|Y) = H(X,Y) - H(Y)$$

$$= 1 + \frac{7}{8} - 3.5 + \frac{9}{16}\log 9 + \frac{3}{16}\log 3$$

$$= \frac{1}{16}\left(14 + 9\log 9 + 3\log 3\right) - 2.5 \approx \underline{0.4553}$$

$$H(Y|X) = H(X,Y) - H(X)$$

$$= 1 + \frac{7}{8} - 3 + \frac{7}{8}\log 7$$

$$= \frac{7}{8}\left(1 + \log 7\right) - 2 \approx \underline{1.3314}$$

$$H(Y) - H(Y|X) = 3.5 - \frac{9}{16}\log 9 - \frac{3}{16}\log 3 - \frac{7}{8}(1 + \log 7) + 2$$

This is ~~the~~ $I(X;Y)$ by the identities

$$= 5.5 + \frac{1}{16}\left(-3\log 3 - 9\log 9 - 14 - 14\log 7\right)$$

$$= 5.5 - \frac{1}{16}\left(3\log 3 + 9\log 9 + 14\log 7 + 14\right)$$

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

$$= 3 - \frac{7}{8}\log 7 + 3.5 - \frac{9}{16}\log 9 - \frac{3}{16}\log 3 - 1 - \frac{7}{8}$$

$$= 5.5 - \frac{14}{16}\log 7 - \frac{9}{16}\log 9 - \frac{3}{16}\log 3 - \frac{14}{16}$$

$$= 5.5 - \frac{1}{16}\left(3\log 3 + 9\log 9 + 14\log 7 + 14\right)$$

$$\approx \underline{0.0883}$$

Calculation sanity check

~~$H(X|Y) = -\sum_{x,y} p(x,y) \log p(x|y)$~~

$p(Y=1 \mid X=b) = \dfrac{p(Y=1, X=b)}{p(X=b)} = \dfrac{1/2}{7/8} = \dfrac{8}{14} = \dfrac{4}{7}$

$p(Y=2 \mid X=b) = \dfrac{p(Y=2, X=b)}{p(X=b)} = \dfrac{1/4}{7/8} = \dfrac{8}{28} = \dfrac{2}{7}$

$p(Y=3 \mid X=b) = \dfrac{p(Y=3, X=b)}{p(X=b)} = \dfrac{1/8}{7/8} = \dfrac{1}{7} = \dfrac{1}{7}$

So

~~$H(X|Y) = -\frac{1}{16} \log$~~

$p(X=a \mid Y=1) = \dfrac{1/16}{9/16} = \dfrac{1}{9}$

~~$p(X=b \mid Y=b) =$~~

$p(X=b \mid Y=1) = \dfrac{1/2}{9/16} = \dfrac{8}{9}$

$p(X=a \mid Y=2) = 0$

$p(X=b \mid Y=2) = \dfrac{1/4}{1/4} = 1$

$p(X=a \mid Y=3) = \dfrac{1/16}{3/16} = 1/3$

$p(X=b \mid Y=3) = \dfrac{1/8}{3/16} = 2/3$

$H(X|Y) = \sum_{x,y} p(x,y) \log p(x|y)$

$= \frac{1}{16} \log 9 + \frac{1}{2} \cdot (\log 9 - \log 8) + 0 + \cancel{\log}\ 0 +$

$\frac{1}{16} \log 3 + \frac{1}{8} (\log 3 - \log 2)$

$= \frac{1}{16} \log 9 + \frac{1}{16} \log 3 + \frac{1}{8} \log 3 - \frac{1}{8} + \frac{1}{2} \log 9 - \frac{3}{2}$

~~$= \frac{1}{16}(\ldots 9(1+8)\log 9 + (1+2)\log 3) + \frac{3}{8} \ldots \frac{1}{8} \ldots \frac{3}{2}$~~

$= \frac{1}{16}(14 + 9\log 9 + 3 \log 3) - 2.5$ 0.4553

~~H(X|Y~~

$$H(Y|X) = \sum_{x,y} p(x,y) \log p(y|x)$$

$$= \frac{1}{16}\log 2 + \frac{1}{2}(\log 7 - \log 4) + 0 + \frac{1}{4}(\log 7 - \log 2)$$

$$+ \frac{1}{16}\log 2 + \frac{1}{8}\log 7$$

$$= \left(\frac{1}{16}+\frac{1}{16}\right)\log 2 + \frac{1}{2}\log 7 - 1 + \frac{1}{4}\log 7 - \frac{1}{4} + \frac{1}{8}\log 7$$

$$= \left(\frac{1}{2}+\frac{1}{4}+\frac{1}{8}\right)\log 7 - 1 - \frac{1}{4} + \frac{1}{8} \quad \left(+\frac{7}{8} - \frac{7}{8}\right)$$

$$= \frac{7}{8}\log 7 - 2 + \frac{7}{8}$$

$$= \frac{7}{8}(1 + \log 7) - 2 \simeq \underline{\underline{1.3314}}$$

They are equivalent ~~so the hotels~~ to using the identities, which was expected. So further calculations will be done using only the identities.

Q3. <u>By ~~o~~ counting</u> &

4 games termination :

A wins in 1 way ~~with prob $(\frac{1}{2})^4$~~ &

B wins in 1 way ~~with prob $(\frac{1}{2})^4$~~, each with prob $\frac{1}{2^4}$

5 games terminate :

i~~A~~ WLOG if A wins, ~~the~~ the final game is won by X so the ~~remaining~~ first 4 games must have 3As and 1B giving us ${}^4C_3$ ways A can win ~~the~~ in 5 games.

~~Similarly~~ So ~~a~~ A wins in ${}^4C_3 = 4$ ways &

B wins in ${}^4C_3 = 4$ ways,

each with prob $\frac{1}{2^5}$

6 games terminate:

WLOG if A wins, the final game is won by A so the first 5 games must be ~~a combination of~~ have 3As and 2Bs giving us ${}^5C_3$ ways A can win in 6 games.

So A wins in ${}^5C_3 = 10$ ways &

B " " 10 ways, each with prob $\frac{1}{2^6}$

7 games terminate:

~~By the same argument as the previous 2 cases~~
WLOG if A ~~wins~~, wins, the final game is won by A so the first ~~5~~ 6 games must have 3As & 3Bs giving us ${}^6C_3$ ways A can win in 7 games

So A wins in ${}^6C_3 = 20$ ways &

B " " 20 ways, each with prob $\frac{1}{2^7}$

~~Since the ways A wins for each length of the series is symmetrical to that of B and of equal number by our independence assumptions, we have that A & B are equally likely to win the series~~

~~$P(X=A) = 0.5$, &~~

~~$P(X=B) = 0.5$.~~

~~The total number of ways any team can win is:~~

~~$1+1+4+4+10+10+20+20 = 70$ ways~~

So, $P(Y=4) = \frac{2}{\cancel{16}} \cdot \frac{1}{2^4} = \frac{1}{2^3} = \frac{1}{8}$

$P(Y=5) = 8 \cdot \frac{1}{2^5} = \frac{1}{2^2} = \frac{1}{4}$

$P(Y=6) = 20 \cdot \frac{1}{2^6} = \frac{10}{2^5} = \frac{5}{2^4} = \frac{5}{16}$

$$p(Y=7) = 40 \cdot \frac{1}{2^7} = \frac{5}{2^4} = \frac{5}{16}$$

$$\cancel{H(X) = \log 2 = 0.5 \log 2 + 0.5 \log 2 = \log 2 = 1}$$

$$H(Y) = \frac{1}{8}\log 8 + \frac{1}{4}\log 4 + 2 \cdot \frac{5}{16}(\log 16 - \log 5)$$

$$= \frac{3}{8} + \frac{2}{4} + \frac{40}{16} - \frac{10}{16}\log 5$$

$$= \frac{27}{8} - \frac{5}{8}\log 5 \approx 3.375 - 1.451205059$$

$$\approx 1.9237795$$

~~H(X)~~ and $\approx \underline{\underline{1.9238}}$

$$H(X) = 2 \cdot \frac{1}{2^4} \cdot \log 2^4 + 8 \cdot \frac{1}{2^5}\log 2^5 + 20 \cdot \frac{1}{2^6}\log 2^6 + 40 \cdot \frac{1}{2^7}\log 2^7$$

$$= \frac{8}{2^4} + \frac{40}{2^5} + \frac{120}{2^6} + \frac{280}{2^7}$$

$$= \frac{1}{2} + \frac{5}{2^2} + \frac{15}{2^3} + \frac{35}{2^4}$$

$$= \frac{8 + 20 + 30 + 35}{2^4}$$

$$= \frac{93}{16} = 5 + \frac{13}{16} = \underline{\underline{5.8125}}$$

$\cancel{H(X|Z) = \mathbb{E}[H(X|Z=z)]}$

$H(Y|X) = 0$, since knowing the outcome also tells us the number of game played.

$$H(X|Y) = H(X) - I(X;Y)$$
$$= H(X) - (H(Y) - H(Y|X))$$
$$= H(X) - H(Y) + H(Y|X)$$
$$\approx 5.8125 - 1.923795$$
$$\approx \underline{\underline{3.8887}}$$

Q4. X is a distribution on 2 events $\{x_1, x_2\}$ where $p(X=x_1) = p$ & $p(X=x_2) = 1-p$.

$$H(X) = -p \log p - (1-p)\log(1-p)$$

Y is a ~~dist~~ continuous uniform distribution on the value of p in the ~~range~~ range $p \in [0,1]$, where the prob density function

$$f(p) = \frac{1}{1-0} = 1 \quad \forall p \in [0,1]$$

So

$$\mathbb{E}_Y[H(X)] = \int_0^1 f(p).(-p\log p - (1-p)\log(1-p))\,dp$$

$$= -\int_0^1 p \log p \, dp - \int_0^1 (1-p)\log(1-p)\, dp$$

(by Integration by parts)

$$\int g(x) \cdot h'(x) dx = g(x) h(x) - \int h(x) \cdot g'(x)\, dx$$

for $\int_0^1 p \log p \, dp$, $g(p) = \log p$, $h'(p) = p$

$g'(p) = \frac{1}{p} \cdot \frac{1}{\ln 2}$, $h(p) = \frac{1}{2}p^2$

~~where ln x = log x~~

$$= \left[\frac{1}{2}p^2 \cdot \log p - \frac{1}{2\ln 2} \int p \, dp \right]_0^1$$

$$= \left[\frac{1}{2}p^2 \cdot \log p - \frac{p^2}{4\ln 2} \right]_0^1$$

$$= -\frac{1}{4\ln 2} - (0) = \frac{-1}{4\ln 2}$$

for $\int_0^1 (1-p) \log(1-p) \, dp$, ~~$g(p) = \log(1-p)$, $h'(p) = 1-p$~~

set $q = 1-p$, when $p = 1$, $q = 0$, & when $p = 0$, $q = 1$. ~~$g'(p) = \frac{-1}{1-p} \cdot \frac{1}{\ln 2}$, $h(p) = p - \frac{1}{2}p^2$~~

so

$$= \int_0^1 q \log q \, dq$$

$$= -\frac{1}{4\ln 2}$$

Finally,

$$\mathbb{E}_X [H(X)] = -\left(\frac{-1}{4\ln 2}\right) - \left(\frac{-1}{4\ln 2}\right)$$

$$= \frac{2}{4\ln 2}$$

$$= \frac{1}{2\ln 2} \simeq 0.721$$

Q5

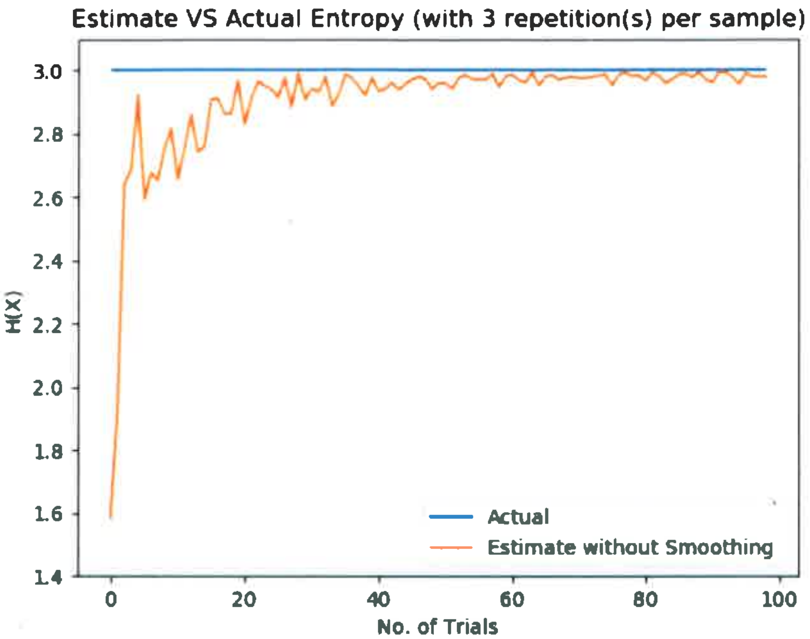


Fig 1: The estimated entropy. Calculated by picking from 8 items, n (no of trials) times, and repeated 3 times to take the average estimated probability of getting each item. The entropy is then calculated and plotted against n. This is done for 0 < n < 100 trials

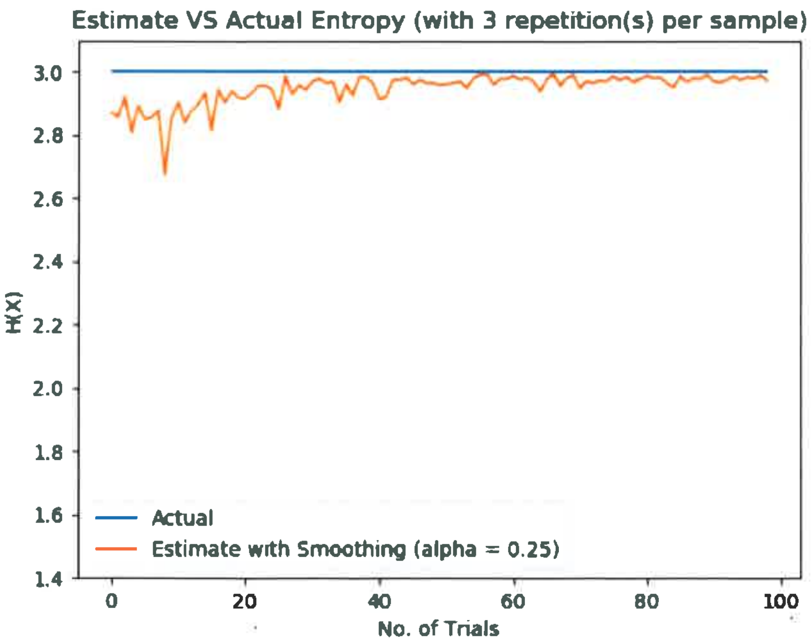


Fig 2: The estimated entropy. Calculated similarly to fig 1, however, Laplace smoothing with alpha = 0.25 is used.

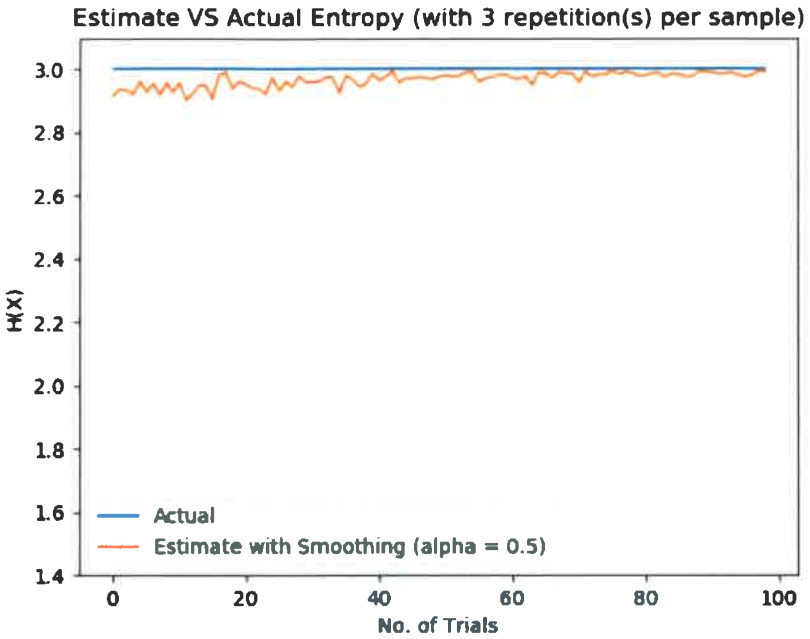


Fig 3: The estimated entropy. Calculated similarly to fig 1, however, Laplace smoothing with alpha = 0.5 is used.

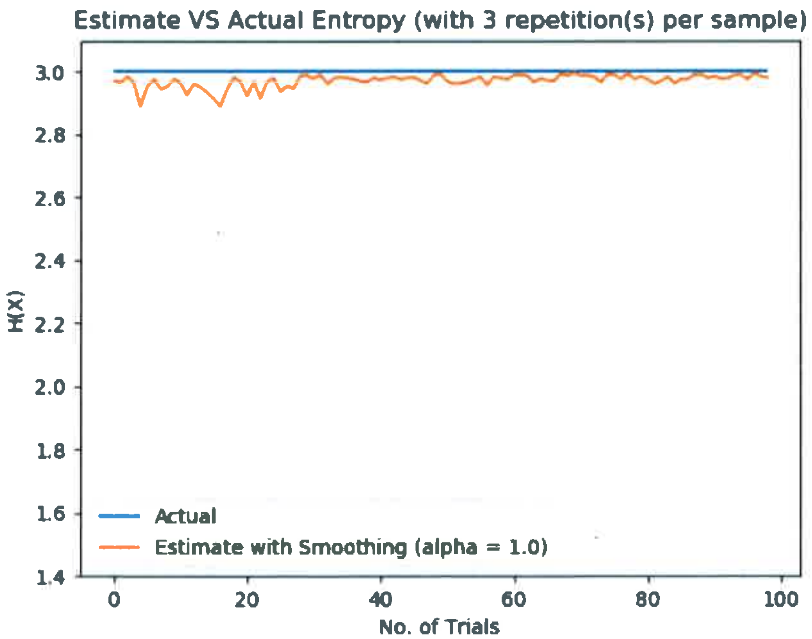


Fig 4: The estimated entropy. Calculated similarly to fig 1, however, Laplace smoothing with alpha = 1.0 is used.

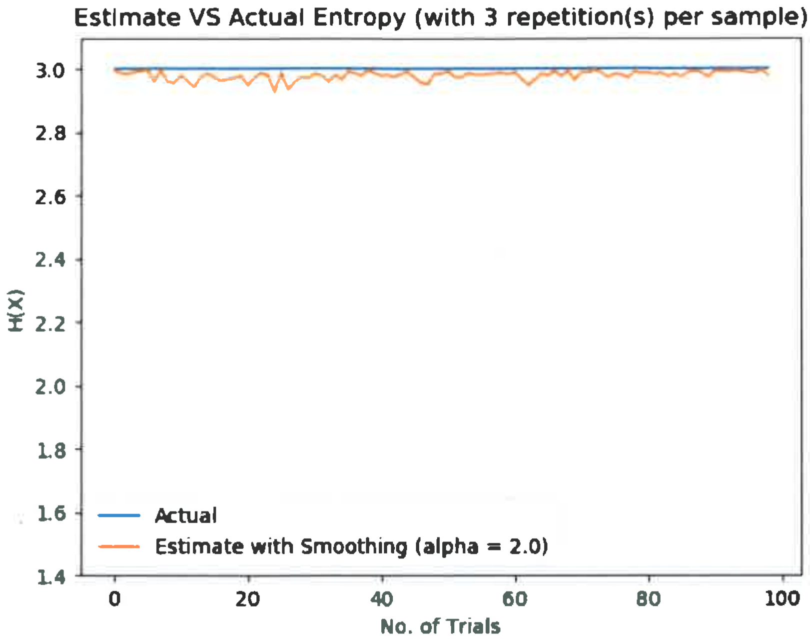


Fig 5: The estimated entropy. Calculated similarly to fig 1, however, Laplace smoothing with alpha 2.0 is used.

<u>Conclusion</u>

Judging by the results, Laplace Smoothing does help as the estimated entropy approaches the actual entropy quicker (in fewer no. of trials) than without the smoothing. As we increase the value of alpha, the estimate approaches the actual entropy much more quickly than without the smoothing. By analysing the equation (equation 3 in the worksheet), we see that a higher value of alpha increases our bias (and strengthens our assumption) of the distribution being a uniformly distributed one. This works well in our case, with a better result with higher alpha, because the distribution we are estimating is, in fact, uniformly distributed.