

Afleveringsopgave: Lykkehjulet

Af Wel Rachid, 20051177

Vi får vist N kort, der hver især dækker over en character. Hver karakter er blevet valgt uafhængigt ud fra den samme distribution p_0, \dots, p_{25} . Vi får lov at vælge et bogstav. Alle pladser hvor dette bogstav befinder sig, skal vendes. Det vi gerne vil åbne er at lære mest muligt om vores string af længde N . Vi skal derudover være opmærksomme på at det at et bestemt bogstav ikke er blandt de N kort, også er en information vi lærer om strengen.

For at beregne hvor meget vi i gennemsnit lære om strengen skal vi modificere vores definition på Entropy. Denne skal nu tage højde for at vi også lærer noget om strengen når vi vælger et bogstav der ikke findes på en bestemt plads.

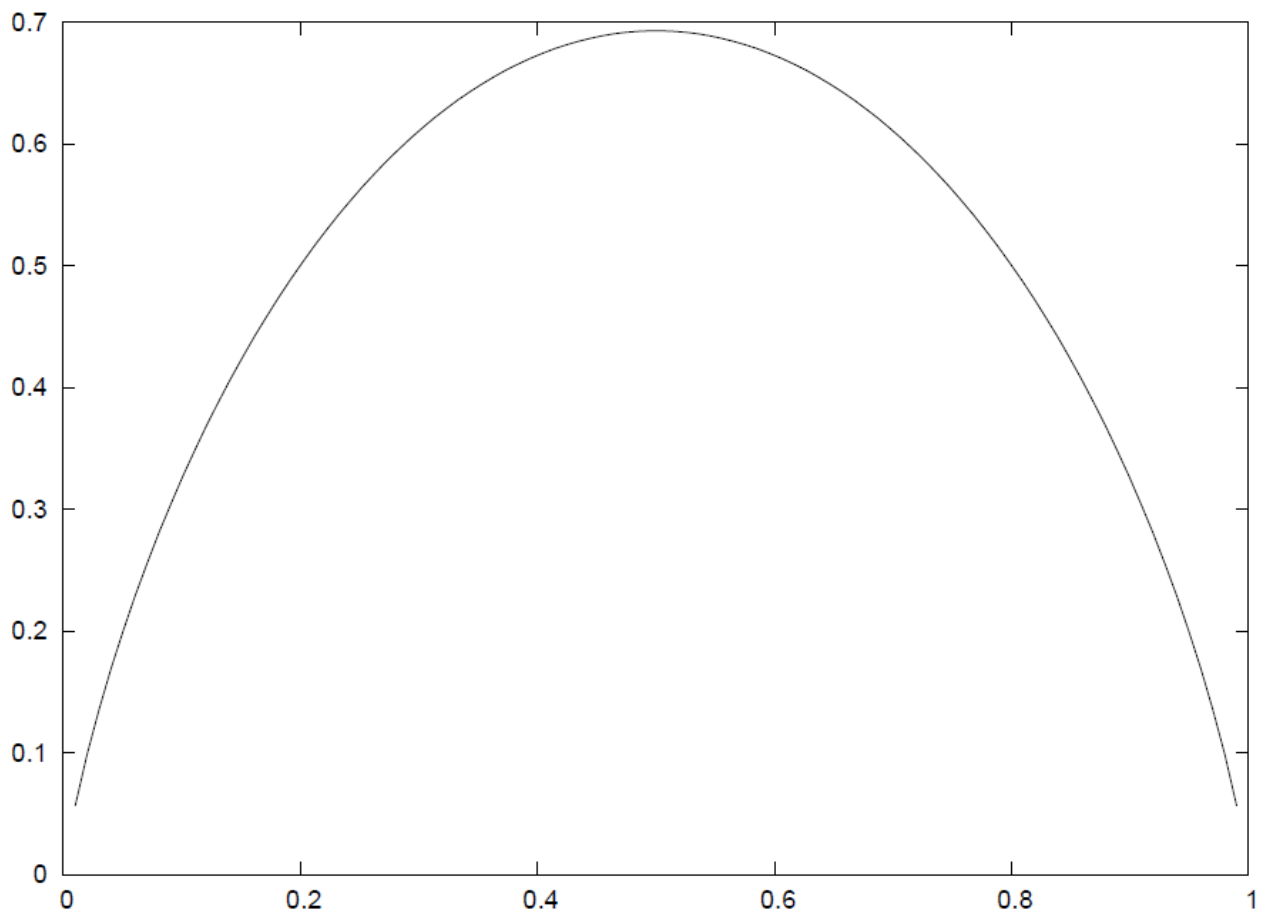
Den oprindelige definition på Entropy ser således ud:

$$H(x) = \sum_{x \in X} (\Pr[x] \log(\Pr[x])), \text{ hvor der her er tale om log base 2}$$

Vores modificerede version kommer således til at rumme en ekstra del som fortæller os noget om strengen når et bogstav ikke findes på en given plads. Denne ser således ud:

$$H(x) = \sum_{x \in X} (\Pr[x] \log(\Pr[x]) - (1 - \Pr[x]) \log(1 - \Pr[x])), \text{ hvor der her er tale om log base 2}$$

Hvis vi prøver at plote den ind i en graf ser den således ud:



Ud fra denne graf kan vi besvare delspørgsmål 1, og sige at vi får ca. 0.7bits information on average.

Dette skyldes at vi har antaget af vores fordeling af distributionen p_0, \dots, p_{25} er den samme for alle p_i .

Delspørgsmål 2 spørger os om hvilken strategi vores resultat antyder for os, hvis vi har en distribution som i det engelske sprog. Her er mit svar at vi bør prøve at tage de bogstaver som giver os flest informationer om strengen. Da de mest sandsynlige bogstaver også er i vores ord, så vil vi få mest information ved at gå efter dem først. Derfor vil mit svar også være at JA, det giver mening at spillerne går efter de bogstaver der er mest frekvente i sproget. Derudover afhænger det ikke af om hvordan frequencies er fordelt, da vi altid vil gå efter de mest sandsynlige.