

1.6.4

(Examen de première session, 2018) En français, la probabilité qu'une lettre prise au hasard dans un texte soit un "O", un "C", un "T" ou un "E" est respectivement égale, approximativement, à 5,02%, 3,18%, 5,92% et 12,20%. Par souci de simplicité, on considère que ces probabilités ne dépendent ni de la place des lettres dans les mots, ni de la nature des lettres voisines.

Sous ces hypothèses, on demande de calculer la quantité d'information contenue dans le mot "OCTET".

Résolution :

On calcul la quantité d'information totale Φ en sommant les quantités d'informations individuelles : β_i .

$$\Phi = \sum_i \beta_i = \sum_i \log_2\left(\frac{1}{p_i}\right)$$

On a :

i	$\beta_i(bits)$
O	4,32
C	4,97
T	4,09
E	3,04
T	4,09

$$\Phi = \sum_i \beta_i \approx 20,51bits$$