

① Escoger el castellano, aunque ya que es <sup>uno de los</sup> idiomas europeos con el que más familiarizado estoy.

② Modelar como fuente discreta sin memoria con distribución uniforme.

Asumimos que todas las letras tienen la misma probabilidad, como hay 27 letras, pues  $\frac{1}{27}$  cada letra.

Calcularemos la entropía, con  $H = -\sum p_i \log_2(p_i)$

$$H = \sum p_i \log_2\left(\frac{1}{p_i}\right) = \frac{27}{27} \cdot \log_2(27) = \log_2(27) = 4.755$$

③ Modelar como fuente discreta sin memoria (DMS) con distribución real de letras, para calcularlo he usado como base de datos el libro Don Quijote de la Mancha de Cervantes, que mediante un programa hecho con python he analizado al pdf, donde me sale que hay 1012920 letras con la los siguientes datos:

a → 122452    b → 13285    c → 37867    d → 55146    e → 139792    f → 4675  
g → 10872    h → ...    i → 4172

Por tanto la distribución de probabilidad para  $\{a, b, c, d, \dots, y, z\}$

$$\text{es } \left\{ \frac{122452}{1012920}, \frac{13285}{1012920}, \dots, \frac{4172}{1012920} \right\}$$

Calculamos la entropía:  $H = -\sum p_i \log_2(p_i) = 4.106 \text{ bits}$

④

Como este punto es muy largo, lo haré principalmente con el ordenador:

Primero para calcular la entropía, que se define como:

$$H = - \sum_i \sum_j p(i) \cdot p(j|i) \cdot \log_2(p(j|i))$$

Mediante python analizaremos las transiciones entre letras y la frecuencia de cada una de las transiciones, obteniendo que para  $\{aa, ab, ac, \dots, zz\}$ , con un total de transiciones de 1012919, la distribución es  $\{\frac{2625}{1012919}, \dots, \dots\}$

$$P(ab) = \frac{2625}{1012919}$$

Encontramos que la entropía es 0.528260