Question 1. À chaque seconde un singe appuie au hasard sur une touche du clavier. Combien de temps θ se passera-t-il en moyenne pour que le mot ABRACADABRA apparaisse à l'écran ?

Soient Σ un alphabet fini de $m \coloneqq |\Sigma|$ lettres, $\mathfrak{p} \coloneqq p_1 \cdots p_k$ un mot de $k \geqslant 1$ lettres formé sur Σ, et $\mathcal{T}$ l'ensemble des mots sur Σ terminant par $\mathfrak{p}$ et ne comportant aucune autre occurrence de $\mathfrak{p}$. Il s'agit de déterminer la longueur moyenne d'un mot de $\mathcal{T}$, autrement dit de calculer

$$\theta \coloneqq \sum_{w\in\mathcal{T}} \frac{|w|}{m^{|w|}} = \sum_{n=0}^{+\infty} \frac{nt_n}{m^n},$$

où t_n, $n \geqslant 0$, est le nombre de mots de $\mathcal{T}$ de longueur n. Pour cela nous allons explicter la somme de la série génératrice $T(X) \coloneqq \sum_{n=0}^{+\infty} t_n X^n$. En effet, $\theta = \frac{1}{m} T'(\frac{1}{m})$!

Soient $\mathcal{S}$ l'ensemble des mots sur Σ ne comportant aucune occurrence de $\mathfrak{p}$, et $S(X)$ la série génératrice associée :

$$S(X) \coloneqq \sum_{w\in\mathcal{S}} X^{|w|} \coloneqq \sum_{n=0}^{+\infty} s_n X^n,$$

où s_n, $n \geqslant 0$, est le nombre de mots de $\mathcal{S}$ de longueur n. Notons que $\mathcal{S}$ contient le mot vide ε ($s_0 = 1$) et que les ensembles $\mathcal{S}$ et $\mathcal{T}$ sont disjoints. Nous allons chercher un système à deux équations reliant $S(X)$ et $T(X)$.

Première équation. Nous avons

$$\mathcal{S} \sqcup \mathcal{T} = \{\varepsilon\} \sqcup (\mathcal{S} \cdot \Sigma).$$

Démonstration. Si $w \in S$, alors soit $w = \varepsilon$, soit w est de la forme $w = w'a$ avec $a \in \Sigma$ et $w' \in \mathcal{S}$ (sinon w contiendrait une occurrence de $\mathfrak{p}$). De même, si $w \in \mathcal{T}$, alors w est de la forme $w = w'a$ avec $a \in \Sigma$ et $w' \in \mathcal{S}$ (sinon w contiendrait au moins deux occurrences de $\mathfrak{p}$). Réciproquement, si $w = \varepsilon$ alors $w \in \mathcal{S}$, et si $w = w'a$ avec $a \in \Sigma$ et $w' \in \mathcal{S}$, alors a termine ou non une occurrence de $\mathfrak{p}$, donc $w \in \mathcal{S} \sqcup \mathcal{T}$.

D'où

$$\begin{aligned} S(X) + T(X) &= \sum_{w\in\mathcal{S}\sqcup\mathcal{T}} X^{|w|} \\ &= 1 + \sum_{w\in\mathcal{S}\cdot\Sigma} X^{|w|} \\ &= 1 + \sum_{w'\in\mathcal{S}} \sum_{a\in\Sigma} X^{|w'|+1} \\ &= 1 + mX \sum_{w'\in\mathcal{S}} X^{|w'|} \\ S(X) + T(X) &= 1 + mXS(X). \end{aligned} \tag{1}$$

Deuxième équation. Soit I l'ensemble des indices $i \in \{1, \ldots, k\}$ tels que $\mathfrak{p}$ commence et termine à la fois par $p_1 \cdots p_i$, c'est-à-dire $p_1 \cdots p_i = p_{k-i+1} \cdots p_k$ (notons que I contient nécessairement k). Alors

$$\mathcal{S} \cdot \{\mathfrak{p}\} = \mathcal{T} \cdot \left(\bigsqcup_{i\in I} \{p_{i+1} \cdots p_k\} \right).$$

Démonstration. Soit $w = w_1 \cdots w_n \in \mathcal{S} \cdot \{\mathfrak{p}\}$. Soit $i \coloneqq \max\{1 \leqslant \ell \leqslant k \mid w_1 \cdots w_{n-k+\ell} \text{ termine par } \mathfrak{p}\}$. Alors $w_1 \cdots w_{n-k+i} \in \mathcal{T}$, et comme w termine par $\mathfrak{p}$ nous avons

$$\begin{aligned} p_1 \cdots p_i p_{i+1} \cdots p_k &= w_{n-k+1} \cdots w_{n-k+i} w_{n-k+i+1} \cdots w_n \\ &= p_{k-i+1} \cdots p_k p_{i+1} \cdots p_k, \end{aligned}$$

donc $p_1 \cdots p_i = p_{k-i+1} \cdots p_k$. Ainsi $i \in I$ et $w \in \mathcal{T} \cdot \{p_{i+1} \cdots p_k\}$.

Réciproquement, soit $w \in \mathcal{T} \cdot \{p_{i+1} \cdots p_k\}$ avec $i \in I$. Alors w est de la forme

$$\begin{aligned} w &= \underbrace{[w' p_1 \cdots p_{k-i}][p_{k-i+1} \cdots p_k]}_{\in\mathcal{T}}[p_{i+1} \cdots p_k] \\ &= \underbrace{w' p_1 \cdots p_{k-i}}_{\in\mathcal{S}} \cdot \mathfrak{p}. && \text{(car } i \in I) \end{aligned}$$

D'où
$$\begin{aligned} X^k S(X) &= \sum_{w\in\mathcal{S}\cdot\{\mathfrak{p}\}} X^{|w|} \\ &= \sum_{\substack{i\in I\\ w\in\mathcal{T}}} X^{k-i}X^{|w|} \\ &= \left(\sum_{i\in I} X^{k-i}\right)\left(\sum_{w\in\mathcal{T}} X^{|w|}\right) \\ X^k S(X) &=: c(X)\,T(X). \end{aligned} \tag{2}$$

Conclusion. En multipliant (1) par X^k, et en utilisant (2), nous avons donc
$$c(X)\,T(X) + X^k T(X) = X^k + mXc(X)\,T(X),$$
soit
$$T(X) = \frac{X^k}{X^k + (1-mX)c(X)}.$$

Par conséquent
$$\theta = \frac{1}{m}T'\left(\frac{1}{m}\right) = \frac{1}{m}\left(\frac{km^{1-k}}{m^{-k}} - m^{-k}\frac{km^{1-k} - mc(\frac{1}{m})}{m^{-2k}}\right) = m^k c\left(\frac{1}{m}\right) = \sum_{i\in I} m^i.$$

Si $\Sigma = \{\mathsf{A},\dots,\mathsf{Z}\}$ et $\mathfrak{p} = \mathsf{ABRACADABRA}$, alors $m = 26$, $k = 11$, $I = \{1,4,11\}$. Ainsi
$$\theta = 26^{11} + 26^4 + 26\ s \quad \approx 3 \text{ milliards d'années.}$$

Références. [FS]

190 Méthodes combinatoires, problèmes de dénombrement.

[FS] Philippe Flajolet et Robert Sedgewick : *Analytic Combinatorics.*