Exercice 4. *Nombre d'occurences.* On se donne un mot

$$\omega = \omega_1 \omega_2 \dots \omega_n$$

de longueur n, constitué de n lettres tirées au hasard avec remise dans un sac. Le sac contient un exemplaire de chaque caractère d'imprimerie (on suppose que ceux-ci sont au nombre de t). On note $N(\omega)$ le nombre d'occurences du mot CARTAN dans ω. Calculer $\mathbb{E}[N]$. *Questions subsidiaires :* calculer $\mathrm{Var}(N)$. Peut-on appliquer l'inégalité de Hoeffding à N directement ? Indirectement ?

Autre question subsidiaire Soit T le nombre de lettre qu'il faut lire pour enfin observer une occurence du mot CARTAN Trouver n tel que $\mathbb{P}(T\leq n) \geq 0{,}99$

On a $N_n = X_1 + X_2 + \dots + X_{n-5}$, avec $X_i \sim \text{Bernoulli}(t^{-6})$, et

$\mathrm{Cov}(X_i, X_j) = t^{-6} \mathbb{1}_{|i-j|=0} - t^{-12} \mathbb{1}_{|i-j|\leq 5}$, donc

$\mathbb{E}(N_n) = (n-5)\,t^{-6}$ et $\mathrm{Var}(N_n) = (n-5)\,t^{-6} - t^{-12}\left((n-5)^2 - (n-10)(n-11)\right) = (n-5)\,t^{-6} - t^{-12}(11n-76)$

On remarque que $\{T\leq n\} \Leftrightarrow \{N_n \geq 1\} \Leftrightarrow \{N_n - \mathbb{E}[N_n] \geq 1 - \mathbb{E}(N_n)\}$

ou encore $\{T>n\} \Leftrightarrow \{N_n \leq 0\} \Leftrightarrow \{N_n - \mathbb{E}[N_n] \leq -\mathbb{E}[N_n]\}$ Donc on peut utiliser, a priori, Hoeffding ou BT pour démontrer que $\mathbb{P}(T>n) \leq 0{,}01$ Par exemple, avec BT,

$$\mathbb{P}(T>n) \leq \mathbb{P}(|N_n - \mathbb{E}(N_n)| \geq \mathbb{E}(N_n)) \leq \frac{\mathrm{Var}(N_n)}{\mathbb{E}(N_n)^2} \simeq \frac{n(t^{-6} - 11t^{-12})}{n^2\, t^{-12}} = \frac{t^6 - 11}{n} \simeq \frac{t^6}{n} \leq 0{,}01$$

à condition que $n \sim 100\, t^6$

Utiliser Hoeffding est + délicat car les X_i ne sont pas $\perp\!\!\!\perp$ Plus précisément,

$$\{X_i \perp\!\!\!\perp X_j\} \Leftrightarrow \{|i-j| \geq 6\}$$

Ainsi on peut appliquer Hoeffding séparément aux sommes $S_1, S_2, S_3, \dots, S_6$, définies par

$S_1 = X_1 + X_7 + X_{13} + \dots$, $S_2 = X_2 + X_8 + X_{14} + \dots$, ..., vérifiant $N = S_1 + S_2 + \dots + S_6$ Chacune des sommes S_i possède entre $\frac{n-10}{6}$ et $\frac{n}{6}$ termes Les S_i vérifient donc, pour $t<0$,

$$\mathbb{P}(S_i - \mathbb{E}(S_i) \leq -t) \leq \exp\left(-12t^2/n\right)$$

Par ailleurs, $N_n - \mathbb{E}(N_n) = \sum_{i=1}^{6} S_i - \mathbb{E}(S_i)$, donc

$$\{N_n - \mathbb{E}(N_n) \leq -t\} \Rightarrow \{\exists i \text{ tq } S_i - \mathbb{E}(S_i) \leq -\tfrac{t}{6}\},$$

et $\mathbb{P}(N_n - \mathbb{E}(N_n) \leq -t) \leq \sum_i \mathbb{P}(S_i - \mathbb{E}(S_i) \leq -t/6) \leq 6\exp(-t^2/3n)$

Ainsi, il suffit de s'assurer que

$$\exp\left(\mathbb{E}(N_n)^2/3n\right) \geq 600,$$

pour obtenir la majoration désirée Il suffit pour cela que

$$n \geq 10 + 3\ln(600)\; t^6 \sim 20\, t^6$$

On gagne donc un facteur 5 sur BT