TD 07 – Convergence de variables aléatoires et (encore) partiel de l'an dernier

Exercice 1. *Suite de bits aléatoires*

On se donne X_i une suite infinie de bits aléatoires non biaisés.

1. Montrer que presque sûrement tout mot fini apparaît dans la suite X_i.
2. En déduire que la presque sûrement tout mot fini apparaît une infinité de fois dans la suite X_i.

Exercice 2. *Algorithme probabiliste pour calculer la médiane*

On étudie un algorithme probabiliste [1] pour déterminer la médiane d'un ensemble $E = \{x_1, \ldots, x_n\}$ de n nombres réels en temps $O(n)$. On rappelle que m est une médiane de E si au moins $\lceil n/2 \rceil$ des élements de E sont inférieurs ou égaux à m, et au moins $\lceil n/2 \rceil$ des élements de E sont supérieurs ou égaux à m. Pour simplifier on suppose n impair (ce qui fait que la médiane est unique) et on suppose aussi que les éléments de E sont tous distincts.

Voici comment fonctionne l'algorithme

(a) Soit $(Y_i)_{1 \le i \le n}$ une suite de v.a. i.i.d. de loi de Bernoulli de paramètre $n^{-1/4}$. On considère le sous-ensemble aléatoire de E défini par $F = \{x_i : Y_i = 1\}$. Si card $F \le \frac{2}{3}n^{3/4}$ ou card $F \ge 2n^{3/4}$ on répond «`ERREUR 1`».

(b) On trie F et on appelle d le $\lfloor \frac{1}{2}n^{3/4} - \sqrt{n} \rfloor$ème plus petit élément de F, et u le $\lfloor \frac{1}{2}n^{3/4} - \sqrt{n} \rfloor$ème plus grand élément de F.

(c) On détermine le rang de d et de u dans E (l'élément minimal a rang 1, l'élément maximal a rang n), que l'on note respectivement r_d et r_u. Si $r_d > n/2$ ou $r_u < n/2$ on répond «`ERREUR 2`».

(d) On note $G = \{x_i \in E : d < x_i < u\}$. Si card $G \ge 4n^{3/4}$ on répond «`ERREUR 3`».

(e) On trie G et on renvoie le $(\lceil n/2 \rceil - r_d)$ème élement de G.

1. Justifier pourquoi l'algorithme retourne la médiane en temps $O(n)$ lorsqu'il ne répond pas de message d'erreur.
2. Montrer que pour $i \in \{1, 2, 3\}$, on a
$$\lim_{n \to \infty} \Pr\left(\text{l'algorithme retourne «}\texttt{ERREUR } i\text{»}\right) = 0.$$
Pour simplifier l'analyse et éviter d'écrire des symobles $\lfloor \cdot \rfloor$ ou $\lceil \cdot \rceil$, on pourra supposer implicitement que des nombres tels que $\sqrt{n}$, $\frac{1}{2}n^{3/4}$, ... sont des entiers

Exercice 3. *Approximation de Poisson*

On se place dans le modèle *Balls and Bins* où l'on jette m balles au hasard dans n paniers. Le problème est que les v.a. X_i représentant le nombre de balles dans le i-ème panier ne sont pas indépendantes (intuitivement, car $X_1 + \cdots + X_n = m$). On voudrait approximer le modèle *Balls and Bins* par le modèle *Approximation de Poisson*, dans lequel $Y_1, \ldots Y_n$ sont des variables aléatoires indépendantes qui suivent chacune une loi de Poisson de moyenne $\mu = m/n$ (la variable Y_i est donc pensée pour être la version simplifiée de X_i).

1. Montrer que $Y = \sum_{i=1}^n Y_i$ suit une loi de Poisson dont on précisera le paramètre.
2. Montrer que la distribution de $(Y_1, \ldots, Y_n)$ conditionnée au fait que $Y = m$ est la même que la distribution de $(X_1, \ldots, X_n)$.
Note : on peut en fait obtenir un résultat légèrement plus général. Si $(X_1, \ldots, X_n)$ représente la charge de n paniers après avoir lancé au hasard k balles, et que les Y_i sont n v.a. indépendantes suivant chacune une loi de Poisson de paramètre m/n, alors la distribution de $(Y_1, \ldots, Y_n)$ conditionnée au fait que $Y = k$ est la même que la distribution de $(X_1, \ldots, X_n)$, indépendamment de la valeur de m.

1. Remarque : il existe un algorithme déterministe de même performance

3. Soit f une fonction sur n variables, à valeurs réelles positives ou nulles. Prouver que

$$\mathbf{E}\left[f(X_1,\ldots,X_n)\right] \leq e\sqrt{m}\mathbf{E}\left[f(Y_1,\ldots,Y_n)\right] \ .$$

   On pourra prouver comme étape intermédiaire que $m! < e\sqrt{m}\left(\frac{m}{e}\right)^m$.

4. En déduire le corollaire suivant : soit $\mathcal{E}$ un événement qui dépend de la charge des paniers. Supposons que $\mathcal{E}$ arrive avec probabilité p dans l'*Approximation de Poisson*, c'est-à-dire si la charge des paniers est $(Y_1,\ldots,Y_n)$. Alors $\mathcal{E}$ arrive avec probabilité au plus $pe\sqrt{m}$ dans le modèle *Balls and Bins*, c'est-à-dire si la charge des paniers est $(X_1,\ldots,X_n)$.

5. *Application :* On jette n balles dans n paniers selon le modèle *Balls and Bins*. Montrer qu'avec probabilité au moins $1-1/n$ (pour n assez grand), la charge maximale est $\geq \ln n/\ln\ln n$.

Exercice 4. *Conditions de convergence*

Soit X_n une suite infinie de variables de Bernoulli indépendantes de paramètres $1-p_n$, avec $0 \leq p_n \leq 1/2$ (i.e. $\mathbf{P}\left\{X_n = 1\right\} = 1-p_n$ et $\mathbf{P}\left\{X_n = 0\right\} = p_n$).

1. Donner une condition nécessaire et suffisante pour que la suite X_n converge en distribution.
2. Donner une condition nécessaire et suffisante pour que la suite X_n converge en probabilité.
3. Donner une condition suffisante (nécessaire ce sera la semaine prochaine) pour que la suite X_n converge presque sûrement.