TD 08 – Convergence de variables aléatoires

Exercice 1. *Second theorème de Borel-Cantelli*

L'objectif de cet exercice est de montrer le second théorème de Borel-Cantelli. Il donne une réciproque du theorème de Borel-Cantelli vu en cours, dans le cas où les événements sont indépendants. Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements *indépendants* de probabilité p_n. On suppose que la somme $\sum_n p_n$ diverge. L'objectif de cet exercice est de montrer qu'alors, presque sûrement, une infinité d'événements A_n se réalisent.

1. Exprimer l'événement "une infinité d'événements A_n se réalisent" en terme d'unions et d'intersections des événements A_n.
2. Soit $B_{k,\ell}$ l'événement $\cap_{k\leq n\leq \ell}\overline{A_n}$. Montrer que pour tout k fixé, $\lim_{\ell\to\infty}\mathbf{P}\left\{B_{k,\ell}\right\}=0$. *Indice : on pourra utiliser l'inégalité* $1+x\leq e^x$ *pour tout* $x\in\mathbb{R}$.
3. On note $B_k=\cap_{n\geq k}\overline{A_n}$. En déduire que $\mathbf{P}\left\{\cup_k B_k\right\}=0$.
4. Conclure que $\mathbf{P}\left\{\text{"une infinité d'événements } A_n \text{ se réalisent"}\right\}=1$.
5. *Application.* Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables de Bernoulli indépendantes de paramètre $\mathbf{P}\left\{X_n=1\right\}=p_n=1/n$. Montrer que presque sûrement la suite X_n contient un nombre infini de '1', mais seulement un nombre fini de '11'.

Exercice 2. *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 nécessaire et suffisante pour que la suite X_n converge presque sûrement.

Exercice 3. *Calcul de limite*

L'objectif de cet exercice est de montrer l'égalité suivante :

$$\lim_n \exp(-n)\sum_{k=0}^{n}\frac{n^k}{k!}=\frac{1}{2}$$

On considère $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes de loi de Poisson de paramètre 1. On note $S_n=\sum_{k=1}^n X_k$.

1. Montrer que pour tout $n\geq 1$, $\mathbf{P}\left\{S_n\leq n\right\}=\exp(-n)\sum_{k=0}^{n}\frac{n^k}{k!}$.
2. Conclure en utilisant le Théorème Central Limite.

Exercice 4. *Théorème de Mycielski*

Recall that the *chromatic number* $\chi(G)$ is the smallest number of colors needed to color the vertices of G such that any two adjacent vertices have different colors. Clearly, graphs with large cliques have a high chromatic number, but the opposite is not true. The goal of this exercise is th prove Mycielski's theorem, which states that for any integer $k \geq 2$, there exists a graph G such that G contains no triangles and $\chi(G) \geq k$.

1. Fix $0 < \varepsilon < \frac{1}{3}$ and let G be a random graph on n vertices where each edge appears independently with probability $p = n^{\varepsilon - 1}$. Show that when n tends to infinity, the probability that G has more than $n/2$ triangles tends to 0.
2. Let $\alpha(G)$ be the size of the largest *independent set* of G (A set of vertices X is *independent* if there is no edge between any two vertices of X in G). Show that $\chi(G) \geq n/\alpha(G)$.
3. Let $a = 3n^{1-\varepsilon} \ln n$. Show that when n tends to infinity,
$$\mathbb{P}(\alpha(G) < a) \to 1.$$
Deduce that there exists n and G of size n such that G has at most $n/2$ triangles and $\alpha(G) < a$.
4. Let G be such a graph. Let G' be a graph obtained from G by removing a minimum number of of vertices so that G' does not contain any triangle. Show that
$$\chi(G') > \frac{n^{\varepsilon}}{6 \ln n}$$
and conclude the proof of Mycielski's Theorem.

Exercice 5. *Filtres de Bloom*

[Disclaimer : l'exercice parle de choses vues en cours que vous n'avez en fait pas vu. Mais ça n'a aucune importance.]

Rappelez-vous les tables de hachage vues en cours et reprenons l'exemple de l'interdiction des mots de passe trop simples. On dispose d'un ensemble F de mots de passe interdits, et l'on veut stocker F de manière intelligente pour pouvoir, à chaque fois qu'un utilisateur choisi un nouveau mot de passe, vérifier si ce mot de passe est admissible. Dans le premier exemple du cours (*Chain Hashing*), on cherche à minimiser le temps d'une requête pour savoir si $x \in F$. Dans le deuxième exemple du cours (*Bit Strings/Fingerprints*), on cherche à minimiser l'espace de stockage de F, quitte à ce que certaines requêtes produisent un faux positif (i.e. répond que $x \in F$ alors que $x \notin F$).

On va s'intéresser ici à un troisième exemple appelés *filtre de Bloom* qui permet d'obtenir un meilleur compromis entre espace de stockage et taux de faux positifs. Un filtre de Bloom est un tableau A à n cases, initialement remplies à 0. On dispose de k fonctions de hachage indépendantes $h_1, \ldots, h_k$ à valeurs dans $\{1, \ldots, n\}$. On suppose comme à l'accoutumée pour les fonctions de hachage, que chaque h_i associe à n'importe quel élément de l'univers un nombre choisi uniformément au hasard dans $\{1, \ldots, n\}$. Soit $F = \{f_1, \ldots, f_m\}$ l'ensemble des m mots interdits. L'étape de pré-processing est la suivante : pour chaque $f \in F$, et pour chaque $i \leq k$, on met $A[h_i(f)]$ à 1 (si cette case était déjà à 1, on ne la touche pas). Supposons maintenant que l'on ait une requête du type $s \in ?F$. On répond de la manière suivante : si tous les $A[h_i(s)]$ valent 1 pour $1 \leq i \leq k$, alors on répond $s \in F$. Sinon, on répond $s \notin F$. On vérifie facilement qu'il est impossible d'obtenir un faux-négatif.

1. Soit X le nombre de cases de A dans lesquelles il reste un 0 après le pré-processing. Quelle est l'espérance de X/n ?
2. Soit $p = e^{-km/n}$. Dans cette question, on suppose pour simplifier que X est égal à pn. Quelle est la probabilité P d'un faux positif ? Comment choisir k pour minimiser P, et qu'obtient-on comme valeur de P ?
3. Justifier pourquoi il a semblé raisonnable de supposer, par simplification, que $X = pn$. Plus exactement, utiliser l'approximation de Poisson pour borner $\mathbf{P}\{|X - np| \geq \varepsilon n)\}$, et commenter.