# TD n°08

28 mars 2018

### Exercice 1 - Approximation de Poisson

On considère à nouveau le modèle *Boules et Bacs* (*Balls and Bins*) : on met aléatoirement $m$ boules dans $n$ bacs. Les nombres de boules dans les différents bacs ne sont pas indépendants. On veut approximer ce modèle par une loi de Poisson. $Y_1, \dots, Y_n$ sont des variables aléatoires indépendantes suivant chacune une loi de Poisson avec paramètre (i.e., espérance) $\mu = m/n$.

**1.1** Montrez que $Y = \sum_{i=1}^{n} Y_i$ suit une loi de Poisson et déterminez son paramètre.

**1.2** Montrez que la distribution de $(Y_1, \dots, Y_n)$ conditionné par $Y = m$ est la même que celle de $(X_1, \dots X_n)$.

**1.3** Soit $f$ une fonction à $n$ variables et à valeurs dans les réels positifs. Montrez que

$$\mathbf{E}\{f(X_1, \dots, X_n)\} \leq e\sqrt{m}\mathbf{E}\{f(Y_1, \dots, Y_n)\} \ .$$

Vous pourrez utiliser que $m! < e\sqrt{m}\left(\frac{m}{e}\right)^m$.

**1.4** Appelons *cas de Poisson* les événements se produisant quand le nombre de boules dans les bacs sont prises comme variables de Poisson indépendantes d'espérance $m/n$, et *cas de Boules et Bacs* quand on jète $m$ boules dans $n$ bacs aléatoirement et indépendamment.
Quelle fonction $f$ choisisseriez-vous dans le résultat précédent pour conclure que :
Chaque événement qui arrive avec probabilité $p$ dans le cas de Poisson, arrive avec probabilité au plus $pe\sqrt{m}$ dans le cas Boules et Bacs.

**1.5** Retrouvez la borne inférieure sur la charge maximum dans le cas où $m = n$, en utilisant l'approximation de Poisson.
Plus précisément, montrez que celle-ci est au moins $\ln n / \ln \ln n$ avec probabilité $1 - 1/n$ pour $n$ suffisamment grand.

### Exercice 2 - Numéros de Sécurité Sociale

Le numéro de sécurité sociale américain est composé de 9 chiffres. Les quatre derniers chiffres servent comme code secret. Supposons que les chiffres soient choisis indépendants et uniformément aléatoirement (et sans vérifier si un même numéro a déjà été donné).

**2.1** On dénote par $n$ le nombre d'individus. Événement : Au moins deux personnes ont le même code secret. Pour quelle valeur de $n$ cet événement est plus probable d'arriver que de ne pas arriver ?

**2.2** Même question avec l'événement : Au moins deux personnes ont le même numéro de sécurité sociale.

### Exercice 3 - Uniformisation

On dispose d'un générateur de bits aléatoires indépendants suivant la même loi de Bernoulli de paramètre $0 < p < 1$. Mais on ne connait pas $p$. Proposez un algorithme qui utilise cette source

d'aléa pour retourner un bit uniforme (i.e., suivant un Bernoulli de paramètre 1/2) et analysez l'espérance du nombre d'appels au générateur.

### Exercice 4 - Filtre de Bloom

On souhaite vérifier que des mots de passe ne sont pas trop faciles à cracker en maintenant un dictionnaire de mots inacceptables. Quand un utilisateur rentre un mot de passe, notre application doit tester son appartenance à l'ensemble des mots inacceptables.
Un filtre de Bloom consiste en un tableau de $n$ bits, $A[0]$ to $A[n-1]$ initialement mis à 0. Le filtre de Bloom utilise $k$ fonctions de hachage aléatoires indépendantes $h_1, \ldots, h_k$. On fait l'hypothèse que ces fonctions de hachage associe à chaque élément de l'univers un nombre aléatoire uniforme dans $\{0, \ldots, n-1\}$. Soit $F = \{f_1, \ldots, f_m\}$ l'ensemble de mots de passe inacceptables.
L'étape de pré-processing est la suivante : Pour chaque élément $f_j \in F$, les bits $A[h_i(f_j)]$ sont mis à 1 pour tout $1 \leq i \leq k$. Observez qu'une entrée peut être mise à 1 plusieurs fois ; mais seulement la première affectation a un effet.
Pour vérifier si un élément $x$ est dans $F$, on teste que toutes les entrées $A[h_i(x)]$ sont à 1 (pour tout $1 \leq i \leq k$). Si ce n'est pas le cas, on conclut que $x \notin F$. Il est facile de voir qu'on ne peut pas avoir de faux négatifs. Mais il peut y avoir des faux positifs : On peut avoir $A[h_i(x)] = 1$ (pour tout $1 \leq i \leq k$) et avoir tort de conclure que $x \in F$.

**4.1** Soit $X$ le nombre de positions dans $A$ qui contiennent un 0 après pré-processing. Que vaut $\mathbf{E}\{X/n\}$ ?

**4.2** Pour simplifier, supposons que $X = pn$ pour $p = e^{-km/n}$. Quelle est la probabilité $P$ d'un faux positif ? Choisissez $k$ pour minimiser $P$. Quelle valeur de $p$ minimise $P$ ?

**4.3** Reconsidérez (et justifiez) notre hypothèse $X = pn$. Utilisez l'approximation de Poisson pour borner $\mathbf{P}\{|X - np| \geq \varepsilon n\}$.

### Exercice 5 - Nombre chromatique

Le nombre chromatique $\chi(G)$ est le plus petit entier naturel $\chi$ tel qu'il est possible de colorer les sommets de $G$ avec $\chi$ couleurs sans avoir les deux extrémités d'une arête de la même couleur. On pourrait penser que si le graphe n'a pas de cycles courts, alors $\chi(G)$ est borné. Cela s'avère pourtant faux.
Prouvez que pour tout entier $k \geqslant 2$, il y a un graphe $G$ sans triangle de nombre chromatique $k$.

### Exercice 6 - Graphe connexe

Un graphe non orienté est dit *connexe* s'il existe un chemin entre toute paire de sommets. Montrez que si $p = (2+\varepsilon)\log n/n$ pour un $\varepsilon > 0$, alors la probabilité qu'un graphe tiré aléatoirement de $G_{n,p}$ est connexe tend vers 1 quand $n$ tend vers l'infini.