TD 03 – Moments d'une variable aléatoire (corrigé)

Exercice 1. *Aiguille de Buffon (le retour)*

Aiguille de Buffon [1]

Considérons l'expérience consistant à jeter une ficelle longueur a sur un parquet composé des planches parallèles de même largeur ℓ.

L'objectif de cet exercice est de calculer le nombre moyen de points d'intersections entre la ficelle et les rainures du parquet, en utilisant juste la linéarité de l'espérance (sans trigonométrie).

1. Soit X la variable aléatoire représentant de nombre de points d'intersection entre la ficelle et les rainures du parquet. Montrer que
$$\mathbf{E}[X] = c \cdot a,$$
pour une certaine constante c que l'on ne demande pas de calculer. *Indice : on pourra décomposer la ficelle en petite tronçons de longueur $\varepsilon < \ell$, que l'on approximera par des segments.*

☞ On utilise l'indice et on décompose la ficelle en a/ε tronçons, de longueur ε (on choisit $\varepsilon < \ell$ et tel que a/ε soit un entier).On prend ε suffisament petit pour que chaque tronçon ressemble à peu près à une droite, et comme $\varepsilon < \ell$, chaque petit tronçon ne peut rencontrer une rainure qu'au plus une fois. On note X_i la variable aléatoire valant 1 si le bout i tombe à cheval sur une rainure et 0 sinon (on a vu qu'il n'y a pas d'autres choix possibles). On a alors $X = \sum_i X_i$. De plus, les X_i suivent tous la même loi (chaque bout de corde à la même probabilité de tomber à un endroit donner). On a donc, pour tout i, $\mathbf{E}[X_i] = p_\varepsilon$ pour une certaine constante $0 \leq p_\varepsilon \leq 1$. Par linéarité de l'espérance, on en déduit

$$\mathbf{E}[X] = \sum_i \mathbf{E}[X_i] = p_\varepsilon \cdot a/\varepsilon = (p_\varepsilon/\varepsilon) \cdot a.$$

La constante de linéarité est donc $c = (p_\varepsilon/\varepsilon)$, qui ne dépend pas de a (à part pour le fait que ε est choisit tel que a/ε soit entier, mais si ε est suffisamment petit ça ne doit pas poser de problème).

2. Trouver une forme qui, où qu'elle soit lancée, aura toujours le même nombre d'intersections avec les rainures du plancher. En déduire la constante c.

☞ Un cercle de diamètre ℓ aura toujours 2 intersections avec les rainures du plancher. Même si on conditionne X par l'événement "la ficelle tombe en formant un cercle", on voit, en utilisant le même raisonnement que dans l'exercice précédent que l'espérance du nombre de points d'intersections est toujours $c \cdot a$. Pour le cas du cercle de diamètre ℓ, on a $a = \pi\ell$, et $\mathbf{E}[X] = 2$. D'où $c = \frac{2}{\pi\ell}$.

Exercice 2. *Running Time*

Soit $\mathcal{A}$ un algorithme qui prend en entrée une chaîne de n bits et dont l'espérance du temps d'exécution est $\mathcal{O}(n^2)$ si l'entrée est choisie aléatoirement de manière uniforme.

1. Soit $f(n)$ une fonction tendant vers $+\infty$ avec n. Montrer que la probabilité que le temps d'exécution soit supérieur à $n^2 f(n)$ tend vers zero quand n tend vers l'infini.

☞ Le but ici est d'utiliser l'inégalité de Markov. Soit X le temps d'exécution de l'algorithme.

$$\mathbf{P}\left\{X \geq n^2 \cdot f(n)\right\} \leq \frac{\mathbf{E}[X]}{n^2 \cdot f(n)} \leq \frac{c \cdot n^2}{n^2 \cdot f(n)} \leq \frac{c}{f(n)} \longrightarrow_{n\to\infty} 0\,.$$

2. Que pouvons nous en déduire sur le temps d'exécution dans le pire cas ?

☞ Pour avoir une borne supérieure sur le temps dans le pire cas, on utilise le fait que les entrées sont distribuées uniformément. Comme chaque entrée est choisie avec probabilité $1/2^n$, on a que si $\mathbf{P}\{X \geq t\}$ est non nul, elle doit être au moins égale à $1/2^n t$ (car au moins une entrée donnera un temps de calcul supérieur à t). On a vu à la question précédente que

$$\mathbf{P}\left\{X \geq n^2 \cdot f(n)\right\} \leq \frac{c}{f(n)}.$$

Pour que cette quantité soit inférieure à $1/2^n$, il faut que $f(n) \geq c2^n$. On en déduit que le temps d'exécution dans le pire cas est borné par $cn^2 2^n = O(n^2 2^n)$.

Exercice 3. *Points fixes d'une permutation*

On note $\mathfrak{S}_n$ le groupe des permutations de $\{1, \ldots, n\}$.

[1] Georges-Louis Leclerc de Buffon (1707–1788), scientifique et écrivain français.

1. Écrivez un algorithme qui prend en entrée un entier n et retourne une permutation aléatoire de loi uniforme sur $\mathfrak{S}_n$. Vous pouvez utiliser la fonction `RandInt` qui prend en entrée un entier m et retourne un entier aléatoire de loi uniforme sur $\{1,\ldots,m\}$. Évidemment, on préfèrera un algorithme aussi rapide que possible.

   ☞ Le plus rapide et élégant est l'algorithme de Knuth ou de Fischer-Yates `https://en.wikipedia.org/wiki/Fisher%E2%80%93Yates_shuffle`

2. On note $F(\sigma)$ le nombre de points fixes d'une permutation σ. Calculez

$$\lim_{n\to+\infty} \mathbb{E}[F(\sigma_n)]$$

   où σ_n suit la loi uniforme sur $\{1,\ldots,n\}$.

   ☞ Par linéarité de l'espérance, on a

$$\mathbb{E}[F(\sigma_n)] = \sum_{k=1}^{n} \mathbb{P}(\sigma_n(k) = k) = 1.$$

3. Calculer la limite, lorsque n tend vers l'infini, de $\mathbf{P}\{F(\sigma_n) = 0\}$. *Indice : on pourra introduire les événements $E_i = \{\sigma_n(i) = i\}$ et utiliser la formule de Poincarré.*

   ☞ Commençons par suivre l'indice, et calculer $\bigcap_{j=1}^{r} E_{i_j}$ pour $1 \le i_1 < \cdots < i_r \le n$. Compter le nombre de permutations de n éléments où $\sigma(i) = i$ revient à compter les permutations de $n-1$ éléments (les $j \neq i$ - on pourrait s'amuser à construire la bijection entre les deux ensembles pour justifier l'égalité).

   Ainsi $\mathbf{P}\{E_i\} = \dfrac{(n-1)!}{n!} = \dfrac{1}{n}$.

   Pour $E_i \cap E_j$, on est ramené aux permutations de $n-2$ éléments.

   Donc $\mathbf{P}\{E_i \cap E_j\} = \dfrac{(n-2)!}{n!} = \dfrac{1}{n(n-1)}$.

   De même, on arrive à $\mathbf{P}\left\{\bigcap_{i_1<\cdots<i_r} E_{i_j}\right\} = \dfrac{1}{n(n-1)\ldots(n-r+1)}$.

   Maintenant qu'on a calculé ces quantités, on remarque que la quantité cherchée est $\mathbf{P}\{F(\sigma_n) = 0\} = 1 - \mathbf{P}\{\bigcup_{1\le i\le n} E_i\}$. Definissons $P_n = \mathbf{P}\{\bigcup_{1\le i\le n} E_i\}$.

   D'après la formule de Poincaré, on a :

$$\begin{aligned} P_n &= \sum_{k=1}^{n} (-1)^{k-1} \sum_{1\le i_1<\cdots<i_k\le n} \mathbf{P}\left\{\bigcap_{1\le j\le k} A_{i_j}\right\} \\ &= \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k!}. \end{aligned}$$

   On reconnaît presque le développement en série $e^{-1} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}$. En fait, on a $P_n - 1 \to -e^{-1}$ donc la limite vaut en fait $1 - e^{-1}$, et $\mathbf{P}\{F(\sigma_n) = 0\}$ tend vers $1/e$ quand n tend vers l'infini.

Exercice 4. *Debiaiser des bits*

Supposons que vous ayez une machine qui produit des bits aléatoires i.i.d. de loi de Bernoulli de paramètre p, mais que vous ne connaissez pas la valeur de $p \in]0,1[$.

1. Proposez un algorithme qui utilise la machine pour produire un bit de loi uniforme sur $\{0,1\}$.

   ☞ L'algorithme naturel est de produire des paires de bits jusqu'à obtenir une paire de bits distincts, et de retourner 0 ou 1 selon que la paire ainsi produite est 01 ou 10. Le nombre de paires de bits générées est une variable aléatoire de loi géométrique de paramètre $q := 2p(1-p)$, en particulier il est fini (et l'algorithme termine) presque sûrement.

2. On souhaite maintenant produire n bits indépendants de loi uniforme sur $\{0,1\}$. Proposez un algorithme, et déterminez une valeur de t telle que la probabilité d'utiliser la machine plus de tn fois soint inférieure à $\frac{1}{100}$ pour n assez grand. Évidemment, plus t est petit, mieux c'est.

   ☞ Si on répète l'algorithme précédent, le nombre d'utilisations N de la machine suit la loi de

$$2(X_1 + \cdots + X_n)$$

   où les variables aléatoires (X_i) sont i.i.d. de loi géométrique de paramètre q. Vous êtes nombreux à avoir majoré $\mathbb{P}(N \ge t)$ par l'inégalité de Markov. C'est correct, mais donne un résultat mauvais : on a $\mathbb{P}(N \ge tn) \le \frac{1}{tp(1-p)}$ ce qui nécessite $t = \frac{100}{p(1-p)}$. Cela peut être amélioré en utilisant l'inégalité de Tchebychev : en effet $\mathbf{E}N = \frac{n}{p(1-p)}$ et $\mathbf{Var}[N] = O(n)$ (peu importe la constante exacte, il suffit de dire qu'une v.a. géométrique a une variance finie), donc pour tout $\varepsilon > 0$,

$$\mathbb{P}(N \geq (1+\varepsilon)\mathbf{E}N) \leq \frac{\mathbf{Var}[N]}{\varepsilon^2(\mathbf{E}N)^2} = O\left(\frac{1}{\varepsilon^2 n}\right).$$

En particulier, tout réel $t > \frac{1}{p(1-p)}$ convient, une amélioration d'un facteur 100...

Le rendement peut être amélioré car l'algorithme de la question 1 gaspille beaucoup de bits sans en extraire d'information. Par exemple, on peut regarder si les paires de (paires de bits identiques) que l'on a jetées sont identiques ou pas, et déclarer que l'on arrête l'algorithme lorsqu'elles sont différentes, en retournant 0 ou 1 selon la situation 00 11 ou 11 00. Cela se prête naturellement à une formulation récursive, que vous trouverez expliquée ici `https://en.wikipedia.org/wiki/Bernoulli_process#Randomness_extraction`. On peut montrer que tout $t > 1/h(p)$ convient, où $h(p) = -p\log_2 p - (1-p)\log_2(1-p)$ est la fonction d'entropie binaire, et que cette borne est optimale. Il est remarquable que pour p proche de $1/2$, le rendement obtenu est proche de 1 !

Exercice 5. *Chebychev d'ordre supérieur*

L'inégalité de Chebyshev utilise la variance d'une variable aléatoire pour borner son écart par rapport à l'espérance. On peut également utiliser des moments d'ordre supérieur.

1. Supposons que l'on ait une variable aléatoire X et un entier pair k pour lequel $\mathbf{E}\left[(X-\mathbf{E}[X])^k\right]$ est finie. Prouver que :

$$\mathbf{P}\left\{|X-\mathbf{E}[X]| > t\sqrt[k]{\mathbf{E}\left[(X-\mathbf{E}[X])^k\right]}\right\} \leq \frac{1}{t^k}.$$

☞ Soit $Y = (X-\mathbf{E}[X])^k$. Supposons dans un premier temps que X n'est pas constante alors Y n'est pas constante égale à 0, et comme Y est positive, on en déduit que $\mathbf{E}[Y] \neq 0$. Par l'inégalité de Markov,

$$\mathbf{P}\left\{Y \geq t^k\mathbf{E}[Y]\right\} \leq \frac{\mathbf{E}[Y]}{t^k\mathbf{E}[Y]} = \frac{1}{t^k}.$$

Or, en passant à la racine k-ieme, on obtient :

$$\mathbf{P}\left\{Y \geq t^k\mathbf{E}[Y]\right\} = \mathbf{P}\left\{\sqrt[k]{Y} \geq t\sqrt[k]{\mathbf{E}[Y]}\right\} = \mathbf{P}\left\{|X-\mathbf{E}[X]| \geq t\sqrt[k]{\mathbf{E}[(X-\mathbf{E}[X])^k]}\right\}$$

où la deuxième égalité est vraie car k est pair. En recombinant les deux relations que l'on a trouvées, on obtient ce qui est demandé :

$$\mathbf{P}\left\{|X-\mathbf{E}[X]| > t\sqrt[k]{\mathbf{E}[(X-\mathbf{E}[X])^k]}\right\} \leq \frac{1}{t^k}.$$

Supposons maintenant que X est constante, alors $X = \mathbf{E}[X]$, et $\mathbf{P}\left\{|X-\mathbf{E}[X]| > t\sqrt[k]{\mathbf{E}[(X-\mathbf{E}[X])^k]}\right\} = 0$, donc le résultat est toujours vrai.

2. Qu'est-ce qui nous empêche d'obtenir une inégalité similaire pour le cas où k est impair ? Trouver un contre exemple pour $k = 1$.

☞ Si k est impair, alors $(X-\mathbf{E}[X])^k$ peut prendre des valeurs négatives, ce qui nous empêche d'appliquer l'inégalité de Markov.

Pour le contre exemple, on peut prendre $X = \pm 1$ avec probabilité $1/2$ pour chaque. On a alors $\mathbf{E}[X] = 0$ et $\mathbf{E}[X-\mathbf{E}[X]] = 0$. Ainsi, pour tout $t > 0$, on a $t\sqrt[1]{\mathbf{E}[(X-\mathbf{E}[X])^1]} = 0$. En particulier, pour $t = 3$, on a $\mathbf{P}\left\{|X-\mathbf{E}[X]| > 3\sqrt[1]{\mathbf{E}[(X-\mathbf{E}[X])^1]}\right\} = \mathbf{P}\{|X| > 0\} = 1 \not\leq \frac{1}{3}$.

Exercice 6. *Coquilles dans un TD*

1. Une feuille de TD contient 4 coquilles. À chaque relecture, une coquille non corrigée est corrigée avec probabilité $1/3$. Les relectures sont indépendantes les unes des autres. Combien de relectures faut-il faire au minimum pour que la probabilité qu'il ne subsiste aucune coquille soit supérieure à 0,9 ?

☞ On a : $\mathbf{P}\{$"La coquille numéro i est corrigée en au plus n relectures"$\}$
$= 1 - \mathbf{P}\{$"Elle n'est pas corrigée après n relectures"$\} = 1 - (\frac{2}{3})^n$.

Donc $\mathbf{P}\{$"Les 4 coquilles sont corrigées en au plus n relectures"$\} = (1-(\frac{2}{3})^n)^4$.

Enfin, puisque n est entier, on obtient : $(1-(\frac{2}{3})^n)^4 \geqslant 0{,}9 \iff n \geqslant 10$ (presque 9, mais rigoureusement c'est 10).

2. À quel autre problème vu en cours vous fait penser la situation précédente ?

☞ C'est le problème du collectionneur de vignettes. La variable X_i comptant le nombre de corrections nécessaires pour corriger la coquille i suit une loi géométrique. La variable qui nous intéresse à la question précédente est $\max(X_i)$. La seule différence avec le problème du collecteur de coupons est qu'ici les variables X_i sont indépendantes (on peut corriger plusieurs coquilles dans la même relecture, alors qu'on ne peut pas avoir plusieurs vignettes avec une vignette).

3. Soit X une variable aléatoire d'espérance $\mu = 10$ et d'écart-type $\sigma = 5$. Montrer que pour tout $n \geq 50$, la probabilité que de l'événement $\{10 - n < X < 10 + n\}$ est au moins égale à 0,99.

☞ On remarque que l'événement en question est égal à $\{|X - 10| < n\}$. Or l'inégalité de Chebyshev nous donne ($n \geq 0$ et $\mu = 10$) :

$$\mathbf{P}\{|X - \mu| \geq n\} \leq \frac{\sigma^2}{n^2}$$

donc $\mathbf{P}\{|X - 10| < n\} \geq 1 - \frac{25}{2500} = 0{,}99$.

Exercice 7. *Covariance*

On définit la covariance de 2 variables aléatoires X et Y par :

$$\mathrm{Cov}(X, Y) = \mathbb{E}\Big((X - \mathbf{E}[X])(Y - \mathbf{E}[Y])\Big).$$

1. Montrer que si X et Y sont indépendants, alors $\mathrm{Cov}(X, Y) = 0$.

☞ Donnons une autre expression pour la covariance :

$$\begin{aligned} \mathrm{Cov}(X, Y) &= \mathbf{E}[XY - X\mathbf{E}[Y] - Y\mathbf{E}[X] + \mathbf{E}[X]\mathbf{E}[Y]] \\ &= \mathbf{E}[XY] - 2\mathbf{E}[X]\mathbf{E}[Y] + \mathbf{E}[X]\mathbf{E}[Y] \\ &= \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y] \end{aligned}$$

Or, si X et Y sont indépendants, on a $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$ d'où $\mathrm{Cov}(X, Y) = 0$.

2. Qu'en est il de la réciproque ?

☞ La réciproque est fausse.

Prendre X tel que $\mathbf{P}\{X = -1\} = \mathbf{P}\{X = 1\} = \frac{1}{2}$ et $Y = 0$ si $X = -1$ et $Y = \pm 1$ avec probabilité 1/2 si $X = 1$. Les variables X et Y ne pas indépendantes, car savoir $Y = 0$ permet de déterminer X. Mais on a $\mathbf{E}[XY] = 0 = \mathbf{E}[X]\mathbf{E}[Y]$.

3. Exprimer $\mathbf{Var}[X + Y]$ en fonction de $\mathbf{Var}[X]$, $\mathbf{Var}[Y]$ et $\mathrm{Cov}(X, Y)$. En déduire une formule plus générale pour $\mathbf{Var}[\sum_{i=1}^n X_i]$.

☞

$$\begin{aligned} \mathbf{Var}[X + Y] &= \mathbf{E}\left[(X + Y)^2\right] - \mathbf{E}[X + Y]^2 \\ &= \mathbf{E}\left[X^2\right] + 2\mathbf{E}[XY] + \mathbf{E}\left[Y^2\right] - \mathbf{E}[X]^2 - 2\mathbf{E}[X]\mathbf{E}[Y] - \mathbf{E}[Y]^2 \\ &= \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathrm{Cov}(X, Y). \end{aligned}$$

Plus généralement :

$$\mathbf{Var}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbf{Var}[X_i] + 2 \sum_{1 \leq i < j \leq n} \mathrm{Cov}(X_i, X_j).$$

4. On se donne X et Y de variances finies, et on définit :

$$r(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

Montrer que $|r(X, Y)| \leq 1$. Que pensez vous du cas où on a l'égalité ?

☞ On commence par remarquer qu'on peut se ramener au cas où X et Y sont centrées (i.e. $\mathbf{E}[\cdot] = 0$) réduites (i.e. $\sigma(\cdot) = 1$) car $r(X, Y)$ reste invariant.

Ensuite, on constate que

$$0 \leq \mathbf{E}\left[(X + \lambda Y)^2\right] = \underbrace{\mathbf{E}\left[X^2\right]}_{1} + 2\lambda \underbrace{\mathbf{E}[XY]}_{r(X,Y)} + \lambda^2 \underbrace{\mathbf{E}\left[Y^2\right]}_{1}$$

pour tout réel λ. Le polynôme de degré 2 en λ admet au plus une racine réelle, donc son discriminant est négatif ou nul :

$$4r(X, Y)^2 - 4 \leq 0 \Rightarrow |r(X, Y)| \leq 1.$$

Si $|r(X, Y)| = 1$, alors X et Y sont reliées (presque partout) par une relation affine. Par exemple, si X, Y centrées réduites et $r(X, Y) = 1$, on a $\mathbf{E}\left[(X + \lambda Y)^2\right] = (1 + \lambda)^2$ donc pour $\lambda = -1$ on en déduite $\mathbf{E}\left[(X - Y)^2\right] = 0$ et $X = Y$ presque sûrement.

Dans le cas général, si $r(X, Y) = \varepsilon \in \{-1, 1\}$, on a :

$$\frac{Y - \mathbf{E}[Y]}{\sigma(Y)} = \varepsilon \frac{X - \mathbf{E}[X]}{\sigma(X)} \text{ p.s.}$$

NB : $r(X, Y)$ s'appelle le coefficient de corrélation linéaire.