Examen final

EI-SE3

Introduction aux Probabilités

Benjamin Auder

18 mai 2015

Exercice 0 [2 points]

On considère (uniquement) deux guichets dans une gare SNCF, dont au moins un est ouvert. On note A (resp. B) l'évènement "le guichet 1 (resp. 2) est ouvert". Une étude statistique a permis d'établir les probabilités suivantes : $\mathrm{P}(A) = 0.55$, et $\mathrm{P}(B) = 0.7$.

Il y a beaucoup d'affluence dans cette gare : au moins un client attend son tour devant chaque guichet ouvert. On note C le nombre total de clients présents aux guichets.

(a) Calculez $\mathrm{P}(A \cup B)$. [1]

> **Solution :**
> "…dont au moins un est ouvert" $\Rightarrow \mathrm{P}(A \cup B) = 1$.

(b) Déterminez $\mathrm{P}(C = 3 | A \cap B)$. [1]

> **Solution :**
> "Au moins un client attend son tour devant chaque guichet ouvert" $\Rightarrow$ Il y a au moins deux clients pour chaque guichet ouvert : un dont on s'occupe, et un qui attend. Donc C sachant $A \cap B$ est supérieur ou égal à 4.
> $\Rightarrow \mathrm{P}(C = 3 | A \cap B) = 0$.

Exercice 1 [11 points]

Cet exercice porte sur les déplacements de n pigeons dans un groupe d'immeubles résidentiels. Ces oiseaux apprécient tout particulièrement les rebords de fenêtres, où ils se cachent derrière les volets. On effectue les hypothèses suivantes.

- Tous les pigeons partent initialement du même point au même instant $t_0 = 0$.
- Il y a m fenêtres numérotées $1, 2, \ldots, m$, parcourues dans cet ordre par chaque pigeon. Une fois arrivé en m on revient vers 1 (*note* : $m \geq 2$).
- Le temps de vol T du point initial à la première fenêtre, puis de la fenêtre i à $(i \mod m) + 1$ suit une loi uniforme $\mathcal{U}(a, b)$ avec $0 < a < b$.

Un pigeon supporte de se poser si quelques congénères sont déjà présents, mais préfère être seul. On note N le nombre de pigeons sur un rebord de fenêtre, et P l'évènement "le pigeon se pose". Alors

- $\mathrm{P}(P | N = 0) = p_0 \in]0, 1[$
- $\mathrm{P}(P | N = 1) = p_1 \in]0, p_0[$
- $\mathrm{P}(P | N = 2) = p_2 \in]0, p_1[$
- $\mathrm{P}(P | N \geq 3) = 0$

Un pigeon reste posé sur un rebord de fenêtre pendant une durée R. suivant une loi exponentielle : $R \sim \mathcal{E}(\lambda)$ avec $\lambda > 0$.

Les questions a) à c) portent sur les fenêtres *du premier tour uniquement*.

(a) *Cas* $n = 1$. [1]

Déterminez le temps moyen nécessaire au pigeon pour parvenir à la k^{eme} fenêtre.

Solution :

La VA donnant le temps d'atteinte de la k^{eme} fenêtre est

$$F = \sum_{i=1}^{k} T_i + \sum_{i=1}^{k-1} P_i R_i \,,$$

avec T_i le temps de vol entre les fenêtres $i-1$ et i, R_i le temps passé à la fenêtre i et P_i une VA de loi de Bernouilli de paramètre p_0. P_i et R_i étant indépendants, $\mathrm{E}[P_i R_i] = \mathrm{E}[P_i]\mathrm{E}[R_i]$. On obtient donc

$$\begin{aligned} \mathrm{E}[F] &= k\mathrm{E}[T] + (k-1)\mathrm{E}[P]\mathrm{E}[R] \\ &= k\frac{a+b}{2} + \frac{(k-1)p_0}{\lambda} \end{aligned}$$

(b) *Cas* $n = 2$. [3]

Calculez la probabilité que les deux pigeons se retrouvent posés ensemble sur le rebord de la première fenêtre. Vérifiez la cohérence du résultat.

Solution :

Pour que les deux pigeons se retrouvent posés ensemble (pendant une durée non nulle), il faut que le premier arrivé s'y pose et qu'il y reste au moins jusqu'à ce que le second arrive, ce dernier se posant également. Cet évènement s'écrit

$$E : \min(T_1, T_2) + R > \max(T_1, T_2) \cap P_1 \cap P_2 \,,$$

avec T_1 et T_2 les temps d'atteinte de la première fenêtre, P_1 et P_2 correspondant aux évènements "le k$^{\text{eme}}$ pigeon se pose". Comme $\max(T_1, T_2) - \min(T_1, T_2) = |T_2 - T_1|$, on peut reformuler $E : |T_1 - T_2| < R \cap P_1 \cap P_2$.

$$\begin{aligned} \mathrm{P}(E) &= \mathrm{P}(P_2 \,|\, |T_1 - T_2| < R \cap P_1)\mathrm{P}(|T_1 - T_2| < R \cap P_1) \\ &= p_1 \mathrm{P}(P_1 \,|\, |T_1 - T_2| < R)\mathrm{P}(|T_1 - T_2| < R) \\ &= p_0\, p_1 \mathrm{P}(|T_1 - T_2| < R) \end{aligned}$$

On détermine ensuite la loi de $\Delta = |T_1 - T_2|$:

$$\begin{aligned} \mathrm{P}(\Delta < x \in [0, b-a]) &= \frac{2}{(b-a)^2} \int_{t_1=a}^{b} \int_{t_2=a}^{t_1} \mathbb{1}_{t_1 - t_2 < x} dt_1 dt_2 \\ &= \frac{2x}{b-a} - \left(\frac{x}{b-a} \right)^2 \text{ après calculs.} \end{aligned}$$

Donc

$$\begin{aligned} \mathrm{P}(\Delta < R) &= \frac{2}{b-a} \int_{t=0}^{b-a} \int_{u=t}^{+\infty} \left(1 - \frac{t}{b-a} \right) \lambda e^{-\lambda t} dt\, du \\ &= 2 \left(\frac{e^{-\lambda\delta} - 1 + \lambda\delta}{\lambda^2 \delta^2} \right) \text{ avec } \delta = b - a \ , \end{aligned}$$

les calculs s'effectuant par changements de variables successifs, et en remarquant que $\frac{d}{dx}(e^x(x+1)) = xe^x$.

On vérifie finalement que

- $\mathrm{P}(\Delta < R|\overline{C}) \underset{\lambda \to +\infty}{\longrightarrow} 0$: si les pigeons ne restent posés qu'un instant, ils ne se croisent probablement pas ;
- $\mathrm{P}(\Delta < R|\overline{C}) \underset{\delta\lambda \to 0}{\longrightarrow} 1$: si les temps de vols sont tous quasi identiques ou si les temps de poses moyens sont très grands, on est à peu près sûr que les pigons vont se croiser (s'ils se posent).

Cette dernière limite s'obtient en écrivant le DL $e^{-\lambda\delta} = 1 - \lambda\delta + \frac{\lambda^2\delta^2}{2} + o(\lambda^3\delta^3)$.

(c) *On se place à partir de maintenant dans le cas général (n quelconque).* [2]

1. Déterminez la loi du temps d'atteinte de la première fenêtre, noté Z_n. (La fenêtre est atteinte quand un des n pigeons y parvient).
2. Montrez que Z_n converge en probabilité vers une variable aléatoire à expliciter. A-t-on aussi convergence preque sûre ?

Solution :

La première fenêtre est atteinte quand le pigeon le plus rapide y parvient. On cherche donc la loi du plus petit temps de vol, qui s'écrit $Z_n = \min(T_1, \ldots, T_n)$ avec T_i = temps de vol du i^{eme} pigeon. Notons F_n (resp. f_n) la fonction de répartition (resp. de densité) de Z_n.

$$\begin{aligned} 1 - F_n(t \in [a, b]) &= \mathrm{P}(\min(T_1, \ldots, T_n) > t) \\ &= \mathrm{P}(T_1 > t \cap \cdots \cap T_n > t) \\ &= \mathrm{P}(T_1 > t) \times \cdots \times \mathrm{P}(T_n > t) \quad \text{(hypothèse d'indépendance)} \\ &= \mathrm{P}(T_1 > t)^n \quad \text{(VAs de même loi)} \\ &= \left(1 - \frac{t-a}{b-a} \right)^n \end{aligned}$$

Ainsi $F_n(t) = 1 - \left(1 - \frac{t-a}{b-a}\right)^n$, et $f_n(t \in [a,b]) = \frac{d}{dt}F_n(t) = \frac{n}{b-a}\left(1 - \frac{t-a}{b-a}\right)^{n-1}$.

F_n converge ponctuellement vers $\mathbb{1}_{]a,b]}$, donc la VA candidate X_a est constante égale à a. On a alors, pour $\varepsilon \in]0, b-a]$

$$\begin{aligned} \mathrm{P}(|Z_n - X_a| > \varepsilon) &= \mathrm{P}(Z_n \in [a+\varepsilon, b]) \\ &= 1 - F_n(X_a + \varepsilon) \\ &= \left(1 - \frac{\varepsilon}{b-a}\right)^n \\ &\underset{n\to+\infty}{\longrightarrow} 0 \ , \end{aligned}$$

ce qui prouve la convergence en probabilité.
La série de terme général $\mathrm{P}(|Z_n - X_a| > \varepsilon)$ converge vers $\frac{b-a}{\varepsilon} - 1$ (suite géométrique), donc il y a aussi convergence presque sûre.

(d) Montrez que tous les pigeons parviennent à se poser (au moins une fois). [3]

Solution :

Notons $A_{k,t}$ l'évènement "il y a au plus deux pigeons sur la fenêtre k à l'instant t". Soit $k_0 \in [\![1, m]\!]$. $A_{k_0,0}$ est réalisé car il n'y a pas de pigeons sur les fenêtres à l'instant 0. Posons $t_0 = 0$. Si à l'instant $t_0 + bm$ il y a au plus deux pigeons sur la fenêtre k_0, alors on pose $t_1 = t_0 + bm$. Sinon, les temps de poses suivant des lois exponentielles ils sont finis avec probabilité 1 : depuis $t = bm$ un pigeon décolle au bout d'un temps $\tilde{t} - t < +\infty$. On pose dans ce cas $t_1 = t_0 + bm + \tilde{t}$, puis on applique la même démarche pour obtenir t_2 à partir de t_1, ...etc.
Une famille infinie d'évènements A_{k_0,t_i} tous réalisés est alors définie par récurrence, avec la condition $t_{i+1} - t_i \geq bm$.

$$\exists\, (t_i \in \mathbb{R}^+)_{i\in\mathbb{N}} \quad / \quad \forall i \in \mathbb{N}\,,\ A_{k_0,t_i} \text{ et } t_{i+1} - t_i \geq bm$$

Soit $\alpha \in [\![1, n]\!]$ le numéro d'un pigeon quelconque. S'il existe un indice i_0 tel que le pigeon α soit posé sur la fenêtre k_0 en t_{i_0}, cela signifie qu'il se pose au moins une fois. On suppose donc que pour chaque indice i le pigeon α n'est pas à la fenêtre k_0 au temps t_i. Notons alors $B_{\alpha,i}$ l'évènement "depuis l'instant t_i et jusqu'à l'arrivée du pigeon α qui se pose à la fenêtre k_0, tout pigeon parvenant à cette dernière ne s'y pose pas". Observons que les $B_{\alpha,i}$ ne sont a priori pas indépendants, car si le pigeon se pose pour un certain indice i_0 cela influe sur la probabilité de l'évènement B_{α,i_0+1}. L'idée consiste à appliquer le lemme de Borel-Cantelli à une certaine sous-famille d'évènements $B_{\alpha,i}$, dont on aura montré l'indépendance et la divergence de la somme des probabilités.

Plaçons-nous au temps t_{i_0} pour un i_0 fixé quelconque. Combien de tentatives de pose à la fenêtre k_0 s'effectuent avant que le courageux pigeon α y parvienne ? Dans

le pire cas il faut passer par $m-1$ fenêtres en volant le plus lentement possible ($T=b$) à chaque fois, tandis que les autres pigeons volent le plus rapidement possible ($T=a$) en partant juste avant la fenêtre k_0. Ces derniers passent chacun au plus $\left\lfloor \frac{b}{a} \right\rfloor$ fois par la fenêtre en question.
En notant $R_{\alpha,i}$ la variable aléatoire donnant le nombre de passages à la fenêtre k_0 entre t_i et l'instant d'arrivée du pigeon α en k_0,

$$\forall r > (n-1)\left\lfloor \frac{b}{a} \right\rfloor, \quad \mathrm{P}(R_{\alpha,i}=r)=0\,.$$

$R_{\alpha,i}$ ne prend donc qu'un nombre fini de valeurs. Comme cet évènement se répète à l'infini, $\exists\, r_0 \in \mathbb{N}$ tel qu'une infinité de réalisations de $R_{\alpha,i}$ soient égales à r_0. On note I l'ensemble des indices correspondants, et on se restreint dorénavant à cet ensemble *défini a posteriori*.

À chaque i dans I on fait correspondre le vecteur $v_i \in [\![-1,3]\!]^{m-1}$, dont la composante $v_i^{(j)}$ est égale au nombre de pigeons posés sur la j^{eme} fenêtre parcourue par le pigeon d'indice α depuis l'instant t_i. Lorsque j correspond à k_0 ou après, $v_i^{(j)}$ vaut -1. v_i ne prend qu'un nombre fini de valeurs : $\exists\, v^* \in [\![-1,3]\!]^{m-1}$ tel qu'une infinité d'indices de I vérifient $v_i = v^*$. On note J ce dernier ensemble d'indices, et on s'y restreint dorénavant.

Observons que pour chaque $i \in \mathbb{N}$, le temps de passage correspondant à $B_{\alpha,i}$ est strictement compris entre t_i et t_{i+1} : en effet le temps de voyage est majoré par bm et on a imposé $t_{i+1}-t_i > bm$. Les évènements $B_{\alpha,i}, i \in J$ correspondent donc à des scénarios identiques ayant lieu sur des plages temporelles ne s'intersectant pas : il sont indépendants, et on peut minorer leur probabilité commune :

$$\mathrm{P}(B_{\alpha,i}, i \in J) \geq p_2\, q_0^{r_0+m-1} > 0\,,$$

avec $q_0 = 1-p_0$.

$$\sum_{i \in J} \mathrm{P}(B_{\alpha,i}) = +\infty$$

On est dans les conditions d'application du lemme de Borel-Cantelli :
$\mathrm{P}(\overline{\lim} B_{\alpha,i}, i \in J) = 1$, ce qui signifie qu'avec probabilité 1, le pigeon numéro α se pose une infinité de fois sur la fenêtre k_0. En conclusion, chaque pigeon se pose − presque sûrement − une infinité de fois sur chaque fenêtre.

Note : le raisonnement ci-dessus n'est plus valide si $p_0 = p_1 = p_2 = 1$.
Le résultat reste tout de même vrai (exercice pour les vacances !)

Bonus : programme en langage R simulant l'expérience.

(e) On considère qu'une fois posés, les pigeons s'endorment pour une durée infinie. À quelles(s) condition(s) parviennent-ils tous à dormir ? [2]

Solution :

Supposons que tous les pigeons dorment. Alors nous avons n pigeons répartis sur m fenêtres avec la contrainte "pas plus de 3 pigeons par fenêtre". Il y a donc au plus $3m$ pigeons : on obtient la condition $n \leq 3m$ (ou encore $m \geq \lceil \frac{n}{3} \rceil$).

Réciproquement, choisissons $n \leq 3m$. On se place à un temps $t \in \mathbb{R}^+$ arbitraire, et suppose que $k < n$ pigeons se sont endormis. $3m > k$, donc il reste au moins une fenêtre comportant au plus deux pigeons endormis. Soit f_0 le numéro d'une telle fenêtre. Raisonnons par l'absurde : si les $n-k$ pigeons restant ne s'endorment jamais, alors ils passent une infinité (dénombrable) de fois à la fenêtre f_0. À chaque passage ils ont une probabilité fixée, supérieure ou égale à $p_2 > 0$ de se poser. Soit α le numéro d'un de ces $n-k$ pigeons insomniaques. On note $B_{\alpha,i}$ l'évènement "le pigeon d'index α se pose à la fenêtre f_0 lors de son i^{eme} passage".

Afin d'obtenir l'indépendance des $B_{\alpha,i}$ il faut alors légèrement modifier la règle du jeu : on décide qu'un pigeon posé (non chassé) redécolle immédiatement. On peut ainsi appliquer le lemme de Borel-Cantelli, étant donné que

$$\sum_{i=1}^{+\infty} \mathrm{P}(B_{\alpha,i}) = +\infty\,.$$

$\Rightarrow$ le pigeon numéro α finit $-$ presque sûrement $-$ par se poser sur le rebord de la fenêtre f_0 : il s'endort au bout d'un temps fini. Cela étant en contradiction avec l'hypothèse (absurde), on en déduit qu'au moins un parmi les $n-k$ oiseaux restant parvient à s'endormir en un temps t_1 fini. Par récurrence immédiate, tous les pigeons parviennent à s'endormir.

Remarque : on peut voir la position du pigeon numéro α à un instant t comme une variable aléatoire $X_{\alpha,t}$ de support $[\![0, m]\!]$ (prenant la valeur 0 si le pigeon est en vol). La fonction qui à α et t associe la VA $X_{\alpha,t}$ est un cas particulier processus aléatoire, indexé par $[\![1, n]\!] \times \mathbb{R}^+$. Ce processus n'est pas évident à étudier car les trajectoires $\tau_\alpha : t \mapsto X_{\alpha,t}$ sont dépendantes (du fait des contraintes imposées).

Exercice 2 [11 points]

On dispose d'une pièce de monnaie truquée, qui tombe sur pile avec probabilité $p \in]0, 1[$, et sur face avec probabilité $1 - p$.

(a) *Dans cette question seulement, on suppose p inconnu.* [1]
Comment déterminer p empiriquement ? Justifiez.

Solution :
Soit X_p une variable aléatoire prenant la valeur 1 lorsque la pièce truquée tombe sur pile, et 0 sinon. X_p suit la loi de Bernouilli de paramètre p : $\mathrm{E}[X_p] = p$. Considérons alors la variable aléatoire $Z_n = \frac{1}{n}\sum_{i=1}^{n} X_p^{(i)}$ où les $X_p^{(i)}$ sont indépendantes et de même loi que X_p. On obtient une réalisation de la VA Z_n en lançant n fois notre pièce truquée. Cette réalisation approxime de mieux en mieux p au fur et à mesure que n augmente, car d'après la loi des grands nombres

$$Z_n \xrightarrow{p.s.} \mathrm{E}[X_p] = p\,.$$

(On peut aussi invoquer directement le théorème de Bernouilli).

(b) Comment simuler un lancer de pièce équilibrée à partir de la pièce truquée ? [2]

Solution :
On reprend la définition de X_p de la réponse précédente. Posons alors

$$X = X_p^{(1)} - X_p^{(2)}\,,$$

avec $X_p^{(1)}$ et $X_p^{(2)}$ deux variables aléatoires indépendantes de même loi que X_p.
X est à valeurs dans $\{-1, 0, 1\}$, et

$$\begin{cases} \mathrm{P}(X=-1) = \mathrm{P}(X_p^{(1)} = 0 \cap X_p^{(2)} = 1) = p(1-p) \\ \mathrm{P}(X=1) = \mathrm{P}(X_p^{(1)} = 1 \cap X_p^{(2)} = 0) = p(1-p) \end{cases}$$

On a donc $\mathrm{P}(X=-1|X\neq 0) = \mathrm{P}(X=1|X\neq 0) = \frac{1}{2}$.

Soit $(Z_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes de même loi que X.
Définissons finalement la VA Z comme suit :

$$Z = Z_{n^*} \text{ avec } n^* = \min\{n \in \mathbb{N}^* \,/\, Z_n \neq 0\}\,.$$

Z a pour valeur la première réalisation de X non nulle lors d'une série de deux lancers de la pièce truquée (il faut $2n^*$ lancers). Alors

$$
\begin{aligned}
\mathrm{P}(Z=1) &= \mathrm{P}\left(\bigsqcup_{n\in\mathbb{N}^*}\left[\bigcap_{i<n}\{Z_i=0\}\cap\{Z_n=1\}\right]\right) \\
&= \sum_{n=1}^{+\infty}\mathrm{P}\left(\bigcap_{i<n}\{Z_i=0\}\cap\{Z_n=1\}\right) \\
&= \sum_{n=1}^{+\infty}\left(\prod_{i=1}^{n-1}\mathrm{P}(Z_i=0)\right)\mathrm{P}(Z_n=1) \\
&= \sum_{n=1}^{+\infty}\mathrm{P}(X=0)^{n-1}\mathrm{P}(X=1) \\
&= \frac{\mathrm{P}(X=1)}{1-\mathrm{P}(X=0)} = \frac{\mathrm{P}(\{X=1\}\cap\{X\neq 0\})}{\mathrm{P}(X\neq 0)} \\
&= \mathrm{P}(X=1|X\neq 0) = \frac{1}{2}
\end{aligned}
$$

en utilisant successivement le fait que les évènements $n^*=k$ soient deux à deux disjoints, l'indépendance des épreuves Z_n, et le fait qu'elles soient de même loi.

Note : on aurait pu prendre d'autres expressions pour X, l'essentiel étant que deux des probabilités $\mathrm{P}(X=k)$ résultantes soient égales. Par exemple $X = 2X_p^{(1)} + X_p^{(2)}$ convient (écriture en base 2, $\mathrm{P}(X=1)=\mathrm{P}(X=2)$).

(c) Soient $b \in [\![2,+\infty[\![$, et $(X_n)_{n\in\mathbb{N}^*}$ une famille de variables aléatoires indépendantes, toutes de loi uniforme sur $[\![0,b-1]\!]$. Montrez que la série $S=\sum_{n=1}^{+\infty}\frac{X_n}{b^n}$ converge presque sûrement, et que S suit une loi uniforme sur $[0,1]$. [3]

Solution :
Posons $R_{n,m}=\sum_{k=n+1}^{m}\frac{X_k}{b^k}$ les restes partiels, avec $n<m$. Soit $\omega\in\Omega$ avec Ω l'espace des réalisations de la suite de VA (X_n).

$$
\begin{aligned}
|R_{n,m}(\omega)| &= R_{n,m}(\omega) \text{ (termes positifs)} \\
&= \sum_{k=n+1}^{m}\frac{\omega_k}{b^k} \\
&\leq \sum_{k=n+1}^{m}\frac{b-1}{b^k} \text{ (inégalités terme à terme)} \\
&= \frac{b-1}{b^n} - \frac{b-1}{b^m} \\
&\underset{m\to+\infty}{\longrightarrow} \frac{b-1}{b^n}
\end{aligned}
$$

Ainsi le reste $R_n(\omega)=\sum_{k=n+1}^{+\infty}\frac{\omega}{b^k}$ est majoré par $\frac{b-1}{b^n}$ qui tend vers zéro :

$S(\omega)$ converge pour tout $\omega \in \Omega$, donc a fortiori presque sûrement.

Tout $x \in [0,1]$ peut s'écrire comme une réalisation $S(\omega)$ (décomposition fractionnaire) $\Rightarrow$ S est surjective. L'évènement $\{S \leq x \in [0,1]\}$ est alors constitué de l'ensemble F_x des décompositions fractionnaires des réels inférieurs ou égaux à x.

On écrit $x = \sum_{n=1}^{+\infty} \frac{x_n}{b^n}$ (décomposition non unique : voir remarque ci-dessous). Si $x = 0$ alors F_x est réduit à la suite nulle, et $\mathrm{P}(S \leq x) = 0$. Supposons donc $x > 0$.

$$\begin{aligned}
\mathrm{P}(S \leq x) &= \mathrm{P}(S < x \in]0,1]) \\
&= \mathrm{P}\left(\bigsqcup_{n \in \mathbb{N}^* / x_n > 0} \{\omega \in \Omega' \,/\, \omega_n < x_n \text{ et } \forall n' < n \,,\, \omega_{n'} = x_{n'}\} \right) \\
&= \sum_{n \in \mathbb{N}^* / x_n > 0} \mathrm{P}(\{\omega \in \Omega' \,/\, \omega_n < x_n \text{ et } \forall n' < n \,,\, \omega_{n'} = x_{n'}\}) \\
&= \sum_{n \in \mathbb{N}^* / x_n > 0} \mathrm{P}(\omega_n < x_n) \times \prod_{n'=1}^{n-1} \mathrm{P}(\omega_{n'} = x_{n'}) \\
&= \sum_{n \in \mathbb{N}^* / x_n > 0} \frac{x_n}{b} \times \prod_{n'=1}^{n-1} \frac{1}{b} = \sum_{n \in \mathbb{N}^* / x_n > 0} \frac{x_n}{b^n} \\
&= \sum_{n=1}^{+\infty} \frac{x_n}{b^n} = x
\end{aligned}$$

(Union dénombrable disjointe, VA indépendantes, toutes de même loi). On obtient $\mathrm{P}(S \leq x \in [0,1]) = x$: fonction de répartition de la loi uniforme sur $[0,1]$.

Remarque : S n'est pas bijective, car les nombres s'écrivant comme une somme finie $\sum_{k=1}^{n} \frac{\omega_k}{b^k}$ (avec $\omega_n = b-1$) possèdent deux développements : celui-ci et $\sum_{k=1}^{n-1} \frac{\omega_k}{b^k} + \sum_{k=n+1}^{+\infty} \frac{b-1}{2^k}$. Cependant l'ensemble E des ω dont tous les termes valent $b-1$ apcr. (à partir d'un certain rang) est dénombrable, donc $S(E)$ est de mesure de Lebesgue nulle. Ainsi S réalise une bijection de $\Omega' = \Omega \backslash E$ dans $[0,1[$ avec $\mathrm{P}(\Omega') = 1$. (Attention 1 n'est plus atteint, car il admet une seule décomposition fractionnaire en base b, exclue de Ω').

(d) En déduire une façon de simuler une variable aléatoire de loi uniforme sur l'intervalle $[0,1]$ à partir de la pièce truquée. [2]

Solution :

On reprend la définition de Z de la réponse b). Soit $(Z_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes de même loi que $\frac{Z+1}{2}$. Les Z_n prennent les valeurs 0 et 1 avec probabilité $\frac{1}{2}$. On pose alors

$$S = \sum_{n=1}^{+\infty} \frac{Z_n}{2^n}$$

D'après la question précédente S suit une loi uniforme sur $[0,1]$, et en notant S_n les sommes partielles

$$\begin{aligned} 0 < S - S_n &= \sum_{k=n+1}^{+\infty} \frac{Z_k}{2^k} \\ &< \sum_{k=n+1}^{+\infty} \frac{1}{2^k} = \frac{1}{2^n} \end{aligned}$$

On peut donc simuler la réalisation de $S_n = \sum_{k=1}^{n} \frac{Z_k}{2^k}$ avec n assez grand (en fonction de la précision souhaitée). Par exemple pour une précision à 10^{-6} près il suffit de choisir $n = 20$ (car $2^{20} > 10^6$).

(e) Toujours avec la pièce truquée, comment simuler une variable aléatoire d-dimensionnelle suivant la loi uniforme dans la boule unité $\mathcal{B}_d(0,1) \subset \mathbb{R}^d$? [3]

1. Via une méthode de rejet.
2. D'une autre manière plus efficace ?

Donnez la complexité en moyenne (en fonction de la précision souhaitée), et, si possible, la loi du nombre de lancers effectués.

Solution :
L'algorithme de rejet consiste à simuler une VA dans $[-1,1]^d$ puis à rejeter les réalisations $(x_1, \dots, x_d)$ ne vérifiant pas $x_1^2 + \dots + x_d^2 = 1$. Il n'est pas efficace lorsque d dépasse 4 ou 5 car le ratio du volume de la sphère unité par celui de l'hypercube $[-1,1]^d$ décroît très rapidement. (À la limite, le volume de l'hypersphère est nul tandis que celui de l'hypercube explose). Plus précisément, le ratio des deux volumes vaut

$$\tau_d = \frac{\pi^{d/2}}{2^d \Gamma(d/2+1)}.$$

(voir la page Wikipedia). Le nombre de tirages T_d nécessaires avant d'en obtenir un dans la sphère suit la loi géométrique de paramètre τ_d. Il reste à déterminer le nombre de lancers de la pièce truquée requis pour obtenir une réalisation de $\mathcal{U}(-1,1)$ (par homothétie depuis $\mathcal{U}(0,1)$).

D'après la réponse b), le nombre de lancers de pièce nécessaires à l'obtention d'un digit binaire s'écrit $2N$ où N suit la loi géométrique de paramètre $\alpha_p = 2p(1-p)$. D'après la réponse précédente il faut exactement $\lceil \log_2 10^P \rceil$ chiffres binaires pour une précision P sur $\mathcal{U}(0,1)$, donc $\lceil \log_2 10^P \rceil + 1$ chiffres si l'on vise $\mathcal{U}(-1,1)$. On en déduit qu'il faut lancer la pièce $L_d = 2(\lceil \log_2 10^P \rceil + 1)NT_d$ fois pour obtenir un point aléatoire dans l'hypersphère.

La loi de cette VA ne se calcule pas facilement : on est ramené au produit de deux lois géométriques, obtenant au mieux

$$\mathrm{P}(L_d = k) = \frac{\alpha_p \tau_d}{(1-\alpha_p)(1-\tau_d)} \sum_{m|k} (1-\alpha_p)^{k/m} (1-\tau_d)^m \,,$$

avec $k = 2(\lceil \log_2 10^P \rceil + 1)\ell$, $m|k$ signifiant "m divise k". L'espérance en revanche s'obtient facilement, N et T_d étant indépendantes :

$$\mathrm{E}[L_d] = 2(\lceil \log_2 10^P \rceil + 1)\mathrm{E}[N]\mathrm{E}[T_d] = \frac{2(\lceil \log_2 10^p \rceil + 1)}{\alpha_p \tau_d}$$

Une autre façon de procéder utilise l'invariance de la loi normale centrée réduite par rotation. Plus précisément, si X suit la loi $\mathcal{N}(0, I_d)$ où I_d est la matrice identité à d dimensions et si U est une matrice de rotation, alors UX suit encore la loi $\mathcal{N}(0, I_d)$. On en déduit une manière d'échantillonner des points aléatoirement sur la sphère unité (*et non à l'intérieur*) :

$$Z_d = \frac{1}{\sqrt{\sum_{i=1}^d X_i^2}} (X_1, \ldots, X_d) \,,$$

avec $X_i \sim \mathcal{N}(0, 1)$.

Ensuite il faut tirer un point sur le rayon $(0, z_d)$ (de longueur 1) selon la loi uniforme dans la sphère, c'est-à-dire un réel y dans $[0, 1]$ selon

$$\mathrm{P}(Y \leq y) = y^d \,,$$

ratio des volumes des hypersphères de rayon y et 1. Cette fonction de répartition s'inverse facilement : il suffit de simuler une variable aléatoire U de loi uniforme sur $[0, 1]$ puis de calculer $U^{1/d}$ pour obtenir un échantillon de Y. En regroupant, on obtient un échantillon uniforme dans l'hypersphère par

$$\frac{U^{1/d}}{\sqrt{\sum_{i=1}^d X_i^2}} (X_1, \ldots, X_d) \,.$$

Les d échantillons de lois normales s'obtiennent par la méthode de Box-Müller à l'aide de $\lceil \frac{d}{2} \rceil$ simulations de VA exponentielles (obtenues depuis une réalisation d'une VA de loi uniforme via l'inversion de la fonction de répartition), et $\lceil \frac{d}{2} \rceil$

simulations de lois uniformes ; soit au total $2\left\lceil \frac{d}{2} \right\rceil$ échantillons uniformes dans $[0,1]$. On en déduit la VA donnant le nombre de lancers de la pièce truquée :

$$\left(4N\left\lceil \frac{d}{2} \right\rceil + \delta_{d=2k}\right)\left(\Upsilon + \lceil \log_2 10^P \rceil\right),$$

Υ étant une *variable aléatoire* indiquant la précision additionnelle moyenne nécessaire sur U pour obtenir une précision P en sortie.

Υ est aléatoire pour deux raisons :

- il faut s'assurer d'avoir $U > 0$ pour calculer $\ln U$;
- $\frac{1}{\sqrt{\sum_{i=1}^{d} X_i^2}}$ est d'autant plus imprécis que la norme de $(X_1, \ldots, X_d)$ est petite.

Cette dernière observation découle du fait que la fonction $t \in]0,+\infty[\mapsto \frac{1}{\sqrt{t}}$ n'est pas lipschitzienne au voisinage de 0. Elle l'est cependant sur tout $]\varepsilon,+\infty[$ avec $\varepsilon > 0$, donc on pourrait donner une borne supérieure probabiliste sur Υ (second exercice pour les vacances :)). On se contente ici d'évaluer la complexité en moyenne.

L'analyse théorique n'étant a priori pas évidente, on utilise une librairie de calcul en précision arbitraire (MPFR) afin de comparer les valeurs des composantes calculées avec une précision initiale P, et avec une précision initiale $2P$ ($2 =$ facteur arbitraire). On adopte ainsi un point de vue inverse : pour P chiffres significatifs sur les VA uniformes dans $[0,1]$ combien en obtient-on sur le résultat ?

Voici le programme en langage C. Arguments :

1. N : nombre de répétitions du calcul des d composantes ;
2. d : dimension ;
3. P : précision initiale.

Le programme affiche l'écart (absolu) moyen entre les calculs effectués en précision P et $2P$, ainsi que l'écart maximal. Sur plusieurs essais avec d variant entre 2 et 100, et P entre 10 et 100, on constate que l'écart moyen n'excède en général pas $0.5.10^{-P}$ tandis que l'écart maximal est au pire de l'ordre de $10^2.10^{-P}$. Cela suggère qu'en pratique il suffit de choisir $\Upsilon = 3$ pour assurer avec une excellente (mais non déterminée) probabilité la précision P sur le résultat.

Exercice 3 [6 points]

On s'intéresse dans cet exercice aux passages aux caisses d'un supermarché. Ce dernier comporte m caisses – toujours ouvertes – gérées par autant d'employé(e)s.

(a) *Chaque soir lors de la fermeture du magasin, les N derniers clients se présentent aux caisses. On suppose que N suit une loi de Poisson de paramètre λ.* [1]
Un client choisit la caisse k avec probabilité $p_k \in [0,1]$. Quelle est la loi du nombre total de clients passant à la caisse numéro k ?

Solution :
Notons N_k le nombre de clients passant à la caisse k :

$$\mathrm{P}(N_k = j \in \mathbb{N}) = \sum_{n=j}^{+\infty} \mathrm{P}(N_k = j | N = n)\mathrm{P}(N = n)\,,$$

donc il suffit de déterminer $\mathrm{P}(N_k = j | N = n \geq j) = \mathrm{P}_{N=n}(N_k = j)$. On remarque pour cela que conditionnellement à l'évènement $\{N = n\}$, la VA N_k s'écrit comme la somme de n épreuves de Bernouilli indépendantes de paramètre p_k. Elle suit donc la loi binomiale $\mathcal{B}(n, p_k)$, et on obtient successivement

$$\begin{aligned}
\mathrm{P}(N_k = j) &= \sum_{n=j}^{+\infty} C_n^j p_k^j (1-p_k)^{n-j} e^{-\lambda} \frac{\lambda^n}{n!} \\
&= e^{-\lambda} \frac{(\lambda p_k)^j}{j!} \sum_{n=j}^{+\infty} \frac{1}{(n-j)!} \lambda^{n-j} (1-p_k)^{n-j} \\
&= e^{-\lambda} \frac{(\lambda p_k)^j}{j!} \sum_{n=0}^{+\infty} \frac{(\lambda - \lambda p_k)^n}{n!} \\
&= e^{-\lambda + \lambda - \lambda p_k} \frac{(\lambda p_k)^j}{j!} = e^{-\lambda p_k} \frac{(\lambda p_k)^j}{j!}
\end{aligned}$$

On reconnaît la loi de Poisson de paramètre λp_k.

(b) *Pour cette question on suppose que N clients sont en train d'attendre aux caisses suivant les mêmes lois que précédemment. Chacun de ces client doit passer Λ_n articles suivant une loi uniforme dans $[\![1, A]\!]$.* [2]
Chaque caisse a un débit déterministe exprimé en "articles par minute" noté a_k.
Le temps C_n nécessaire au paiement d'un client suit une loi Gamma $\Gamma(\alpha, \beta)$.
Un retardataire arrive aux caisses avec Λ_0 articles. Déterminez l'espérance de son temps d'attente, et interprétez le résultat obtenu.

Solution :
Notons T la variable aléatoire du temps d'attente du retardataire, et T_k le temps d'attente sachant que l'on a sélectionné la caisse k.

$$\mathrm{E}[T] = \sum_{k=1}^{m} p_k \mathrm{E}[T_k]$$

$$T_k = C_n + \frac{\Lambda_0}{a_k} + \sum_{n=1}^{N_k} \left(C_n + \frac{\Lambda_n}{a_k} \right) ,$$

car il faut ajouter le temps de passage du retardataire à celui des clients déjà présents dans la file. On écrit alors

$$\begin{aligned} \mathrm{E}[T_k] &= \sum_{r=0}^{+\infty} \mathrm{E}[T_k|N_k = r]\mathrm{P}(N_k = r) \\ &= \sum_{r=0}^{+\infty} \mathrm{E}\left[C_n + \frac{\Lambda_0}{a_k} + \sum_{n=1}^{r} \left(C_n + \frac{\Lambda_n}{a_k} \right) \right] e^{-\lambda p_k} \frac{(\lambda p_k)^r}{r!} \\ &\quad \text{(d'après la question précédente)} \\ &= \sum_{r=0}^{+\infty} \left[\frac{\alpha}{\beta} + \frac{\Lambda_0}{a_k} + r \left(\frac{\alpha}{\beta} + \frac{A}{2a_k} \right) \right] e^{-\lambda p_k} \frac{(\lambda p_k)^r}{r!} \\ &\quad \text{(par linéarité de l'espérance)} \end{aligned}$$

Une fois les calculs effectués on obtient

$$\mathrm{E}[T] = \underbrace{\frac{\alpha}{\beta}}_{\text{paiement}} + \underbrace{\sum_{k=1}^{m} p_k \frac{\Lambda_0}{a_k}}_{\text{passage des articles}} + \sum_{k=1}^{m} p_k(\lambda p_k) \underbrace{\left(\frac{\alpha}{\beta} + \frac{A+1}{2a_k} \right)}_{\text{temps moyen par client}} ,$$

p_k désignant la proportion de clients à la caisse numéro k, et λp_k le nombre moyen de clients y passant.

Remarque : la loi du temps d'attente serait beaucoup plus compliquée à exprimer (pas sûr d'ailleurs qu'on puisse l'expliciter ...)

(c) *On suppose désormais que personne n'attend aux caisses.* [3]
Les hypothèses de la question précédente restent valides.
Xénia et Youri sont les deux derniers à arriver pour passer leurs achats : Xénia choisit une caisse suivant la distribution $p_1, \ldots, p_m$, puis Youri en sélectionne une au hasard parmi celles restant inoccupées. On note X et Y les variables aléatoires des temps d'attente respectifs. Exprimez la loi jointe, puis simplifiez la formule obtenue dans le cas $\alpha = 1$. X et Y sont-elles indépendantes ?

Solution :
On applique la formule des probabilités totales avec le système complet $K_Z = 1, \ldots, K_Z = m$ où K_Z désigne le numéro de la caisse choisie par Z.

$$\begin{aligned}
\mathrm{P}(X \leq x \cap Y \leq y) &= \sum_{k=1}^{m} \mathrm{P}(X \leq x \cap Y \leq y | K_X = k)\mathrm{P}(K_X = k) \\
&= \sum_{k=1}^{m}\sum_{\ell=1}^{m} \mathrm{P}(X \leq x \cap Y \leq y | K_X = k \cap K_Y = \ell)\mathrm{P}(K_X = k)\mathrm{P}(K_Y = \ell | K_X = k) \\
&= \sum_{k=1}^{m}\sum_{\ell=1}^{m} \mathrm{P}(X \leq x | K_X = k)\mathrm{P}(Y \leq y | K_Y = \ell)\mathrm{P}(K_X = k)\mathrm{P}(K_Y = \ell | K_X = k) \ ,
\end{aligned}$$

car les évènements $\{X \leq x\}$ et $\{Y \leq y\}$ sont indépendants une fois les caisses connues. On a $\mathrm{P}(K_X = k) = p_k$. Il faut déterminer les autres quantités.

$$\mathrm{P}(K_Y = \ell | K_X = k) = \frac{\delta_{\ell \neq k}}{m-1} \,.$$

$$\begin{aligned}
\mathrm{P}(X \leq x | K_X = k) &= \mathrm{P}\left(C + \frac{\Lambda}{a_k} \leq x\right) \\
&= \sum_{r=1}^{A} \mathrm{P}\left(C \leq x - \frac{r}{a_k}\right)\frac{1}{A} \\
&= \frac{1}{A}\sum_{r=1}^{A}\int_{t=0}^{x-\frac{r}{a_k}} \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1}e^{-\beta t}dt \\
&= \frac{\beta^\alpha}{A\Gamma(\alpha)}\sum_{r=1}^{A}\int_{u=0}^{\beta\left(x-\frac{r}{a_k}\right)} \left(\frac{u}{\beta}\right)^{\alpha-1} e^{-u}\frac{du}{\beta} \\
&\quad \text{changement de variable } \beta t = u \\
&= \frac{1}{A\Gamma(\alpha)}\sum_{r=1}^{A}\gamma\left(\alpha, \beta\left(x - \frac{r}{a_k}\right)\right) \ ,
\end{aligned}$$

en utilisant la fonction gamma incomplète.
Le calcul est le même pour $\mathrm{P}(Y \leq y | K_Y = \ell)$: on obtient

$$\mathrm{P}(Y \leq y | K_Y = \ell) = \frac{1}{A\Gamma(\alpha)}\sum_{r=1}^{A}\gamma\left(\alpha, \beta\left(y - \frac{r}{a_\ell}\right)\right) \ .$$

Finalement,

$$\begin{aligned}
\mathrm{P}(X \leq x \cap Y \leq y) = &\frac{1}{A^2\Gamma(\alpha)^2(m-1)}\sum_{k=1}^{m}\sum_{\ell \neq k} p_k \\
&\times \sum_{r_1, r_2 \in [\![1,A]\!]} \gamma\left(\alpha, \beta\left(x - \frac{r_1}{a_k}\right)\right)\gamma\left(\alpha, \beta\left(y - \frac{r_2}{a_\ell}\right)\right)
\end{aligned}$$

Il reste à dériver cette expression pour obtenir la densité. On utilise pour cela

$\frac{d}{dz}\gamma(\alpha, h(z)) = h(z)^{\alpha-1}e^{-h(z)}h'(z)$ avec $h : z \mapsto \beta z + w$.

$$f_{X,Y}(x,y) = \frac{\beta^\alpha}{A^2\Gamma(\alpha)^2(m-1)} \sum_{k=1}^{m} \sum_{\ell \neq k} p_k \times \sum_{r_1,r_2 \in [\![1,A]\!]} \left[\psi_{r_1,a_k}^{\alpha-1}(x) e^{-\beta\psi_{r_1,a_k}(x)} \gamma(\alpha, \beta\psi_{r_2,a_\ell}(y)) + \gamma(\alpha, \beta\psi_{r_1,a_k}(x)) \psi_{r_2,a_\ell}^{\alpha-1}(y) e^{-\beta\psi_{r_2,a_\ell}(y)}\right],$$

avec $\psi_{r,a}(z) = z - \frac{r}{a}$. Enfin, pour $\alpha = 1$ (cas de la loi exponentielle $\mathcal{E}(\beta)$), certaines choses se simplifient :

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{\beta^\alpha}{A^2(m-1)} \sum_{k=1}^{m} \sum_{\ell \neq k} p_k \\ &\quad \times \sum_{r_1,r_2 \in [\![1,A]\!]} e^{-\beta\psi_{r_1,a_k}(x)}(1 - e^{-\beta\psi_{r_2,a_\ell}(y)}) + (1 - e^{-\beta\psi_{r_1,a_k}(x)})e^{-\beta\psi_{r_2,a_\ell}(y)} \\ &= \frac{\beta^\alpha}{A^2(m-1)} \sum_{k=1}^{m} \sum_{\ell \neq k} p_k \sum_{r_1,r_2 \in [\![1,A]\!]} e^{-\beta\psi_{r_1,a_k}(x)} + e^{-\beta\psi_{r_2,a_\ell}(y)} - 2e^{-\beta(\psi_{r_1,a_k}(x) + \psi_{r_2,a_\ell}(y))}. \end{aligned}$$

Dans le cas général X et Y ne sont pas indépendants car les choix de caisses ne le sont pas (cf. calcul ci-dessous), et la loi du temps d'attente est entièrement déterminée par la caisse (via le paramètre a_k). Cependant si par exemple tous les a_k sont égaux, alors X et Y sont bien sûr indépendants.

$$\begin{aligned} \mathrm{P}(K_Y = \ell) &= \sum_{k=1}^{m} \mathrm{P}(K_Y = \ell | K_X = k)\mathrm{P}(K_X = k) \\ &= \sum_{k=1}^{m} \frac{\delta_{\ell \neq k}}{m-1} p_k \\ &= \frac{1}{m-1} \sum_{k \neq \ell} p_k = \frac{1 - p_\ell}{m-1}. \end{aligned}$$

$$\Rightarrow \mathrm{P}(K_X = k)\mathrm{P}(K_Y = k) = \frac{p_k(1-p_k)}{m-1} \neq \mathrm{P}(K_X = k \cap K_Y = k) = 0$$

Note : l'objectif était d'aller vers ce genre de problème, mais le contexte de l'exercice est maladroitement choisi et j'ai manqué de temps.