

## TD 02 – Variables Aléatoires (corrigé)

**Exercice 1.***Independence*

1. Show that the events  $\{A_i\}_{1 \leq i \leq n}$  are mutually independent if and only if

$$\mathbf{P} \left\{ \bigcap_{i=1}^n B_i \right\} = \prod_{i=1}^n \mathbf{P} \{B_i\}$$

where for every  $i$ , either  $B_i = A_i$  or  $B_i = A_i^c$ . We use the notation  $A^c$  for the complement of  $A$  in  $\Omega$ .

$\Rightarrow$  Induction par le nombre de  $A_i^c$  en utilisant  $P(A_i) + P(\cup A_i^c) = 1$ .

$\Leftarrow$  Prouver que

$$\mathbf{P} \left\{ \bigcap_{i \in I \subseteq \{1, \dots, n\}} B_i \right\} = \prod_{i \in I \subseteq \{1, \dots, n\}} \mathbf{P} \{B_i\}.$$

Induction par  $n - |I|$ .

For random variables  $X_1, \dots, X_n$ , recall that we defined independence by asking that

$$\mathbf{P} \{X_1 \leq t_1, \dots, X_n \leq t_n\} = \mathbf{P} \{X_1 \leq t_1\} \cdots \mathbf{P} \{X_n \leq t_n\} \quad (1)$$

for all  $t_1, \dots, t_n$  reals.

2. Show that in the case where  $X_i$  are all discrete random variables taking values in a discrete<sup>1</sup> countable set  $C$ , then this is equivalent to asking that for all  $x_1, \dots, x_n$  in  $C$ ,

$$\mathbf{P} \{X_1 = x_1, \dots, X_n = x_n\} = \prod_{i=1}^n \mathbf{P} \{X_i = x_i\}. \quad (2)$$

$\Rightarrow$  Commençons par montrer que (2)  $\Rightarrow$  (1). Pour simplifier les notations, on note  $C_i$  l'ensemble (dénombrable)  $C_i = C \cap (-\infty, t_i]$ . On a alors

$$\begin{aligned} \mathbf{P} \{X_1 \leq t_1, \dots, X_n \leq t_n\} &= \sum_{a_1 \in C_{t_1}, \dots, a_n \in C_{t_n}} \mathbf{P} \{X_1 = a_1, \dots, X_n = a_n\} && \text{(partition disjointe)} \\ &= \sum_{a_1 \in C_{t_1}, \dots, a_n \in C_{t_n}} \mathbf{P} \{X_1 = a_1\} \cdots \mathbf{P} \{X_n = a_n\} && \text{avec 2} \\ &= \left( \sum_{a_1 \in C_{t_1}} \mathbf{P} \{X_1 = a_1\} \right) \cdots \left( \sum_{a_n \in C_{t_n}} \mathbf{P} \{X_n = a_n\} \right) \\ &= \mathbf{P} \{X_1 \leq t_1\} \cdots \mathbf{P} \{X_n \leq t_n\}. \end{aligned}$$

Montrons maintenant que (1)  $\Rightarrow$  (2). On le fait par récurrence, en montrant que  $\mathbf{P} \{X_1 = x_1, \dots, X_i = x_i, X_{i+1} \leq t_{i+1}, \dots, X_n \leq t_n\} = \mathbf{P} \{X_1 = x_1\} \cdots \mathbf{P} \{X_i = x_i\} \cdot \mathbf{P} \{X_{i+1} \leq t_{i+1}\} \cdots \mathbf{P} \{X_n \leq t_n\}$ . Pour  $i = 0$ , c'est vrai par hypothèse (c'est (1)). Supposons l'hypothèse de récurrence vraie pour  $i < n$  et montrons là pour  $i + 1$ . Comme les variables sont discrètes, on sait qu'il exist  $x'_{i+1} < x_{i+1}$  tel que pour tout  $x \in ]x'_{i+1}, x_{i+1}[$ , on ait  $\mathbf{P} \{X_{i+1} = x\} = 0$ . On a alors que  $\{X_{i+1} = x_{i+1}\} = \{X_{i+1} \leq x_{i+1}\} \setminus \{X_{i+1} \leq x'_{i+1}\}$ . D'où

$$\begin{aligned} \mathbf{P} \{X_1 = x_1, \dots, X_{i+1} = x_{i+1}, X_{i+2} \leq t_{i+2}, \dots, X_n \leq t_n\} &= \mathbf{P} \{X_1 = x_1, \dots, X_{i+1} \leq x_{i+1}, X_{i+2} \leq t_{i+2}, \dots, X_n \leq t_n\} \\ &\quad - \mathbf{P} \{X_1 = x_1, \dots, X_{i+1} \leq x'_{i+1}, X_{i+2} \leq t_{i+2}, \dots, X_n \leq t_n\} \\ &= \mathbf{P} \{X_1 = x_1\} \cdots (\mathbf{P} \{X_{i+1} \leq x_{i+1}\} - \mathbf{P} \{X_{i+1} \leq x'_{i+1}\}) \cdot \mathbf{P} \{X_{i+1} = x_{i+1}\} \cdots \mathbf{P} \{X_n \leq t_n\} \\ &= \mathbf{P} \{X_1 = x_1\} \cdots \mathbf{P} \{X_{i+1} = x_{i+1}\} \cdot \mathbf{P} \{X_{i+1} = x_{i+1}\} \cdots \mathbf{P} \{X_n \leq t_n\}. \end{aligned}$$

On conclut que les deux formulations sont équivalentes.

Remarque. Si  $C$  n'est pas discret (c'est-à-dire qu'il y a des points d'accumulation), on peut s'en sortir en passant à la limite.

**Exercice 2.***Répétitions dans une suite de bits aléatoires*

Soient  $(X_1, \dots, X_n)$  des variables aléatoires i.i.d. de loi uniforme sur  $\{0, 1\}$ . Une répétition est un sous-mot de  $X_1 X_2 \cdots X_n$  du type  $00 \cdots 0$  ou  $11 \cdots 1$ . Ainsi la suite  $00011001$  contient 4 répétitions de longueur 2.

1. 'discrete' means that there is some  $\varepsilon > 0$  such that all points of  $C$  are at distance at least  $\varepsilon$ .

1. Pour  $p > 1$  fixé, quelle est l'espérance du nombre de répétitions de longueur  $p$ ? Montrez que pour  $p = \lfloor \log_2 n \rfloor$ , cette espérance est de l'ordre de 1.  $\text{☞}$  Une application directe du principe de linéarité de l'espérance (écrire  $N$  comme une somme d'indicatrices...) donne que l'espérance du nombre  $N$  de répétitions de longueur  $p$  vaut  $\mathbb{E}[N] = \frac{n-p+1}{2^{p-1}}$ . Lorsque  $p = \lfloor \log_2 n \rfloor$ , on a  $2 + o(1) \leq \mathbb{E}[N] \leq 4 + o(1)$ . De nombreuses copies ont affirmé que  $\mathbb{E}[N] \sim 2$ , ce qui est incorrect : si  $\{ \cdot \}$  désigne la partie fractionnaire, la quantité  $2^{\lfloor \log_2 n \rfloor}$  oscille entre 1 et 2...
2. Montrez que pour  $p \leq 0.99 \log_2 n$ , la probabilité d'obtenir au moins une répétition de longueur  $p$  tend vers 1 quand  $n$  tend vers l'infini.  $\text{☞}$  Soit  $k = \lfloor n/p \rfloor$ . On considère les événements  $(A_i)_{0 \leq i \leq k-1}$  définis par  $A_i = \{X_{pi+1} = \dots = X_{p(i+1)}\}$ . Ces événements sont indépendants (groupement par paquets) et de même probabilité  $1/2^{p-1}$ . On a

$$\mathbb{P}(\text{au moins une répétition de longueur } p) \geq \mathbb{P}\left(\bigcup A_i\right) = 1 - \left(1 - \frac{1}{2^{p-1}}\right)^k \geq 1 - \exp(k/2^{p-1}).$$

On vérifie que cette quantité tend vers 1 lorsque  $n$  tend vers l'infini dès lors que  $p \leq 0.99 \log_2 n$ .

### Exercice 3.

StagiairesL3

Bob veut recruter un stagiaire L3 parmi  $n$  candidats, et bien sûr il veut recruter le meilleur. Les candidats se présentent un par un pour l'interview, dans un ordre aléatoire. Quand il interviewe un candidat, Bob lui donne un score (pas d'ex-aequo). La règle du jeu de l'ENS de Lyon est la suivante : après avoir interviewé un candidat, soit on l'embauche, soit on perd toute chance de l'embaucher. Malin, Bob utilise la stratégie suivante : d'abord, interviewer  $m$  candidats, et les rejeter tous ; puis après le  $m$ -ème candidat, embaucher le premier candidat interviewé qui est meilleur (plus gros score) que tous ceux déjà interviewés.

1. Montrer que la probabilité que Bob choisisse le meilleur candidat est

$$P(n, m) = \frac{m}{n} \sum_{j=m+1}^n \frac{1}{j-1}$$

$\text{☞}$  Soit  $E_j$  l'évènement "le  $j$ -ième candidat est le meilleur et est recruté".  $P(E_j) = \frac{1}{n} \frac{m}{j-1}$  si  $j > m$ . En effet, le meilleur passe en  $j$ -ième position avec probabilité  $\frac{1}{n}$ , et dans ce cas, le meilleur parmi les  $j-1$  premiers, est parmi les  $m$  premiers candidats avec probabilité  $\frac{m}{j-1}$ .

2. En déduire que  $\lim_n \max_m P(n, m) \geq 1/e$ .

$\text{☞}$  Les bornes se trouvent par des calculs d'intégrales :

$$\frac{m}{n} (\ln(n) - \ln(m)) \leq P(n, m) \leq \frac{m}{n} (\ln(n-1) - \ln(m-1))$$

Le maximum de la fonction  $\frac{\ln(x)}{x}$  est atteint pour  $x = e$ . Donc on prend  $m = \frac{n}{e}$ , et on obtient  $\lim_n \max_m P(n, m) \geq 1/e$ .

### Exercice 4.

Inclusion

Let  $X$  and  $Y$  be independently and uniformly chosen subsets of  $\{1, \dots, n\}$ .

1. Compute  $\mathbf{P}\{X \subseteq Y\}$ .

$\text{☞}$  Choisir  $X$  uniformément au hasard dans  $\{1, \dots, n\}$ , revient à faire  $n$  lancers de dés de telle sorte que l'issue du  $i$ ème tirage décide si  $i \in X$  ou pas. Pour  $i \in \{1, \dots, n\}$ , soit  $X_i$  (resp.  $Y_i$ ) la variable aléatoire qui vaut 1 si  $i \in X$  (resp.  $i \in Y$ ) et 0 sinon. On dit que  $i$  fait défaut si  $Y_i = 0$  et  $X_i = 1$ . On voit que  $\mathbf{P}\{i \text{ fait défaut}\} = 1/4$ . Maintenant,  $X \subseteq Y$  si et seulement aucun  $i$  ne fait défaut, donc

$$\mathbf{P}\{X \subseteq Y\} = \prod_{i=1}^n \mathbf{P}\{i \text{ ne fait pas défaut}\} = \prod_{i=1}^n \left(1 - \frac{1}{4}\right) = \left(\frac{3}{4}\right)^n.$$

### Exercice 5.

Noir et blanc

Suppose we start with a bin containing two balls, one white and one black. I repeat the following procedure until the bin contains  $n$  balls. At each step, I take a ball uniformly at random from the bin and put the ball back into the bin and add another ball of the same color to the bin.

1. Show that the number of white balls is equally likely to be any number between 1 and  $n-1$ .

$\text{☞}$  Soit  $X_n$  la variable aléatoire qui compte le nombre de balles blanches au début de la  $(n-1)$ ème étape (c'est-à-dire lorsqu'il y a  $n$  boules dans la corbeille). Ainsi  $\mathbf{P}\{X_2 = 1\} = 1$  puisqu'au début de la première étape, il y a exactement une balle blanche dans la corbeille.

Soit  $Y_n$  la variable aléatoire qui vaut 1 si la balle tiré à la  $(n-1)$ ème étape est blanche, et 0 sinon.

Montrons par récurrence sur  $n$  que  $\forall i \in \{1, \dots, n-1\}, \mathbf{P}\{X_n = i\} = \frac{1}{n-1}$ .

Pour  $i \in \{2, \dots, n-2\}$ , on a  $X_n = i$  si et seulement l'un des deux cas suivants se présente :

- Il y avait précédemment  $i$  balles blanches, et l'on a tiré une boule noire ;
- Il y avait précédemment  $i - 1$  balles blanches, et l'on a tiré une boule blanche.

Comme ces deux événements sont disjoints, on a :

$$\begin{aligned} \mathbf{P}\{X_n = i\} &= \mathbf{P}\{X_{n-1} = i \text{ et } Y_{n-1} = 0\} + \mathbf{P}\{X_{n-1} = i - 1 \text{ et } Y_{n-1} = 1\} \\ &= \mathbf{P}\{Y_{n-1} = 0 \mid X_{n-1} = i\} \cdot \mathbf{P}\{X_{n-1} = i\} \\ &\quad + \mathbf{P}\{Y_{n-1} = 1 \mid X_{n-1} = i - 1\} \cdot \mathbf{P}\{X_{n-1} = i - 1\} \\ &= \frac{(n-1-i)}{(n-1)} \cdot \frac{1}{n-2} + \frac{i-1}{n-1} \cdot \frac{1}{n-2} \text{ par hypothèse de récurrence} \\ &= \frac{n-2}{n-1} \cdot \frac{1}{n-2} = \frac{1}{n-1} \end{aligned}$$

De plus, par le même genre d'arguments,

$$\mathbf{P}\{X_n = 1\} = \mathbf{P}\{X_{n-1} = 1 \text{ et } Y_{n-1} = 0\} = \frac{1}{n-2} \cdot \frac{n-2}{n-1} = \frac{1}{n-1}$$

et

$$\mathbf{P}\{X_n = n - 1\} = \mathbf{P}\{X_{n-1} = n - 2 \text{ et } Y_{n-1} = 1\} = \frac{1}{n-2} \cdot \frac{n-2}{n-1} = \frac{1}{n-1}$$

## Exercice 6.

*Aiguille de Buffon*<sup>2</sup>

*Aiguille de Buffon*<sup>2</sup>

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

Dans un premier temps, on suppose  $a < l$  et on va s'intéresser à la probabilité que l'aiguille tombe à cheval sur deux planches.

- Proposer une modélisation pour ce problème.

☞ Faire discuter les étudiants et se mettre d'accord sur l'introduction des deux quantités suivantes :

- la distance  $r$  du centre de l'aiguille à la rainure la plus proche,
- l'angle  $\theta \in [0, \pi/2]$  formé par l'aiguille et l'une des rainures (prendre l'angle géométrique : on se fiche du signe et des demi-tours que l'aiguille a effectuée durant sa chute).

Alors,  $r$  suit la loi uniforme  $\mathcal{U}([0, l/2])$  et  $\theta$  suit la loi uniforme  $\mathcal{U}([0, \pi/2])$ .

- Trouver une relation traduisant le fait que l'aiguille est tombée sur deux planches. En déduire la probabilité de cet événement.

☞

FAIRE UN DESSIN.

L'aiguille repose sur deux planches lorsque  $r < \frac{a}{2} \cos \theta$ . Ainsi, la probabilité cherchée est :

$$\begin{aligned} p &= \int_0^{\pi/2} \int_0^{l/2} 1_{r < \frac{a}{2} \cos \theta} dF_r dF_\theta \\ &= \frac{4}{\pi l} \int_0^{\pi/2} \int_0^{l/2} 1_{r < \frac{a}{2} \cos \theta} dr d\theta \\ &= \frac{4}{\pi l} \int_0^{\pi/2} \frac{a}{2} \cos \theta d\theta \\ &= \frac{4}{\pi l} \frac{a}{2} = \frac{2a}{\pi l}. \end{aligned}$$

- Dans le cas où  $a \geq l$ , calculer la probabilité que l'aiguille tombe sur au moins deux planches. Qu'obtient on pour  $a = l$ ? et pour  $a \gg l$ ?

☞ Cette fois, il est plus simple de calculer la probabilité que l'aiguille tombe sur exactement une planche. Déjà, il faut que  $a \cos \theta < l$  sinon la "largeur" de l'aiguille est plus grande que la largeur d'une planche. Ensuite, il faut que  $\frac{a}{2} \cos \theta < r$  pour que l'aiguille soit bien sur une seule planche.

FAIRE UN DESSIN.

Ainsi, la probabilité cherchée vérifie :

$$\begin{aligned} 1 - p &= \int_0^{\pi/2} \int_0^{l/2} 1_{\frac{a}{2} \cos \theta < r \text{ et } a \cos \theta < l} dF_r dF_\theta \\ &= \frac{4}{\pi l} \int_0^{\pi/2} \int_0^{l/2} 1_{\frac{a}{2} \cos \theta < r \text{ et } a \cos \theta < l} dr d\theta \\ &= \frac{4}{\pi l} \int_{\arccos \frac{l}{a}}^{\pi/2} \left( \frac{l}{2} - \frac{a}{2} \cos \theta \right) d\theta \\ &= \frac{2}{\pi l} \int_{\arccos \frac{l}{a}}^{\pi/2} (l - a \cos \theta) d\theta \\ &= \frac{2}{\pi l} \left( \left( \frac{\pi l}{2} - a \right) - \left( l \arccos \frac{l}{a} - a \sin \arccos \frac{l}{a} \right) \right) \\ &= 1 - \frac{2a}{\pi l} \left( 1 + \frac{l}{a} \arccos \frac{l}{a} - \sqrt{1 - \frac{l^2}{a^2}} \right). \end{aligned}$$

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

Ainsi, on obtient :

$$p = \frac{2a}{\pi l} \left( 1 + \frac{l}{a} \arccos \frac{l}{a} - \sqrt{1 - \frac{l^2}{a^2}} \right).$$

Si  $a = l$ , on retrouve  $p = \frac{2}{\pi}$  comme à la question précédente. Si  $a \gg l$ , alors on a  $u = \frac{l}{a} \ll 1$ , et un développement limité donne :

$$p \approx \frac{2}{\pi} \frac{1}{u} \left( 1 + u \left( \frac{\pi}{2} - u \right) - \left( 1 - \frac{1}{2} u^2 \right) + o(u^2) \right) = 1 - \frac{u}{\pi} = 1 - \frac{l}{\pi a}.$$

4. Maintenant qu'on a jeté les aiguilles, il ne reste plus qu'à jeter les ficelles. Une fois tombée à terre, la ficelle va former une courbe quelconque de longueur  $a$  dans le plan formé par le parquet, et on va s'intéresser cette fois au nombre de points d'intersection entre la ficelle et les rainures du parquet.

Calculer l'espérance du nombre de points d'intersection en fonction de la longueur de la ficelle.

Indice : on pourra approximer la ficelle par une succession de petits segments.



L'idée dans le cas de la ficelle est d'approcher la courbe obtenue suite au lancé par une courbe affine par morceaux, avec des morceaux de même longueur  $\varepsilon \rightarrow 0$ . Chaque morceau peut alors être vu comme une aiguille de taille  $\varepsilon$ .

On associe à chaque morceau la variable aléatoire  $X_i$  qui vaut 1 si le morceau touche deux planches, et 0 sinon (on néglige le cas où un morceau est pile sur une rainure car la probabilité d'un tel événement est nulle). La linéarité de l'espérance nous dit alors que :

$$E_a = \mathbf{E}[X_1 + \dots + X_N] = \mathbf{E}[X_1] + \dots + \mathbf{E}[X_N].$$

De là :

- soit on en déduit la linéarité de  $E_*$  ( $E_{a+b} = E_a + E_b$  puis  $E_a = K \cdot a$ ) et on se sert du cas de l'aiguille courte ( $a < l$ ) ou du cercle pour conclure que  $K = \frac{2}{\pi l}$ . Pour le cercle, si on regarde ce qui se passe pour un cercle de rayon  $\rho = \frac{l}{2}$ , alors la ficelle est de longueur  $a = \pi l$  et quelle que soit sa position dans l'espace, elle va avoir 2 intersections avec les rainures du parquet. D'où  $2 = K \cdot \pi l$  et  $K = \frac{2}{\pi l}$ .
- soit on constate que comme  $\varepsilon \rightarrow 0$ , les  $X_i$  correspondent au cas de l'aiguille courte de taille  $\varepsilon$  pour laquelle on a 0 intersection avec probabilité  $1 - \frac{2\varepsilon}{\pi l}$  et 1 intersection avec probabilité  $\frac{2\varepsilon}{\pi l}$ . Donc  $\mathbf{E}[X_i] = \frac{2\varepsilon}{\pi l}$  et en sommant on retrouve  $E_a = N \cdot \frac{2\varepsilon}{\pi l} = \frac{2a}{\pi l}$ .

Conclusion :  $E_a = \frac{2a}{\pi l}$ .

### Exercice 7.

*Le problème des rencontres*

On se donne une urne contenant  $n$  boules numérotées de 1 à  $n$ . On va alors procéder à une succession de tirages sans remise jusqu'à vider l'urne.

On s'intéresse aux événements  $E_i = \ll \text{la } i\text{ème boule tirée porte le numéro } i \gg$ .

1. Proposer un espace de probabilité pour modéliser cette expérience.

- On prend :
- $\Omega$  = l'ensemble des permutations de  $\{1, 2, \dots, n\}$ ,
  - $\mathcal{A} = \mathcal{P}(\Omega)$ ,
  - $\mathbf{P}\{\{\sigma\}\} = 1/n!$  (équipartition).

2. Calculer la probabilité des événements suivants :  $E_i$ ,  $E_i \cap E_j$  pour  $i < j$ , et enfin  $\bigcap_{j=1}^r E_{i_j}$  pour  $1 \leq i_1 < \dots < i_r \leq n$ .

$E_i$  arrive lorsqu'on tire la boule  $i$  à l'étape  $i$ . Or, 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).

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

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

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

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

3. Calculer la probabilité que l'évènement  $E_i$  se produise pour au moins un  $i$ . Quelle est la limite de cette probabilité lorsque  $n$  tend vers l'infini.

La quantité cherchée est donc  $P_n = \mathbf{P}\{\bigcup_{1 \leq i \leq n} E_i\}$ .

D'après la formule de Poincaré, on a :

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

On reconnaît presque le développement en série  $e^{-1} = \sum_{k=0}^n \frac{(-1)^k}{k!}$ . En fait, on a  $P_n - 1 \rightarrow -e^{-1}$  donc la limite vaut en fait  $1 - e^{-1}$ .

4. Combien y a-t-il de façons de placer huit tours sur un échiquier de telle sorte qu'aucune d'entre elles en attaque une autre? Qu'en est-il si on impose en plus que la diagonale principale soit vide?

☞ Pour placer 8 tours avec aucune en prise, il faut mettre 1 tour par ligne et par colonne. Il suffit donc par exemple de choisir successivement la ligne pour la tour sur la colonne  $i$  pour  $1 \leq i \leq 8$ . C'est un tirage sans remise dans  $\{1, 2, \dots, 8\}$ , donc il y a  $8! = 40320$  possibilités.

Pour le deuxième cas, on est ramené au problème des rencontres pour  $n = 8$  (on impose en effet que pour tout  $i$ , le  $i$ ème tirage ne donne pas la ligne  $i$ ). Donc on a  $8!/P_8 = 14833$  possibilités.