# TD 02 – Variables Aléatoires (corrigé)

**Exercice 1.** *Independence*

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

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

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

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

Induction par $n-|I|$.

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

$$\mathbf{P}\left\{X_1\le t_1,\dots,X_n\le t_n\right\} = \mathbf{P}\left\{X_1\le t_1\right\}\cdots\mathbf{P}\left\{X_n\le t_n\right\} \tag{1}$$

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[1] countable set $C$, then this is equivalent to asking that for all $x_1,\dots,x_n$ in $C$,

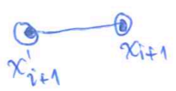
$$\mathbf{P}\left\{X_1 = x_1,\dots,X_n = x_n\right\} = \prod_{i=1}^n \mathbf{P}\left\{X_i = x_i\right\}. \tag{2}$$

☞ Commençons par montrer que (2) $\Rightarrow$ (1). Pour simplifier les notations, on note $C_t$ l'ensemble (dénombrable) countable $C_t = C\cap(-\infty,t]$. On a alors

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

(and here, $a_1,\dots,a_n$ are "discrete" by definition)

Montrons maintenant que (1) $\Rightarrow$ (2). On le fait par récurrence, en montrant que $\mathbf{P}\left\{X_1=x_1,\dots,X_i=x_i,X_{i+1}\le t_{i+1},\dots X_n\le t_n\right\} = \mathbf{P}\left\{X_1=x_1\right\}\cdots\mathbf{P}\left\{X_i=x_i\right\}\cdot\mathbf{P}\left\{X_{i+1}\le t_{i+1}\right\}\cdots\mathbf{P}\left\{X_n\le t_n\right\}$. 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 $\mathbf{P}\left\{X_{i+1}=x\right\}=0$. On a alors que $\{X_{i+1}=x_{i+1}\} = \{X_{i+1}\le x_{i+1}\}\setminus\{X_{i+1}\le x'_{i+1}\}$. D'où




$$\begin{aligned}
\mathbf{P}\left\{X_1=x_1,\dots,\underline{X_{i+1}=x_{i+1}},X_{i+2}\le t_{i+2},\dots X_n\le t_n\right\} &= \left.\begin{aligned}&\mathbf{P}\left\{X_1=x_1,\dots,X_{i+1}\le x_{i+1},X_{i+2}\le t_{i+2},\dots X_n\le t_n\right\}\\ &-\mathbf{P}\left\{X_1=x_1,\dots,X_{i+1}\le x'_{i+1},X_{i+2}\le t_{i+2},\dots X_n\le t_n\right\}\end{aligned}\right\}\text{ holds by recurence}\\
&= \mathbf{P}\left\{X_1=x_1\right\}\cdots\left(\mathbf{P}\left\{X_{i+1}\le x_{i+1}\right\}-\mathbf{P}\left\{X_{i+1}\le x'_{i+1}\right\}\right)\cdot\mathbf{P}\left\{X_{i+1}=x_{i+2}\right\}\cdots\mathbf{P}\left\{X_n\le t_n\right\}\\
&= \mathbf{P}\left\{X_1=x_1\right\}\cdots\mathbf{P}\left\{X_{i+1}=x_{i+1}\right\}\cdot\mathbf{P}\left\{X_{i+1}=x_{i+2}\right\}\cdots\mathbf{P}\left\{X_n\le t_n\right\}.
\end{aligned}$$

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 $X_1X_2\cdots X_n$ du type $00\cdots0$ ou $11\cdots1$. Ainsi la suite 00011001 contient 4 répétitions de longueur 2. (00, 00, 11, 00)

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

litte-o : strict upper bound
big-o : inclusive upper bound

$f(x)$ is $o(g(x))$ means that $g(x)$ grows much faster than $f(x)$

$\mathbb{E}[N] = \frac{2(n-p+1)}{2^p}$ either 00...0 or 11...1 / # possible subwords of length $p$

$A_i$ : the event that a subword starting at the $i^{th}$ element is a repetition of length $p$

$1\ 2\ 3 \cdots p\ p+1\ p+2 \cdots n-p+1\ n-p+2 \cdots n$ — $A_1, A_2, A_3, \ldots, A_{n-p}, A_{n-p+1}$

N : # repetitions of length p

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. ☞ 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 quantitté $2^{\{\log n\}}$ oscille entre 1 et 2...

$E[X_i] = P\{A_i\}$

$E[X] = \sum_{i=1}^{m} Pr\{A_i\}$

where $X_i$ : the indicator function of r.v. for event $A_i$

$X$ : r.v. counting #events occuring

2. Montrez que pour $p \leq 0.99 \log_2 n$, la probabilité d'obtenir au moins (at least) une répétition de longueur $p$ tend vers 1 quand $n$ tend vers l'infini. ☞ Soit $k = \lfloor n/p \rfloor$. On considère les événements $(A_i)_{0 \leq i \leq k-1}$ définis par $A_i = \{X_{pi+1} = \cdots = X_{pi+p}\}$. Ces événements sont indépendants (groupement par paquets) et de même probabilité $1/2^{p-1}$. On a

↳ the event that $X_{pi+1} X_{pi+2} \cdots X_{pi+p}$ is a repetition

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

$A_0$ or $A_1$ or ... or $A_{k-1}$ is a repetition

$A \cup B = (A^c \cap B^c)^c$

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.** (same as HW) *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. (hire the best candidate we see that is better than all candidates that have already seen before)

ex-aquo = seri

embaucher : hire

malin : smartly

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

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

→ because the candidate hired is among candidates numbered $m+1$ to $n$

☞ 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 (indeed), le meilleur passe en j-ième position avec probabilité $\frac{1}{n}$, et dans ce cas, le meilleur parmi les $j-1$ premiers (# candidates have seen), est parmi les $m$ premiers candidats avec probabilité $\frac{m}{j-1}$.

bcs we choose the first one who is better than the m first candidates already seen.

2. En déduire que $\lim_n \max_m P(n,m) \geq 1/e$.
☞ 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 (achieved) 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 $\{1, \ldots, n\}$.

1. Compute $\mathbf{P}\{X \subseteq Y\}$.
☞ Choisir $X$ uniformément au hasard dans $\{1, \ldots, n\}$, revient à faire $n$ lancers de dés (dice) de telle sorte que l'issue du $i$ième tirage ($i^{th}$ draw) décide si $i \in X$ ou pas. Pour $i \in \{1, \ldots, n\}$, 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 $\mathbf{P}\{i \text{ fait defaut}\} = 1/4$. Maintenant, $X \subseteq Y$ si et seulement aucun $i$ ne fait défaut, donc

wrong as this means that $i \in X$ but $i \notin Y$ (while we want to compute $P\{X \subseteq Y\}$)

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

because for $Y_i = 0, X_i = 0$; $Y_i = 1, X_i = 0$; $Y_i = 1, X_i = 1$ we still have $X \subseteq Y$

**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 $n-1$.
☞ Soit $X_n$ la variable aléatoire qui compte le nombre de balles blanches au début de la $(n-1)$ième 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 $(n-1)$ième étape est blanche, et 0 sinon.
Montrons par récurrence sur $n$ que $\forall i \in \{1, \ldots, n-1\}, \mathbf{P}\{X_n = i\} = \frac{1}{n-1}$. * (this is what we have to show)
Pour $i \in \{2, \ldots n-2\}$, on a $X_n = i$ si et seulement l'un des deux cas suivants se présente :

# whit balls at step $(n-1)$ is equal to $i$

— 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 évenements sont disjoints, on a :

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

$\leftarrow P(A \cap B) = P(A|B) \cdot P(B)$

De plus, par le même genre d'arguments,

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

et

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

substituting to the formula above

**Exercice 6.** *Aiguille de Buffon*

*Aiguille de Buffon*[2]

Jarum

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.

1. Proposer une modélisation pour ce problème.

☞ Faire discuter les étudiants et se mettre d'accord sur l'introduction des deux quantités suivantes :
1. la distance $r$ du centre de l'aiguille à la rainure la plus proche,
2. 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 $\mathcal{U}([0, \pi/2])$.

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

largeur = $l$

☞

FAIRE UN DESSIN.

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

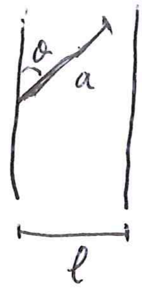
$$\begin{aligned}
p &= \int_0^{\pi/2} \int_0^{l/2} \mathbb{1}_{r < \frac{a}{2}\cos\theta} \, dF_r \, dF_\theta \\
&= \frac{4}{\pi l} \int_0^{\pi/2} \int_0^{l/2} \mathbb{1}_{r < \frac{a}{2}\cos\theta} \, dr \, d\theta \\
&= \frac{4}{\pi l} \int_0^{\pi/2} \frac{a}{2} \cos\theta \, d\theta \\
&= \frac{4}{\pi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

3. 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.

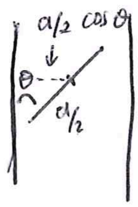
FAIRE UN DESSIN.

Ainsi, la probabilité cherchée vérifie :




$a \cos\theta < l$

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




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 $a >> l$, alors on a $u = \frac{l}{a} << 1$, et un développement limité donne :

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

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 $\varepsilon \to 0$. 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 :

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

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 $\varepsilon \to 0$, 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 : $E_a = \dfrac{2a}{\pi l}$.

**Exercice 7.** *Le problème des rencontres*

box kelereng

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 évènements $E_i$ = « la $i$ème boule tirée porte le numéro $i$».

tanpa pengembalian

1. Proposer un espace de probabilité pour modéliser cette expérience.
☞ On prend :
— $\Omega$ = l'ensemble des permutations de $\{1,2,\ldots,n\}$, (the event space)
— $\mathcal{A} = P(\Omega)$, (allowable events)
— $\mathbf{P}\{\{\sigma\}\} = 1/n!$ (équipartition). (probability function)

2. Calculer la probabilité des évènements suivants : $E_i$, $E_i \cap E_j$ pour $i < j$, et enfin $\bigcap_{j=1}^{r} E_{i_j}$ pour $1 \leq i_1 < \cdots < i_r \leq n$.
☞ $E_i$ arrive (happens) lorsqu'on tire la boule $i$ à l'étape $i$. Or (however), compter (counting) le nombre de permutations de $n$ éléments où $\sigma(i) = i$ revient (is equivalent) 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 $\mathbf{P}\{E_i\} = \dfrac{(n-1)!}{n!} = \dfrac{1}{n}$.
brought back
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 $\mathbf{P}\left\{\bigcap_{i_1<\cdots<i_r} E_{i_j}\right\} = \dfrac{1}{n(n-1)\ldots(n-r+1)}$. $= \dfrac{(n-r)!}{n!}$

3. Calculer la probabilité que l'évènement $E_i$ se produise (occurs) pour au moins (at least) un $i$ (for one $i$). Quelle est la limite de cette probabilité lorsque $n$ tend vers l'infini.
☞ La quantité cherchée est donc $P_n = \mathbf{P}\{\bigcup_{1\leq i\leq n} E_i\}$.
$\hookrightarrow$ $E_1$ or $E_2$ or ... or $E_n$ occurs

D'après la formule de Poincaré, on a :

$\sum_{1\leq i_1<\cdots<i_k\leq n}$ means the sum for all subsets of $\{1,\ldots,n\}$ of size $k$

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

$\hookrightarrow$ by question 2, this is equal to $\binom{n}{k}\cdot\dfrac{(n-k)!}{n!} = \dfrac{1}{k!}$

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

Poincaré : $P\left\{\bigcup_{1\leq i\leq n} A_i\right\} = \sum_{k=1}^{n}(-1)^{k-1}\sum_{1\leq i_1\leq\cdots\leq i_k\leq n} P\left\{\bigcap_{1\leq j\leq k} A_{i_j}\right\}$

Bonus

8 towers chessboard in such a way

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 ?

what if we also require

☞ 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 $\{1, 2, \ldots, 8\}$, 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.

# TD 02, Random Variable

## Exc 1 [Independence]

1). Implement the results from the TD 1

2). Two different definitions of independence

(1) $P\{X_1 \le t_1, \dots, X_n \le t_n\} = \prod_{i=1}^{n} P\{X_i \le t_i\}$ for $t_1, \dots, t_n \in \mathbb{R}$

(2) $P\{X_1 = x_1, \dots, X_n = x_n\} = \prod_{i=1}^{n} P\{X_i = x_i\}$ for $x_1, \dots, x_n \in C$

Where $C$ is a discrete countable set ("discrete" means that $\exists\, \varepsilon > 0$ s.t. all points of $C$ are at distance at least $\varepsilon$)

$\Rightarrow$ $\boxed{(2) \Rightarrow (1)}$ Denote $C_t = C \cap (-\infty, t]$ which is countable set

$$P\{X_1 \le t_1, \dots, X_n \le t_n\} = \sum_{a_1 \in C_{t_1}, \dots, a_n \in C_{t_n}} P\{X_1 = a_1, \dots, X_n = a_n\}$$

bcs it's disjoint partitions

$$= \sum_{a_1 \in C_{t_1}, \dots, a_n \in C_{t_n}} \prod_{i=1}^{n} P\{X_i = a_i\}$$

by (2), and the fact that $C_{t_1}, \dots, C_{t_n}$ are discrete

$$= \prod_{i=1}^{n} \left( \sum_{a_i \in C_{t_i}} P\{X_i = a_i\} \right)$$

grouping each index $i$

$$= \prod_{i=1}^{n} P\{X_i \le t_i\}$$

$\Leftarrow$ $\boxed{(1) \Rightarrow (2)}$ We proceed by recurence, showing that:

$$P\{X_1 = x_1, \dots, X_i = x_i, X_{i+1} \le t_{i+1}, \dots, X_n \le t_n\} = \prod_{j=1}^{i} P\{X_j = x_j\} \prod_{j=i+1}^{n} P\{X_j \le t_j\}$$

* For $i = 0$: $P\{X_1 \le t_1, \dots, X_n \le t_n\} = \prod_{j=1}^{n} P\{X_j \le t_j\}$:

true by definition (1)

* Suppose that the equation holds for $i < n$. We'll show for $i+1$
* Since the variables $x_i$'s are discrete, $\exists\, x'_{i+1} < x_{i+1}$ such that

$\forall x \in ]x'_{i+1}, x_{i+1}[$, we have $P\{X_{i+1} = x\} = 0$


Hence: $\{X_{i+1} = x_{i+1}\} = \{X_{i+1} \le x_{i+1}\} \setminus \{X_{i+1} \le x'_{i+1}\}$

* So: $P\{X_1 = x_1, \dots, X_{i+1} = x_{i+1}, X_{i+2} \le t_{i+2}, \dots, X_n \le t_n\}$

$= P\{X_1 = x_1, \dots, X_{i+1} \le x_{i+1}, X_{i+2} \le t_{i+2}, \dots, X_n \le t_n\}$

$- P\{X_1 = x_1, \dots, X_{i+1} \le x'_{i+1}, X_{i+2} \le t_{i+2}, \dots, X_n \le t_n\}$

$$= \prod_{j=1}^{i} P\{X_j = x_j\} \cdot P\{X_{i+1} \le x_{i+1}\} \cdot \prod_{j=i+2}^{n} P\{X_j \le t_j\}$$

$$- \prod_{j=1}^{i} P\{X_j = x_j\} \cdot P\{X_{i+1} \le x'_{i+1}\} \cdot \prod_{j=i+2}^{n} P\{X_j \le t_j\}$$

$$= \prod_{j=1}^{i} P\{X_j = x_j\} \cdot \left(P\{X_{i+1} \le x_{i+1}\} - P\{X_{i+1} \le x'_{i+1}\}\right) \cdot \prod_{j=i+2}^{n} P\{X_j \le t_j\}$$

$$= \prod_{j=1}^{i+1} P\{X_j = x_j\} \cdot \prod_{j=i+2}^{n} P\{X_j \le t_j\}$$

So the two formulations are equivalent

## Exc 2 [Répétitions dans une suite de bits aléatoires]

$(X_1, \dots, X_n)$ be r.v uniformly distributed in $\{0,1\}^n$.

Repetition : a word $X_1 X_2 \dots X_n$ of type $00\cdots0$ or $11\cdots1$

1). For $p > 1$ fixed, we want to compute :

$\mathbb{E}[N]$ ; $N$ : # repetitions of length $p$

* Word : $X_1 X_2 \cdots X_n$

Define for $1 \leq i \leq n-p+1$, $A_i = \{\text{subword } \underbrace{X_i X_{i+1} \cdots X_{i+p-1}}_{\text{length } p} \text{ is a repetition}\}$

$X_i$ : the indicator function for event $A_i$, i.e. $X_i = \begin{cases} 1 & \text{if it is a repetition} \\ 0 & \text{otherwise} \end{cases}$

By linearity of expectation : $\mathbb{E}[N] = \sum_{i=1}^{n-p+1} \mathbb{E}[X_i] = \sum_{i=1}^{n-p+1} P\{A_i\}$

* $P\{A_i\} = \frac{2}{2^p}$ for every $i$ (bcs there are two possible repetitions, namely $\underbrace{00\cdots0}_{p}$ and $\underbrace{11\cdots1}_{p}$ among $2^p$ possible subwords of length $p$)

So $\mathbb{E}[N] = (n-p+1) \cdot \frac{2}{2^p} = \frac{n-p+1}{2^{p-1}}$

* For $p = \lfloor \log_2 n \rfloor$, we have $2 + o(1) \leq \mathbb{E}[N] \leq 4 + o(1)$

$f(n)$ is $o(1)$ means that $\lim_{n\to\infty} f(n) = 0$
i.e $\forall \varepsilon > 0, \exists N_\varepsilon$ st $\forall n > N_\varepsilon$, we have $f(n) \leq \varepsilon$

## Exc 3 [Inclusion]

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

We'll compute $P\{X \subseteq Y\}$.

* Choose $X$ uniformly at random in $\{1, \dots, n\}$ by drawing dice $n$ times s.t. the result of the $i$th draw tells us whether $i \in X$ or $i \notin X$.
(We do similarly for $Y$)

* For $i \in \{1, \dots, n\}$, let $X_i$ and $Y_i$ be r.v where :

$X_i = \begin{cases} 1 & \text{if } i \in X \\ 0 & \text{oth.} \end{cases}$ $\qquad Y_i = \begin{cases} 1 & \text{if } i \in Y \\ 0 & \text{oth} \end{cases}$

* Note that $X \not\subseteq Y$ if $\exists i$ s.t. $Y_i = 0$, $X_i = 1$ (happens w.p $\frac{1}{4}$)

So $P\{X \subseteq Y\} = \prod_{i=1}^{n} P\{\underbrace{Y_i \neq 0 \text{ or } X_i \neq 1}\} = \prod_{i=1}^{n} \left(1 - \frac{1}{4}\right) = \left(\frac{3}{4}\right)^n$

possibilities : $X_i = 0, Y_i = 0$
$X_i = 0, Y_i = 1$
$X_i = 1, Y_i = 1$

# Exc 2 (continue)

2) We show that for $p \leq 0{,}99 \log_2 n$,

$$P\{\exists \text{ a repetition of length } p\} \xrightarrow[n \to \infty]{} 1$$

the event that $X_{pi+1} \cdots X_{pi+p}$ is a repetition

* Let $k = \lfloor \frac{n}{p} \rfloor$
* Consider the events $(A_i)_{0 \leq i \leq k-1}$ defined as : $A_i = \{X_{pi+1} = \cdots = X_{pi+p}\}$

The events $A_i$'s are independent, which happens w.p $\frac{2}{2^p} = \frac{1}{2^{p-1}}$

$$\underbrace{X_1 \cdots X_p}_{\text{for } p=0} \ \underbrace{X_{p+1} \cdots X_{2p}}_{p=1} \ \underbrace{X_{2p+1} \cdots X_{3p}}_{p=2} \cdots \ \cdots$$ (the grouping)

* So $P\{\exists \text{ a repetition of length } p\} \geq P\{\text{one of the } A_i\text{'s is a repetition}\}$

$$= P\left\{\bigcup_{i=0}^{k-1} A_i\right\}$$

$$= 1 - P\left\{\bigcap_{i=0}^{k-1} A_i^c\right\}$$

$$= 1 - \left(1 - \frac{1}{2^{p-1}}\right)^k$$

$$\geq 1 - \exp\left(\frac{k}{2^{p-1}}\right)$$

this value tends to 1 when n tends to $\infty$

# Exc 4 [Noir et Blanc]



* Start with a bin containing 2 balls, $w$ and $b$.
* Repeat :- take a ball uniformly at random from the bin and
  - put the ball back in to the bin
  - add another ball of same color

  until the bin contains $n$ balls.

* We show : $P\{\#\text{ white balls}\} = \frac{1}{n-1}$

* Let $X_n$ : rv for # white balls in the beginning of the $(n-1)^{th}$-step. (i.e. when $\exists$ $n$ balls in the bin).

  Let $Y_n$ : r.v st. $Y_n = \begin{cases} 1 & \text{if the ball taken at } (n-1)^{th} \text{ step is white} \\ 0 & \text{otherwise.} \end{cases}$

* We show by recurrence that :

$$\forall i \in \{1, \dots, n-1\}, \quad P\{X_n = i\} = \frac{1}{n-1}$$

* For $i \in \{2, \dots, n-2\}$, we have $X_n = i$ iff one of the following happens : (*) $\exists$ $i$ white balls, and the ball taken is black
  (*) $\exists$ $(i-1)$ white balls, and the ball taken is white

* The two events are disjoint so :

$$\begin{aligned} P\{X_n = i\} &= P\{X_{n-1} = i \;\&\; Y_{n-1} = 0\} + P\{X_{n-1} = i-1 \;\&\; Y_{n-1} = 1\} \\ &= P\{Y_{n-1} = 0 \mid X_{n-1} = i\} \cdot P\{X_{n-1} = i\} \\ &\quad + P\{Y_{n-1} = 1 \mid X_{n-1} = i-1\} \cdot P\{X_{n-1} = i-1\} \\ &= \left[\frac{n-1-i}{n-1} \cdot \frac{1}{n-2}\right] + \left[\frac{i-1}{n-1} \cdot \frac{1}{n-2}\right] \\ &= \frac{n-2}{n-1} \cdot \frac{1}{n-2} = \frac{1}{n-1} \end{aligned}$$

bcs they are ~~equally~~ uniformly distributed

because $\exists$ $n-1$ balls and among them, $i-1$ is $w$.

* Moreover :

$$P\{X_n = 1\} = P\{X_{n-1} = 1 \;\&\; Y_{n-1} = 0\} = \frac{1}{n-2} \cdot \frac{n-2}{n-1} = \frac{1}{n-1}$$

$$P\{X_n = n-1\} = P\{X_{n-1} = n-2 \;\&\; Y_{n-1} = 1\} = \frac{1}{n-2} \cdot \frac{n-2}{n-1} = \frac{1}{n-1}$$

$\square$

# Exc 5 [le problème des rencontres]

* Given a box containing $n$ balls numbered from 1 to $n$.
  We take one ball consecutively (without remise) until the box is empty
* Let $E_i$ : {the $i^{th}$ ball taken is of number $i$}

1) Probability space :

$\Omega$ : the set of permutations of $\{1,\dots,n\}$ (the space of events)

$\mathcal{A}$ : the set of allowable events

$P\{\{\sigma\}\} = \frac{1}{n!}$ for each event $\sigma \in \Omega$

2). * $P\{E_i\} = \frac{(n-1)!}{n!} = \frac{1}{n}$ because counting the # permutations where $\sigma(i) = i$ is the same as counting # permutations of $(n-1)$ elements

* $P\{E_i \cap E_j\} = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$

* $P\left\{\bigcap_{i_1 < \dots < i_r} E_{i_j}\right\} = \frac{1}{n(n-1)\dots(n-r+1)}$ $\left(\text{or } \frac{(n-r)!}{n!}\right)$

3) We calculate $P\{E_i$ occurs for at least one $i\}$ at least one of $E_i$'s occurs
i.e. we want to compute : $P_n = P\left\{\bigcup_{1 \le i \le n} E_i\right\}$

* By Poincaré formula :

$$P_n = \sum_{k=1}^{n} (-1)^{k-1} \underbrace{\sum_{1 \le i_1 < \dots < i_k \le n} P\left\{\bigcap_{1 \le j \le k} E_{i_j}\right\}}_{*}$$

* Note that, here $\sum_{1 \le i_1 < \dots < i_k \le n}$ means the sum for all subsets of $\{1,\dots,n\}$ of size $k$

So by question 2, the terms * is equal to :

$$\binom{n}{k} \cdot \frac{(n-k)!}{n!} = \frac{n!}{k!\,(n-k)!} \cdot \frac{(n-k)!}{n!} = \frac{1}{k!}$$

So $P_n = \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k!}$

* We know that $e^{-1} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}$, so $P_n - 1 \xrightarrow[n \to \infty]{} -e^{-1}$, i.e. $\boxed{P_n \xrightarrow[n \to \infty]{} 1 - e^{-1}}$

# Exc 5 (continue)

= Benteng

4) a) # ways to place 8 rooks on a chessboard s.t. no two of them attack each other

* 2 rooks cannot be on the same rows/columns
(i.e. for every Row and every column, there exists exactly one rook)

* So it's enough to choose successively the Row for the rook in column $i$ for each $1 \le i \le 8$.

* So there are $8!$ ways.

b) What if we also require that the main diagonal is empty

* Define $E_i = \{$the rook in column $i$ is assigned to the $i^{th}$ row$\}$

* So $P\{$the main diagonal is empty$\} = P\{\overline{E_i}$ happens for all $i$'s$\}$ ...
$$1 - P\{\bigcup_{1 \le i \le 8} E_i\} = 1 \sum_{i=1}^{8} \frac{(-1)^{i-1}}{i!} = (1 - P_8)$$

* So # ways $= 8! \cdot (1 - P_8)$

# Exc 6 [Balls in the bins]



* Tony has n baskets, he will make n throws, each time he will choose which basket uniformly and independently at random and he never fails.
* $X_i$ : # balls thrown in the $i^{th}$ basket

1). $P\{X_1 = 1 \mid X_1 + X_2 + X_3 = 1\}$

$$P\{X_1 = 1 \mid X_1 + X_2 + X_3 = 1\} = \frac{P\{X_1 = 1 \ \& \ X_1 + X_2 + X_3 = 1\}}{P\{X_1 + X_2 + X_3 = 1\}} = \frac{P\{X_1 = 1, X_2 = 0, X_3 = 0\}}{P\{X_1 + X_2 + X_3 = 1\}}$$

Note that, for $i \in \{1,2,3\}$, $p_i = P\{X_i = 1, X_{i+1 \bmod 3} = 0, X_{i+2 \bmod 3} = 0\}$

We have $p_1 = p_2 = p_3$ (by symmetry)

and $p_1 + p_2 + p_3 = P\{X_1 + X_2 + X_3 = 1\}$

So: $P\{X_1 = 1 \mid X_1 + X_2 + X_3 = 1\} = \dfrac{p_1}{p_1 + p_2 + p_3} = \dfrac{1}{3}$

2). $\mathbb{E}[X_1 \mid X_2 = 0]$

↳ It's equivalent to throwing n balls into $(n-1)$ bins (because we just forget the bin number 2)

Formula of expectation $\mathbb{E}[X] = \sum_{\ell \in \Omega} \Pr\{\ell\} \cdot X(\ell)$

So $\mathbb{E}[X_1 \mid X_2 = 0] = \dfrac{n}{n-1} = 1 - \dfrac{1}{n-1}$

Recall that expectancy = average

3). An expression for $P\{X_1 \geq X_2\}$

* We let $a = P\{X_1 > X_2\} = P\{X_2 > X_1\}$, and $b = P\{X_1 = X_2\}$
* So : $2a + b = 1$, and $P\{X_1 \geq X_2\} = a + b$
* Moreover, $P\{X_1 = X_2\} = \sum_{k=0}^{\lfloor n/2 \rfloor} P\{X_1 = k, X_2 = k\}$

$$= \sum_{k=0}^{\lfloor n/2 \rfloor} \overset{(1)}{\binom{n}{k}} \overset{(2)}{\binom{n-k}{k}} \overset{(3)}{\left(\frac{1}{n}\right)^k} \overset{(4)}{\left(\frac{1}{n}\right)^k} \overset{(5)}{\left(1 - \frac{2}{n}\right)^{n-2k}}$$

$$= \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{n!}{(k!)^2 (n-2k)!} \cdot \frac{1}{n^{2k}} \left(1 - \frac{2}{n}\right)^{n-2k}$$

(1) Choose k throws among n throws to go to bin 1
(2) Choose k throws among $(n-k)$ to go to bin 2
(3) $P\{$k throws go to bin 1$\}$, which is $\frac{1}{n}$ for each throw
(4) $P\{$ —— 2 $\}$, ——
(5) With probability $(1 - \frac{2}{n})$, a throw goes to a bin nb $i \neq 1,2$ for the other $(n-2k)$ bins